NatureScot Research Report 1304 - Scottish marine biodiversity data review - stakeholder analysis to inform the marine species and habitats data infrastructure and management requirements for improved availability, accessibility and use in Scotland

nature.scot/doc/naturescot-research-report-1304-scottish-marine-biodiversity-data-review-stakeholder-analysis-inform

Contents

Keywords

marine; biodiversity; dataflow; improvement; infrastructure; species; habitats; available; accessible; data

Executive summary

Purpose of this review

High quality, current and accessible marine species and habitat data are essential to support marine environmental policy and planning decisions. Achieving greater exchange and interoperability of data within the marine sector will help support the transformation required to meet the ambitious commitments set by Scottish Government to reach Net Zero-emissions by 2045 and tackle both the climate emergency and biodiversity loss.

An investigation of the terrestrial and freshwater <u>Biological Recording Infrastructure in Scotland</u> was undertaken by the Scottish Biodiversity Information Forum (SBIF) and published in 2018 (the "SBIF Review"). NatureScot commissioned this comparable analysis, as an adjunct to the original SBIF Review, for Scottish marine biodiversity data. The aim of this marine review was to explore and determine limitations to the existing infrastructure, through engagement with key stakeholders, and present recommendations that will make the management and use of marine species and habitat data more consistent, joined up and accessible to the marine community in future.

Key findings

A large amount of Scottish marine data flows into the Scottish and wider UK infrastructures (e.g., databases, repositories, portals, and the Marine Environmental Data and Information Network (MEDIN)) at varying levels of efficiency; ranging from well-established automated workflows to ad-hoc or non-automatic workflows, depending on the biodiversity receptor (mammals; benthic; birds; fish) in question and the organisations contributing the data. Difficulties in identifying, accessing and using marine biodiversity data persist; it is widely recognised that current data flows could be simplified and that there are still barriers to be overcome with data sharing, spatial resolution and coverage. The existing framework and mechanisms to mobilise and access the wide range of existing marine biodiversity datasets can be labour intensive and inefficient.

A key strength of the established data flow and systems is the ability to support the large volume of species and habitat data that are recorded and shared by Government bodies, but there is a widespread lack of clarity regarding roles, responsibilities and processes. Historical under-funding is a contributory factor, which has also limited capacity to capitalise on new infrastructure and advances in technology (e.g., cloud-based systems and widespread use of Application Programming Interfaces (APIs)). The lack of dedicated resource and skills within the existing infrastructure for efficient provision and management of commercial and third sector (NGO and citizen science) data is an issue, combined with cultural and behavioural barriers to data sharing, and prevents this data being easily and fully incorporated into the marine evidence base.

Technical and cultural barriers to data sharing in particular impact on the availability, quality, and accessibility of data for collation and use by others. The review found that relatively large quantities of data were still stored locally and not fully incorporated into the data flow network. Gaps in data availability were identified, in part, are likely to be a result of a combination of data not being properly shared, organised or due to resource driven workflow time-lags.

In summary, the key barriers to sharing marine species and habitats data identified in this review, preventing wider re-use, relate to:

1. **Cultural and behavioural barriers**, which range from reticence to share data for commercial reasons to barriers resulting from concerns over how the data might be used or misused / misinterpreted; some data providers requiring embargoes on releasing data, concerns about not receiving proper credit, and lack of incentives for sharing data.
2. **Practical barriers,** which range from a lack of data integration (linked to the adoption of standards and consistent data formats) causing bottlenecks in the process to not understanding how to make data available in meaningful ways, due to:

- Lack of technical expertise and/or capacity to deliver;
- Lack of understanding of the existing infrastructure and processes;
- Insufficient knowledge of how to adopt standardised classification systems and/or produce sufficient metadata linked to datasets;
- The variety of metadata and data standards, vocabularies, and ontologies that exist for marine biodiversity data, means that choosing the one(s) that best fits the data, methodology, and data management goals can be a time-consuming process;

3. **Inadequate strategies and resources** that result in data often being made available in an opportunistic manner rather than being focused on need and frequently without sufficient resources:

- Underestimation of the costs of managing data, particularly maintenance of 'live' information, can lead to abandonment or slow decay of data collected through discrete projects.
- Lack of time and resources to devote to learning new standards and performing data transformation is also a recurring and long-standing issue.

Priority actions

The resulting top priorities to better coordinate and streamline the flow of Scottish biodiversity data, identified through this analysis of stakeholder needs, are:

- **Governance and stakeholder engagement:** Ensure future governance of marine data management in Scotland;
- **Data sharing:** Unification and/or rationalisation of databases and portals with clear guidance on where to submit datasets and the dissemination opportunities; data sharing needs to become routine;
- **Data accessibility:** Clear sign-posting to data resources, with consistent use of persistent identifiers to ensure that data are easily accessible for use in collations without duplication;
- **Data availability:** Address gaps in data availability through increased data sharing and publication and removal of data flow bottlenecks; improved access to industry survey data and ensure that academic research data are available within the Scottish / UK data landscape as well as internationally;
- **Collaboration and commitment:** Widespread adoption of Findable, Accessible, Interoperable and Reusable (FAIR) data principles to improve current practices (e.g., on data verification, adoption of standards, culture of sharing and 'open-ness' of data).

High-level recommendations

Committed investment into skills resource and data infrastructures now will lead to long-term impact and savings (time and money) in the future, by creating more streamlined data workflows and accessible and reusable data. A prioritisation matrix was used to score the high-level recommendations, firstly based on their impact or value (reward) and secondly on the effort or investment (time, money) needed to complete them.
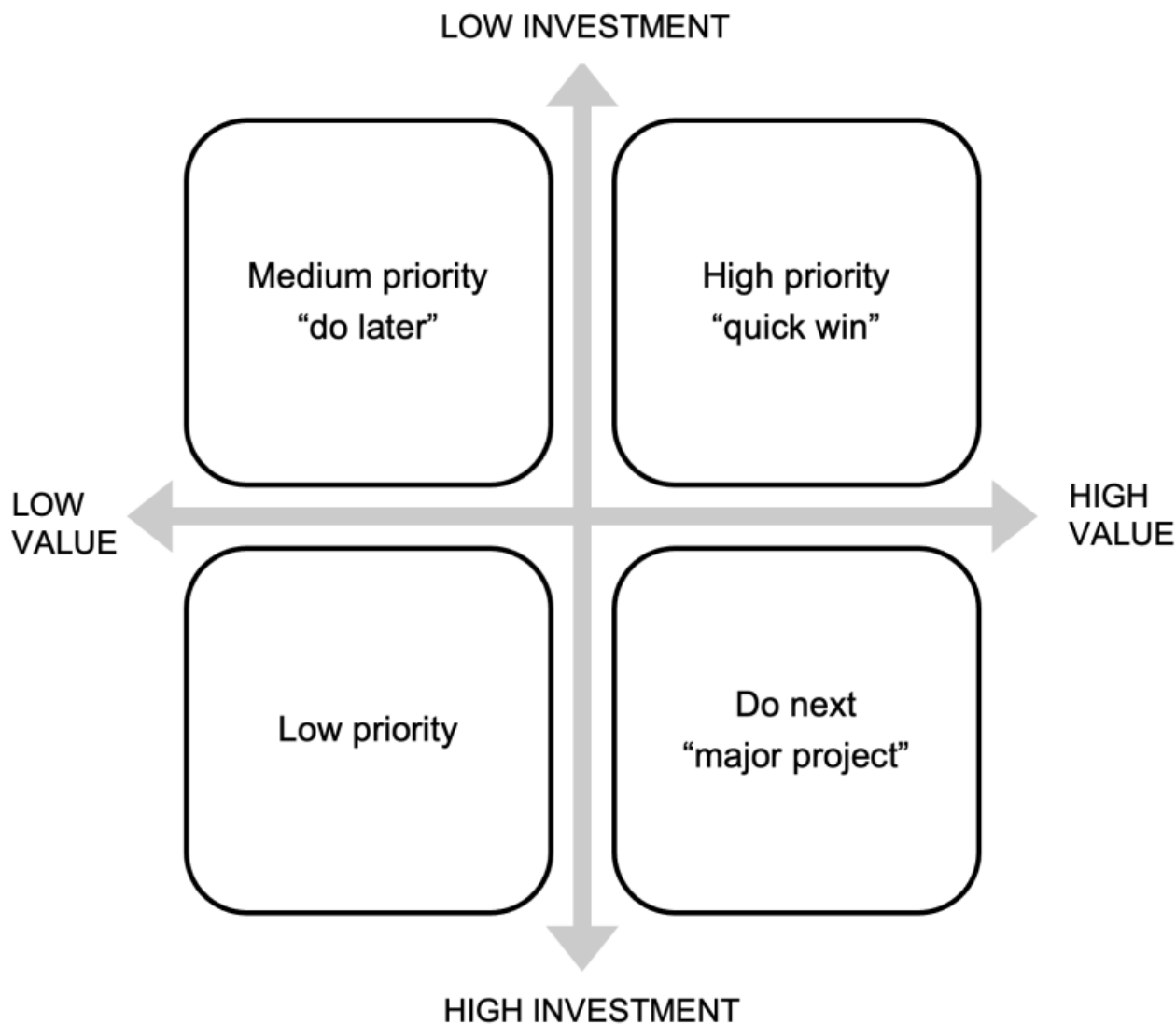
Prioritisation matrix used to score the high-level recommendations
Click for a full description

The **25 recommendations** outlined below address the issues and barriers associated with the existing marine data infrastructure and data availability, accessibility and quality; brigaded under six themes. The recommendations are scored as per the prioritisation matrix; 11 out of the 25 recommendations made in this review are considered 'quick wins', where low investment is required but a high value is gained.

Some recommendations are specific to Scottish data flow, however the majority could or would have UK-wide implications and/or benefits. Many of the recommendations are centred on continued negotiation and discussions with stakeholders, cultural change and behavioural adaptation, building on and improving existing skills, systems and workflows, and better sign-posting; rather than huge infrastructure change.

The recommendations are listed according to theme and scored to indicate their priority.

THEME 1: Continued engagement with key stakeholders

Recommendation	Prioritisation score
RECOMMENDATION 16: Develop proactive engagement with data custodian stakeholders who weren't fully involved in the review	Do next "major project"

Recommendation	Prioritisation score
RECOMMENDATION 25: Ensure future governance of marine data management in Scotland	Do next "major project"

THEME 2: Clarifying and streamlining data flows

Recommendation	Prioritisation score
RECOMMENDATION 3: Adopt primacy of affiliated data submission routes	High "quick win"
RECOMMENDATION 4: Map out marine data flows holistically	Low priority
RECOMMENDATION 5: Adopt primacy of Marine Recorder Online	High "quick win"
RECOMMENDATION 7: Agree a single, central route for casual records	High "quick win"
RECOMMENDATION 9: Formalise data flows between DASSH and the NBN Atlas	Do next "major project"
RECOMMENDATION 11: Clarify workflow responsibilities for mobilising benthic records to NBN Atlas	High "quick win"
RECOMMENDATION 20: Provision of biodiversity records collected under licence or for consent into the MEDIN Data Archive Centre network	High "quick win"

THEME 3: Improving the quality of existing data management

Recommendation	Prioritisation score
RECOMMENDATION 6: Clarify responsibility for tagging of records of conservation importance	High "quick win"
RECOMMENDATION 8: Each record submitted to have a persistent identifier (PID) to prevent duplication	High "quick win"
RECOMMENDATION 12: Progress a verification protocol for imagery derived data that complements the existing NMBAQC scheme component for grab and core sediment derived data	Do next "major project"
RECOMMENDATION 15: Plan for and fund the management and sharing of all new data being collected	High "quick win"

Recommendation	Prioritisation score
RECOMMENDATION 21: Maintain data version control through encouraging active custodianship	High "quick win"
RECOMMENDATION 22: Optimise re-use of data through adherence with FAIR Data Principles	High "quick win"

THEME 4: Investing in infrastructure and resource (people skills and funding)

Recommendation	Prioritisation score
RECOMMENDATION 2: Scottish (and UK) Government recognise and resource key skills and infrastructure across the full data lifecycle	Do next "major project"
RECOMMENDATION 19: Invest in data engineers and allocate resource for system decommissioning	Do next "major project"

THEME 5: Improving existing and creating new data infrastructure

Recommendation	Prioritisation score
RECOMMENDATION 1: Undertake a UK-wide marine biodiversity data infrastructure assessment	Do next "major project"
RECOMMENDATION 10: Develop infrastructure to support viewing and download of habitat records	Do next "major project"
RECOMMENDATION 13: Provide infrastructure and data management support for citizen science marine biodiversity recording	Do next "major project"
RECOMMENDATION 17: Develop simplified user interfaces onto repositories to support wider data submission	Do next "major project"
RECOMMENDATION 23: Develop existing portal infrastructure to support efficient searching, data display and dataset collation	Do next "major project"
RECOMMENDATION 24: Embed marine expertise in, and interoperability of, the National and Regional (LERC) hubs infrastructure in Scotland	Medium "Do later"

THEME 6: Simplifying existing and creating new guidance

Recommendation	Prioritisation score
RECOMMENDATION 14: Simplify the requirements for submitting data into DASSH whilst maintaining data quality	Medium "Do later"
RECOMMENDATION 18: Develop guidance on optimal data submission pathways	High "quick win"

Abbreviations & acronyms

API – Application Programming Interface

BBD – Better Biodiversity Data

BEIS – Business, Energy and Industrial Strategy

BioDIG – Biological Data and Information Group

BODC – British Oceanographic Data Centre

BTO – British Trust of Ornithology

CC-BY / CC-BY-NC – Creative Commons-by attribution / non-commercial

CES – Crown Estate Scotland

DAC – Data Archive Centre

DASSH – Archive for Marine Species and Habitats Data

DATRAS – Database of Trawl Surveys

Defra – Department for Environment, Food and Rural Affairs

DOI – Digital Object Identifier

DwC-A – Darwin Core Archive

eDNA – Environmental DNA (Deoxyribonucleic acid)

EIA – Environmental Impact Assessment

EMODNet – European Marine Observation and Data Network

eNGO – Environmental Non-Government Organisation

EurOBIS / OBIS – European Ocean Biodiversity and Information System

FAIR – Findable, Accessible, Interoperable, Reusable

FAME/STAR – Future of the Atlantic Marine Environment / Seabird Tracking and Research

GBIF – Global Biodiversity Information Forum

GeMS – Geodatabase of Marine Features in Scotland

GOOS – Global Ocean Observing System

HBDSEG – Healthy and Biologically Diverse Seas and Evidence Group

HWDT – Hebridean Whale and Dolphin Trust

ICES – International Council for the Exploration of the Sea

IDDP – Integrated Digital Data Platform

INSPIRE – Infrastructure for Spatial Information in the European Community

JCDP – Joint Cetacean Data Programme

JNCC – Joint Nature Conservation Committee

LERC – Local Environmental Record Centre

MarClim – Marine Biodiversity and Climate Change

MARG – Monitoring and Assessment Reporting Group

MarLIN – Marine Life Information Network

MarPAMM – Marine Protected Areas Management and Monitoring

MASTS – Marine Alliance for Science and Technology for Scotland

MBA – Marine Biological Association

MCS – Marine Conservation Society

MEDIN – Marine Environmental Data and Information Network

MPA – Marine Protected Area

MRO – Marine Recorder Online

MSBIAS – Marine Species of the British Isles and Adjacent Seas

MSCC – Marine Science Co-ordination Committee

MSFD – Marine Strategy Framework Directive

MS / MSS – Marine Scotland / Marine Scotland Science

NAFC – North Atlantic Fisheries College

NBN – National Biodiversity Network

NCBI – National Centre for Biotechnology Information

NMBAQC – NE Atlantic Marine Biological Analytical Quality Control

NMPi – National Marine Plan Interactive

OECD – Organisation for Economic Co-operation and Development

OGL – Open Government Licence

OIC – Orkney Islands Council

PAG – Project Advisory Group

PAM – Passive Acoustic Monitoring

PID – Persistent Identifier

PITT – Passive Integrated Transponder Tag

PMF – Priority Marine Feature

QA – Quality Assurance

QC – Quality Control

REF – Research Excellence Framework

RSPB – Royal Society for Protection of Birds

SAMS – Scottish Association for Marine Science

SBIF – Scottish Biodiversity Information Forum

SCANS – Small Cetaceans in European Atlantic waters and the North Sea

SEPA – Scottish Environmental Protection Agency

SIRMP – Shetland Islands Regional Marine Plan

SMA – Scotland's Marine Atlas

SMRU – Sea Mammal Research Unit (University of St Andrews)

SNCB (CNCB) – Statutory Nature Conservation Body (Country Nature Conservation Body)

SOTEAG – Shetland Oil Terminal Environmental Advisory Group

SWT – Scottish Wildlife Trust

UHI – University of Highlands and Islands

UK – United Kingdom

UKMMAS – UK Marine Monitoring and Assessment Strategy

WDC – Whale and Dolphin Conservation

WeBS – Wetland Bird Survey

WFS – Web Feature Service

WMS – Web Mapping Service

WoRMS – World Register of Marine Species

Definitions

Definition of terms regarding marine biodiversity data flows in this review:

Application Programming Interface (API): An API is a software intermediary (set of definitions and protocols) that delivers a request to a server and then relays a response back to the client, allowing interaction between two applications.

Automated workflow: Based on pre-defined tasks the workflow runs on its own without any human intervention.

Database: A digital infrastructure set-up for accessing, storing, managing and curating data. Provides a 'digital' collation of multiple data sources that are of the same format, standard etc. Data are actively managed, curated and processed and data flows are mediated.

Data aggregator: An infrastructure (or organisation) that digitally collates data from many sources, provides some value-added processing, and repackages the result in a usable form.

Data custodian: The data holder. Responsible for maintaining the storage, security and integrity of the data.

Data end user: The individual or organisation who will ultimately be using the data (raw or products). *For example:* data analysts will likely want to download the data so they can make more sophisticated use of the data, whereas general public are likely to want simple data exploration tools, wrapped up in a narrative that puts the data into context.

Data engineer: A data engineer develops and constructs data products and services and integrates them into systems and business processes.

Data manager: The individual responsible for developing and/or governing data-oriented systems and for maintaining, curating and mobilising data resources.

Data originator: The organisation or individual which commissions and/or produces the data.

Data provider: The organisation or individual which submits and/or publishes data.

Data products: Model or analysis output that uses processed (derived) data that has spatial and temporal resolution and been subject to quality control (completeness, consistency and space/time uniformity) as input (EMODNet definition).

Data validation: The process of determining whether data falls within the acceptable range of values for a given field. Validation typically occurs when a record is initially created or updated.

Data verification: Verification performs a check of the data to ensure that it is accurate, consistent, and reflects its intended purpose. Verification may take place as part of a recurring data quality process and plays an especially critical role when data is migrated from or merged with other data sources.

Duplicated effort: Where data is being submitted to the same place via different pathways or where their databases / portals are not linked up so data must be submitted to several different systems (databases, repositories etc).

Dysfunctional workflow: Where data should be submitted but isn't / where data flows should happen but don't.

Functional requirements: defines what the system should do; usually defined as a process (i.e., system features and user requirements).

Marine community: Stakeholders that have varying levels of interest and investment in the data flow lifecycle (e.g., collection, management, use). Each party relies on and interacts with one another to obtain access to and use of the various types of existing marine data.

Missing links: Where data exists, but is not used / where data systems (databases, repositories etc) exist but are not connected.

Non-functional requirements: defines how the system should do it; everything that makes the process happen (i.e., system properties and user expectations).

Open data: Data that is made available under open licence, such as the Open Government Licence (OGL) or Creative Commons (CC), in a in a common, machine-readable format so that anyone can access, use and share it.

(Open) data portal: An online application interface which supports end users in searching (including via API), accessing or downloading data that they need. Portals are a way of sharing data (openly) for the benefit of others.

Processed data: Full spatial and temporal resolution geo- and time-referenced data that has been QC checked (edited, cleaned, transformed etc from its raw form).

Raw data: Unprocessed data (this could include photos, video footage, sound recordings or simply manual observations or counts of species at a location at a given time) at full resolution, including synchronisation methods and excluding communication artefacts (EMODNet definition).

Record: Comprised of uniquely named components (data fields) within a database structure. Each row in a database table is a record; it contains information on a single data item.

Repository: Any central data storage infrastructure (also known as a data archive or library). A data repository can aggregate data from multiple sources, without the data being necessarily related. *For example*, a 'data lake' is a large data repository that stores unstructured data that is classified and tagged with metadata.

Web Feature Service (WFS): A WFS returns features with geometry and attributes that clients can use in geospatial analysis. WFS services also support filters that allow spatial and attribute queries to be performed on the data.

Web Mapping Service (WMS): A WMS is a standard protocol for serving georeferenced map images over the internet; typically produced by a map server from data provided by a GIS database.

Acknowledgements

Introduction

Marine ecosystems are in decline, yet we also have more marine data, and more data portals, than ever before. High quality, current and accessible marine species and habitat data are essential to support marine environmental policy and planning decisions. Likewise, an efficient marine biodiversity data flow network is key to informing strategy on how the environment is managed to benefit people and meet future challenges. Achieving greater exchange and interoperability of data within the marine sector will help support the transformation required to meet the ambitious commitments set by Scottish Government to reach Net Zero-emissions by 2045 and tackle both the climate emergency and biodiversity loss.

This review examines the existing data flow pathways and infrastructure supporting the collation and use of marine species and habitats data in Scotland. The review was developed by firstly investigating the barriers and issues that data end users, data managers and data providers encounter, and then mapping out the current marine species and habitats data landscape (systems, tools, data and challenges) in Scotland across key receptors (marine mammals, birds, benthic species and habitats, fish and shellfish) and sectors (public, industry, environmental Non-Governmental Organisations (eNGOs) and academia). This enabled clear identification of where improvements are required to achieve greater exchange and access to biodiversity data within the marine community. The analysis has been done in collaboration with key sector stakeholders; broadly defined as the users and providers of Scottish marine biodiversity data, and a suite of recommendations are made on the basis of the stakeholder needs and requirements identified.

This introduction describes the background and rationale for the analysis, the primary objectives, and the structure of this document.

Background and rationale

There is considerable interest in improving the current framework of data use, from collection and submission, to access, collation and application. Data providers have become increasingly willing, and in some cases compelled, to share information. However, difficulties in identifying, accessing and using marine biodiversity data persist; it is widely recognised that current data flows could be simplified and that there are still barriers to be overcome with data sharing, spatial resolution and coverage. The existing framework and mechanisms to mobilise and access the wide range of existing marine biodiversity datasets can be labour intensive and inefficient.

Limited data discovery and re-use, as a result of restricted data sharing, limited metadata and complex data workflows, was a driving force behind the formation of the Scottish Biodiversity Information Forum (SBIF) in 2010 following a public petition (PE1229) calling on the Scottish Parliament to:

"Urge the Scottish Government to establish integrated local and national structures for collecting, analysing and sharing biological data to inform decision making processes to benefit biodiversity."

Around the same time as the public petition, the Scottish Marine Science Strategy (2010-2015) also highlighted issues with marine data and called for an organised infrastructure to cope with the volumes, technological developments and legislative requirements. It further promoted openness to allow the re-use of data and highlighted the policy need for better integrated advice. The UK Government's commitment, primarily through the Marine Science Co-ordination Committee (MSCC), to opening up access to as much marine data as possible is detailed in the UK Marine Science Strategy. This commitment includes opening up access to industry data.

The Marine (Scotland) Act and the reporting requirements of the Marine Strategy Framework Directive (MSFD), the OSPAR Convention, and most recently the commitment to enhance marine environmental protection are part of the Bute House Agreement. These drive the need for scientifically robust quality assured marine data to underpin legislation, planning and Marine Protected Area (MPA) designation and management, and to support the Scottish Government's Blue Economy vision for Scotland.

The 'SBIF Review'

An investigation of the Biological Recording Infrastructure in Scotland was undertaken by the Scottish Biodiversity Information Forum (SBIF) and published in 2018 (the "SBIF Review") setting out recommendations to achieve five main outcomes, including transformed data flows, service provision, governance, and funding by 2025. The SBIF Review demonstrated many of the problems arising from current data workflows and highlighted the need for further discussion around improvements to the biological recording infrastructure across the rest of the UK. However, the review was focussed on the terrestrial and freshwater environment with limited engagement with key stakeholders in the marine sector, and as a result there is currently no document setting out what the requirements and priorities are for the marine biodiversity data management and infrastructure in Scotland.

Objectives and scope of this analysis

NatureScot commissioned this comparable* analysis for Scottish marine biodiversity data as an adjunct to the SBIF Review to explore and present recommendations that will make the management of marine species and habitat data more consistent, joined up and accessible to the marine community. Encouraging FAIR (Findable, Accessible, Interoperable, Reusable) Data Principles and supporting the UK Marine Environmental Data and Information Network (MEDIN) objectives for more integrated data management and sharing, as well as compliance with the INSPIRE Directive to make data publicly accessible, are core to this current review.

The timeframe of engagement with stakeholders in this Marine Review was conducted over a significantly shorter (9 months) period than that in the SBIF Review which took two years to complete. The stakeholder landscape in terrestrial and freshwater recording is also significantly different compared with marine; eNGOs and Local Environmental Record Centre (LERC) volunteer recorders versus largely government-led organisations, respectively. Therefore, the approach taken to engage with stakeholders in this marine review does not entirely replicate that of the SBIF Review.

This report considers how marine biodiversity data recording works in Scotland at present and provides an overview of the issues and challenges identified by stakeholders relating to finding, accessing and using data. Data workflows, data submission and data access points were all considered within scope of this analysis. However, the technical detail of the database infrastructures supporting the storage and management of data were excluded from scope as this area of improvement work is covered though various other ongoing UK projects (e.g., the re-development of Marine Recorder; the development of the Joint Cetacean Data Programme (JCDP); and the JNCC's Big Picture work). All types of marine (including within the intertidal and estuarine areas) species and habitats data were considered, with focus given to the following receptors: seabed species and habitats, mammals (cetaceans and seals), fish, and marine birds.

The overarching aim of this analysis work was to provide recommendations for improvements to coordinate and streamline the flow of Scottish marine species and habitats records between organisations into existing downstream infrastructures. The four primary objectives were to:

1. Engage with stakeholders across sectors to document where the marine community feel the barriers or gaps to efficient collation, access and use of marine biodiversity data exist;
2. Explore and understand: (1) what individuals / organisations do with the marine biodiversity data that they hold; (2) where commonalities in the data workflows of big marine players exist and how well these are currently integrated; (3) and what data it is that data users need to access and use both now and in the future;
3. Work together with the UK Marine Environmental Data and Information Network (MEDIN) and its archive for marine species and habitats data (DASSH) to foster join-up between Scottish and UK aspirations for marine biodiversity data recording.
4. Optimise interfaces between marine and terrestrial sectors and identify mutual opportunities to create synergy and facilitate join-up and/or sharing of resources.

The Healthy & Biologically Diverse Seas Evidence Group (HBDSEG) for the UK Marine Monitoring and Assessment Strategy (UKMMAS) has established a Biodiversity Data and Information Group (BioDIG) that are actively looking at improvements to UK marine biodiversity data flows. It is therefore important that the recommended improvements, made in this report, for marine biodiversity data flows in Scotland for each biodiversity receptor are made in collaboration with the vision for UK data flows and MEDIN.

Report structure

Section 1

Introduction - presents the overall aim of the project and provides background and rationale to undertaking the review, outlining the key drivers behind improving the accessibility and availability of marine species and habitats data.

Section 2

Methodology - describes the methods used to review the existing infrastructure for managing and accessing Scottish marine species and habitats data through literature review and a stakeholder engagement plan comprising questionnaire, video conference interviews and a workshop with key sector organisations and individuals.

Section 3

The existing Scottish biodiversity data landscape - describes each database and portal that receives Scottish marine species and habitats data and provides a data flow mapping sketch for key biodiversity receptors. Stakeholder questionnaire feedback is used to describe current data flow publication routes and access points.

Section 4

Scottish benthic data flow options appraisal - presents options for improvements to the existing 'as-is' flow of Scottish benthic data into UK infrastructure to facilitate data sharing, access and re-use.

Section 5

Issues and challenges identified by stakeholders - provides an analysis of the stakeholder questionnaire results and evidence from 1-2-1 stakeholder discussions to present the main barriers inhibiting the accessibility and availability of marine species and habitats data.

Section 6

Summary of stakeholder needs and requirements - presents the functional and non-functional improvements required by stakeholders to enable more efficient access and use of species and habitats data in Scotland. These are underpinned by the results from the stakeholder questionnaire and data flow mapping exercise.

Section 7

Applicability of the SBIF Review recommendations - presents a brief summation of the applicability of the original 2018 SBIF Review recommendations and current work through the SBIF Better Biodiversity Data (BBD) Project for terrestrial and freshwater biological data recording to the marine biodiversity data community's needs and requirements.

Section 8

Prioritisation of recommendations made in this marine review - presents the priority of recommendations made in the report using a prioritisation matrix to identify recommendations that are 'Quick Wins'; low investment, high value, and those that are bigger projects; high investment, high value. The recommendations are also assessed for dependencies on each other.

Section 9

Conclusions and next steps - summarises the key findings of the report, the key priorities and next steps to implementing the recommendations and finding solutions to the issues and barriers identified.

Methodology – approach to stakeholder engagement

Review of published literature and other initiatives

A review of existing relevant work that is ongoing or has been published was done at the time of scoping and kept under review to ensure that the marine analysis work built on and complemented existing work and avoided duplication. Publications most relevant to this review, include:

- SBIFs' "A Review of the Biological Recording Infrastructure in Scotland" and subsequent SBIF Better Biodiversity Data (BBD) Project. The BBD project will establish a National Biodiversity Data Hub for Scotland that can provide leadership and coordination to support delivery of citizen science (via Local Environmental Record Centres (LERC)) biodiversity data at both national and regional levels. The project aims to establish a consortium, similar to the arrangements in Wales, with the existing LERCs in Scotland forming the basis of the National Hub. The intention is that a business model will be developed to generate sufficient income from added value data services to data clients, such that the Hub should be able to support itself.
- A JNCC-led piece of work (2021/2022), funded by the Department for the Environment, Food and Rural Affairs (Defra), to investigate the flow of data from monitoring programmes undertaken by individual UK Statutory Nature Conservation Bodies (SNCBs) into UK Marine Strategy (UK MS) indicator assessments. Data flow diagrams for UK benthic habitats, cetaceans and seals have been created and recommendations made for streamlining and expanding the flow of data in to UK MS assessments.
- The JNCC's review of the current biodiversity data usage across the four UK Country Nature Conservation Bodies, which included investigation of marine data usage. The report identifies and discusses the main limitations of the current marine and terrestrial data workflows and suggests future improvements.

- A review of the species data landscape in England was commissioned in October 2020 by the Cabinet Office's Geospatial Commission, which included a review of the SBIF recommendations for Scotland applicability to England. The scope of this review includes a short section on marine biological data; however, it is focussed only on species, not habitat, data.
- A joint OECD Working Paper in collaboration with the UK Marine Environmental Data and Information Network (MEDIN) and the Global Ocean Observing System (GOOS), reported the value chains in public marine data (a UK case study).
- Marine Scotland and Crown Estate Scotland's industry 'developer data archiving' project (initiated Q4 2021/2022) to explore and define a formal process for handling the data, evidence and information that is provided or requested as part of the Scottish offshore wind and marine renewables consenting process. The aim is to develop guidance that will help manage the flow of information from applicants through to delivery of useful, compatible and open data mobilised into UK MEDIN Data Archive Centres (or the most appropriate repositories).
- A Review of Access to Industry Marine Environmental Data, funded by the Marine Management Organisation, Marine Scotland, The Crown Estate and MEDIN, was published in 2015 (ABPmer, 2015). The report identified where data are not made publicly available and the barriers that are preventing the provision of these datasets.

Project governance

This current review was overseen by a Project Advisory Group (PAG) of 15 members from governmental bodies, private sector, academic institutions and eNGOs. The PAG was established from members of the UK BioDIG, a MEDIN-led technical sub-group of HBDSEG, who have an interest in Scottish marine biodiversity data plus other key marine data stakeholders that have experience in collating, using and/or managing marine data. The organisations represented are listed in Annex A.

This PAG was tasked with providing strategic direction for the review. For example, by helping to identify key issues with existing marine data workflows and opportunities for improvement and collaboration with other existing projects at a Scottish and UK scale. The PAG was also asked to identify prioritised recommendations for improvements in Scottish marine data management including the effective translation of stakeholder requirements and needs into changes in the existing data flow pathways and infrastructure, while ensuring that data and systems are interoperable at a UK scale.

A small internal NatureScot project working group provided project support.

Evidence gathering and stakeholder consultation

To understand how the marine biodiversity data infrastructure is currently operating and to seek ideas for improvements, stakeholder engagement consisted of: 1-2-1 informal interviews with 54 individuals from 35 organisations; continuous and iterative follow-up stakeholder discussions; a questionnaire advertised publicly through MASTS, MEDIN and the NBN; and a workshop. This engagement process took place between April 2021 and December 2021, following an initial project scoping exercise. A list of the stakeholder organisations involved in the review are listed in Annex B.

The existing Scottish biodiversity data landscape

The first step in undertaking this marine review was to obtain a clear understanding of the current state of the data flow pathways and infrastructure supporting the sharing, collation and use of marine species and habitats data in Scotland.

This section gives an overview of how marine species and habitats data flow from recorders to end users along the biodiversity data pathway. It explains how data are currently managed, collated and distributed in Scotland and identifies the main organisations involved. The findings in this section are a result of a literature review, stakeholder interviews and a stakeholder questionnaire.

Delivering a successful data flow pathway means:

- Identifying and adopting appropriate standards at all stages;
- Supporting quality assurance and verification stages;
- Encouraging data mobilisation and archiving from all sectors;
- Reducing the amount of data duplication in the network;
- Making data available to end users in a timely manner.

The data flow mapping of the Scottish marine species and habitat data landscape presented in this review focussed on the data lifecycle stages (figure 1) that relate to the stewardship (i.e., quality assurance, collation, archiving and communication to ensure that data are made accessible with appropriate acknowledgement of the data originator) and end use (availability and accessibility) of marine data.
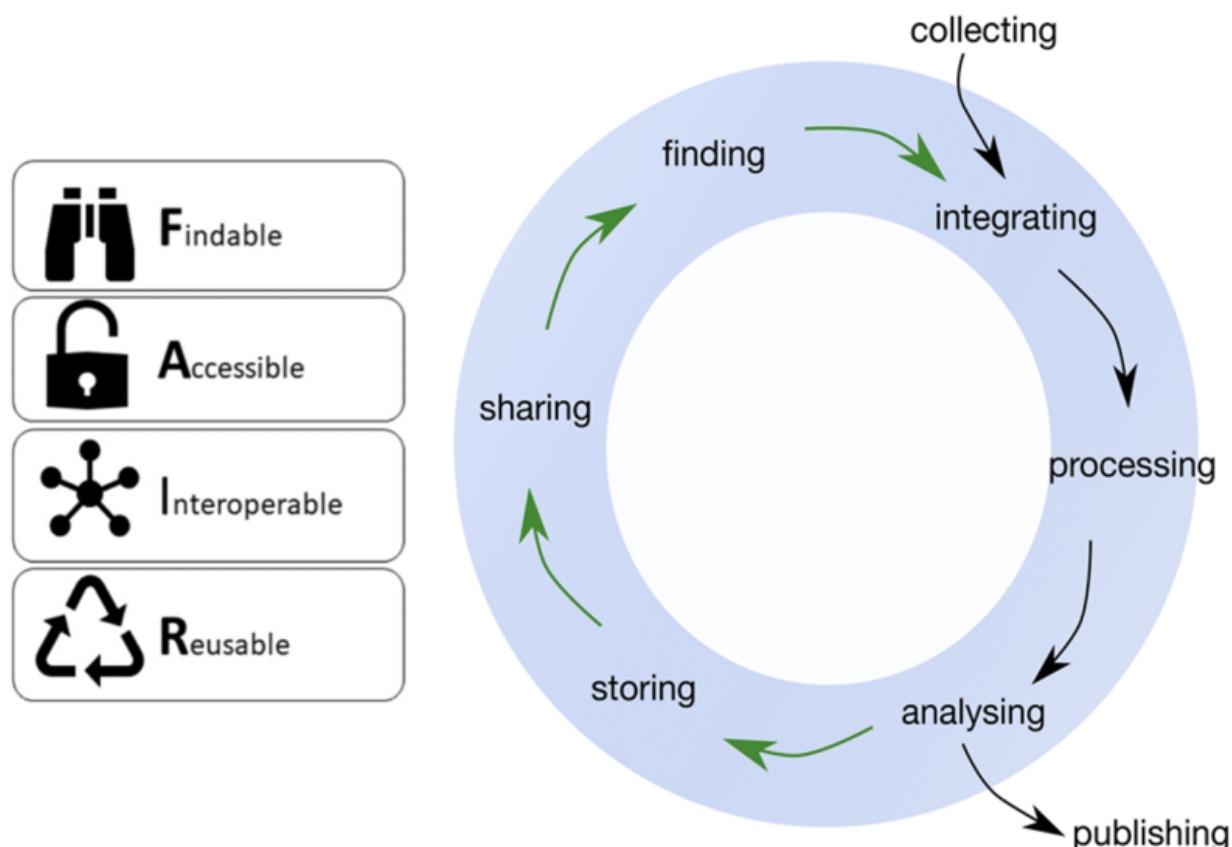

Figure 1: A simple data lifecycle framework
Click for a full description

Who is involved?

Government bodies are often instrumental in implementing and coordinating the infrastructures necessary to share and disseminate data and knowledge. NatureScot, the Joint Nature Conservation Committee (JNCC), Marine Scotland (MS), Scottish Environmental Protection Agency (SEPA) and Crown Estate Scotland (CES) are the key 'big players' engaged with this Scottish review, each bringing an individual perspective and different drivers behind their need to commission, access and use marine biodiversity data.

A large proportion of seabed species and habitat data (particularly historically) is professionally collected by government bodies and some academic institutions, rather than by volunteers, except in coastal and nearshore environments. This is at least partly a consequence of the high cost of sending large vessels offshore and data storage infrastructure costs. The UK national recording scheme, Seasearch, which focuses on volunteer diver collected benthic data is an exception to this. Technological advances are however facilitating improvements and diversification in data collection activities, e.g., by citizen science initiatives, environmental Non-Governmental Organisation (eNGOs) and regional government-funded research and monitoring projects.

Marine mega-fauna data recording is more closely aligned to some terrestrial recording models, being dominated by eNGO volunteer schemes. Cetacean data recording in particular relies heavily on regional citizen science initiatives, local volunteers and/or discrete projects.

Recent years have seen a huge expansion in offshore activities, including fisheries, with a variety of industry sectors now making use of the seabed (e.g., aggregate dredging, renewable energy generation, aquaculture, oil and gas, cables and pipelines). For many of these activities there are requirements to collect biodiversity data (on marine mammals, ornithology, benthic species and habitats, fish and shellfish) for characterisation, licensing and compliance monitoring. Independently these biodiversity datasets may have temporal and/or spatial limitations, but when standardised and combined with broad-scale survey data they provide an invaluable resource for a range of analyses to inform conservation and policy needs.

Data standards and controlled vocabularies

Data standards (including those for discovery metadata) are essential to enable easy discovery, aggregation and re-use of data. For datasets to be collated, there needs to be a commonality at least between core fields within the data and having defined standards to facilitate this. There is general agreement in the marine community, and reinforced by the stakeholders engaged in this review, that data standards should be in accordance with the FAIR principles and that data standards help facilitate the future collation of independent datasets collected in response to a range of different scientific and legislative drivers.

Standardising data offers advantages in four main areas: data quality, suitability of data for analyses, ease of data ingestion, and ease of mobilisation and compatibility. These components all facilitate data management, aggregation and interoperability and provide end users with a greater degree of confidence in the quality of data; saving time and contributing towards maximising the use of independent datasets.

A number of data standards are used in the marine biodiversity data flow pathway; some are applicable only to a particular process or data type and there is variable adoption of standards across the data flow landscape. The standards most relevant to marine biodiversity data are:

- Darwin Core (this standard has a species-centric focus; however, the custom ExtendedMeasurementOrFact (eMOF) extension of Darwin Core (OBIS-ENV) allows habitat data to be described using Darwin Core Archive (DwC-A));
- MEDIN Discovery metadata standard and data guidelines;
- The controlled taxonomic vocabulary Marine Species of the British Isles and Adjacent Seas (MSBIAS) (a sub-set of the World Register of Marine Species (WoRMS)).

Annex C contains further detail of each.

Review of portals and data aggregators

This sub-section summarises the portals and data aggregators that receive and publish Scottish marine data.

The UK Marine Environmental Data and Information Network (MEDIN) was established in 2008 as the 'hub' for UK marine data and provides the framework for the management and harmonisation of marine data in the UK (see Annex D for more detail). The ethos of MEDIN is to make data freely and openly available to end users. The MEDIN model operates using a series of standards and specifications, underpinned by a network of seven thematic accredited Data Archive Centres (DACs) (figure 2) and a Discovery Metadata Portal. Metadata published to the MEDIN Portal can also be published through to data.gov.uk, depending on the preference of the individual or organisation generating the metadata. Note that there is no efficient mechanism to publish Scottish marine biodiversity data, generated using the MEDIN online metadata tool, to the Scottish Spatial Data Infrastructure (SSDI) Metadata Portal.

UK organisations involved in collecting marine environmental data (in UK and non-UK waters (e.g., international research projects)) are encouraged to submit data to a relevant MEDIN DAC; this includes public, private and third sector organisations.
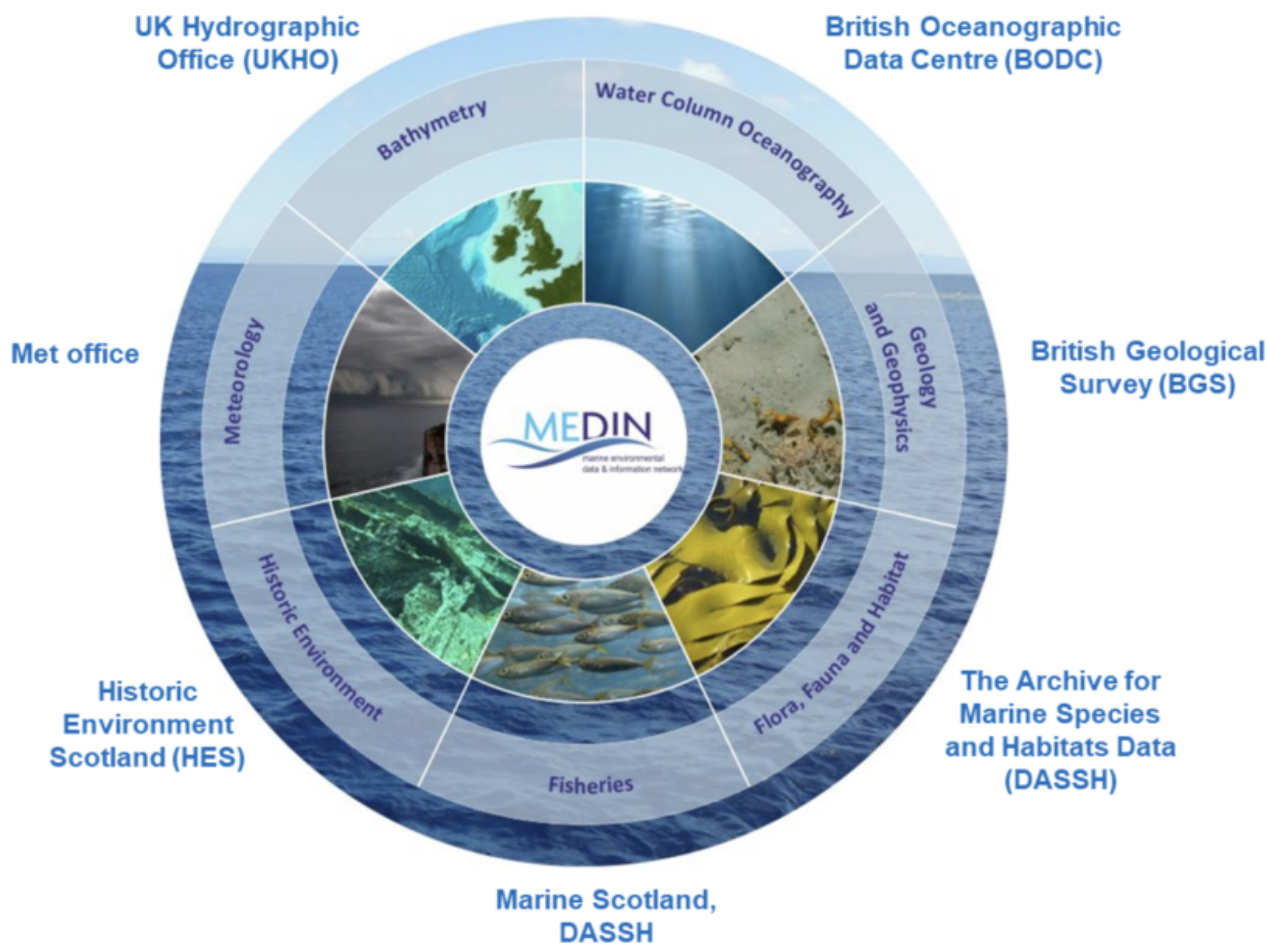
Figure 2: The seven accredited MEDIN data archive centres (DACs) for Scottish marine data
Click for a full description

The two marine biodiversity data focussed (DASSH and Marine Scotland (fisheries)) DACs are most relevant to the scope of this review and are described below; the other five DACs are focussed on physical data and out of scope.

The MEDIN accredited Scottish Fisheries Data Archive Centre (FishDAC) is hosted by Marine Scotland.

- Data submitted to international surveys at the International Council for the Exploration of the Sea (ICES) Data Centre are being made visible through the FishDAC and available for download.
- The FishDAC data is also published on Marine Scotland's own marinedata portal with DOIs.

DASSH is the archive for flora, fauna and habitat data (including benthic, mammal and seabird data). The DAC is hosted by the Marine Biological Association (MBA) and is core-funded by Defra and the Scottish Government (Marine Scotland) and MEDIN.

- DASSH is considered to be an integral part of the marine biodiversity data flow pathway, operating internationally as the UK node of the Ocean Biodiversity Information System (OBIS) and are a partner in EMODnet Biology.
- Data submitted to DASSH in an agreed standard format (including MEDIN data guidelines, DwC-A and Marine Recorder) can be ingested at zero-cost to the data provider with DASSH providing long-term archiving and publication services.
- DASSH mobilise marine species sightings data to the NBN Atlas; fully attributed data is archived within DASSH, whilst a summary view of the data is available through the NBN Atlas.

Other key portals and data aggregators that receive and publish Scottish marine biodiversity data include:

- Marine Scotland's National Marine Plan interactive (NMPi) portal [part of existing MS Open Data Network (MSODN)]
- NatureScot's Natural Spaces portal
- National Biodiversity Network (NBN) Atlas
- European Ocean Biodiversity Information System (EurOBIS) / European Marine Observation and Data Network (EMODNet)
- International Council for the Exploration of the Sea (ICES)

A description of each platforms purpose (niche) is detailed in Annex D.

The marine species and habitat data flows in Scotland (and the UK) can be complex, even for a single species group or receptor (figure 3 and Biodiversity receptor data flow mapping).

The flow of marine biodiversity data from repositories into portals and archive centres currently involves a series of non-automated or semi-automated workflows to clean, format and mobilise the data. There is a risk that lots of money is currently being spent replicating and maintaining high granularity systems at great cost to multiple organisations, draining resources for the workflow maintenance and development aspects needed to create a streamlined network infrastructure. The move to cloud-based platforms with spatial functionality and more user-focussed interfaces will streamline this process and link up systems, enabling increased use of Application Programming Interfaces (APIs) and web services to transfer and mobilise data products; the aim being to submit once, disseminate and use many times.

Figure 3 shows key repositories (e.g., receptor-specific databases, NBN Atlas and DASSH) to which Scottish marine biodiversity data are submitted to by data providers; the existing data flow pathways into the MEDIN DAC network; Scottish - UK portals (e.g., NMPi, NBN Atlas), international portals (e.g., GBIF, OBIS) and data aggregators (e.g., EMODNet, EurOBIS).
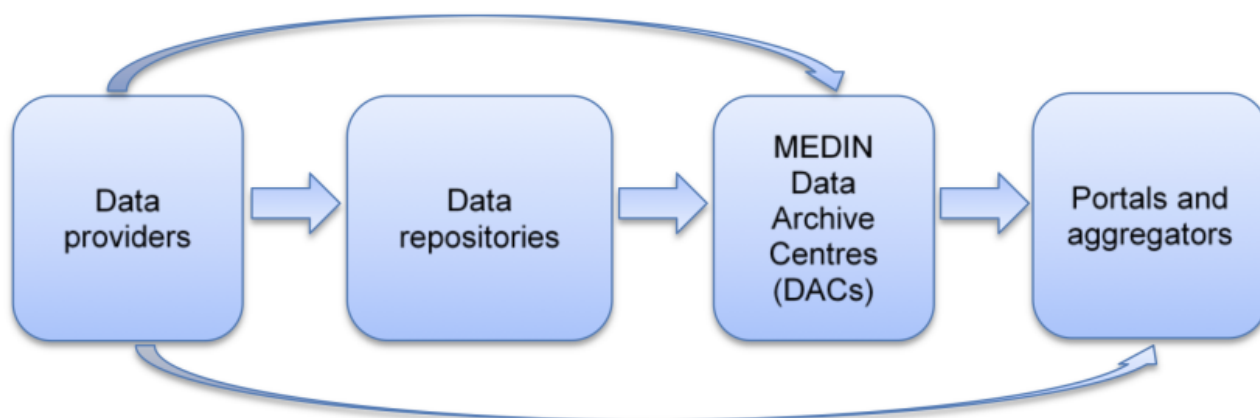
Figure 3: Simple (high-level) diagram of key components of the Scottish data flow landscape
Click for a full description

How data are currently published and accessed

This sub-section summarises the current data flow publication routes and data access points, based on stakeholder views expressed in the questionnaire and 1-2-1 discussions.

Data publication

MEDIN data archive centres, Marine Scotland's NMPi, and their own organisations' data portals (e.g., NatureScot's Natural Spaces) are the key portals and aggregators that Scottish marine biodiversity data is published to, based on stakeholder responses to the question *'Where do you send / mobilise or publish your marine biodiversity data to?'* (figure 4).

Data management systems (e.g., Marine Recorder) were not included; the above question was intended to capture information on where data providers make their marine records publicly available. It is difficult to determine from this result whether questionnaire respondents considered the workflows that exist between

DASSH and the NBN Atlas, for example. Caution must therefore be given to interpretation of the proportion of responses to, for example, NBN Atlas because there may be instances where respondents have included or excluded records that are submitted to DASSH (and then published onwards to the NBN Atlas) from their thinking when responding.

The establishment of MEDIN and its network of DACs as a 'hub' for marine data, combined with the fact that large quantities of habitat records are collected as an integral part of marine surveys (which are unsuitable for submission to the NBN Atlas), likely contributes to the lesser relative contribution to NBN Atlas and Local Environmental Record Centres (LERC) in figure 4. The smaller proportion of citizen science applications, LERC and NBN Atlas being cited is also likely due to the low number of third sector respondents. Figure 4 also highlights the relatively large proportion of marine biodiversity data that is being collected but not mobilised to established infrastructure, and is held on personal hard-drives or internal servers.
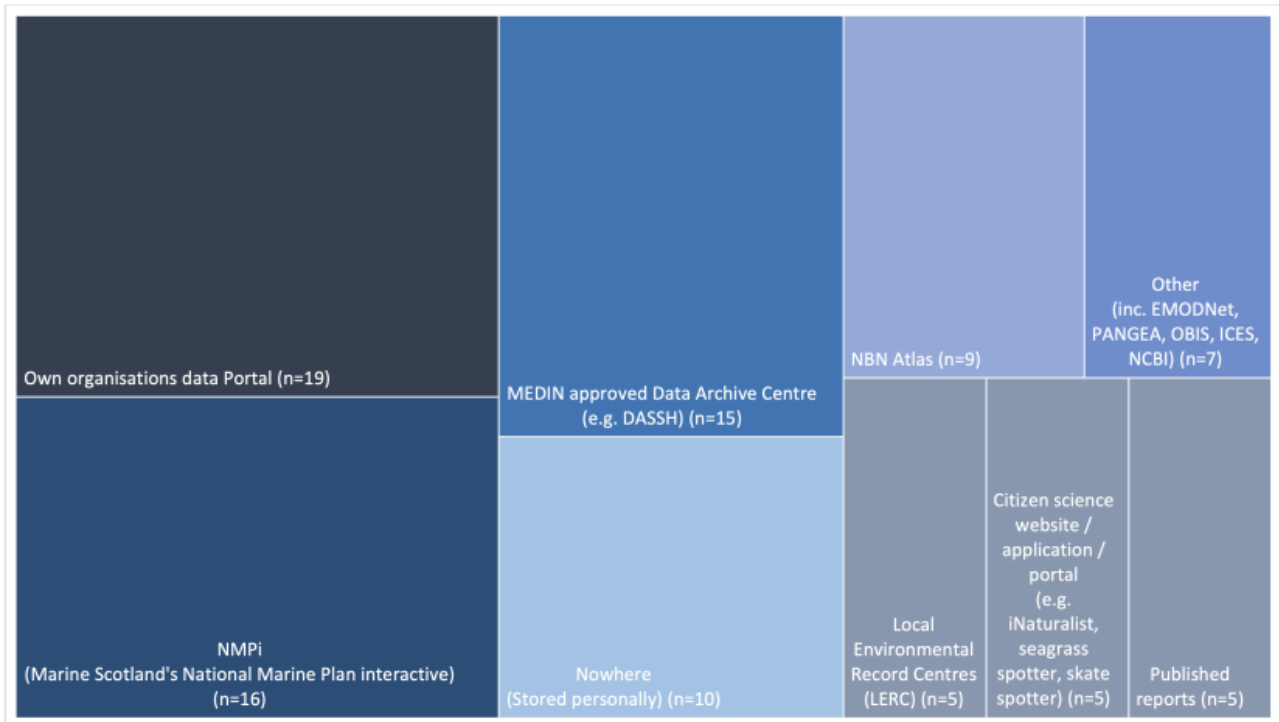
n= the number of respondents across all sectors

Figure 4: Portals and aggregators that respondents publish their Scottish marine biodiversity data to, across all sectors (public, academic, commercial and third)
Click for a full description
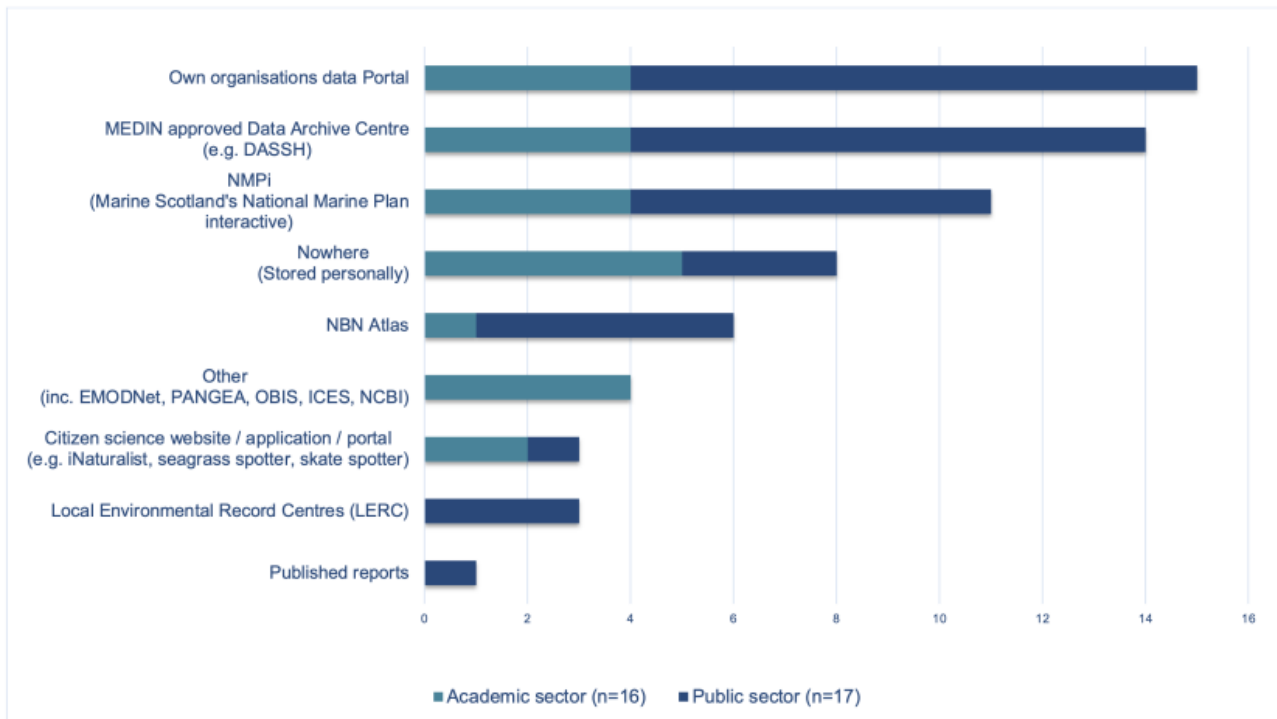
Figure 5: Portals and aggregators that academic and public sector respondents publish their marine biodiversity data to
Click for a full description

Figure 5 provides an insight into how the two sectors with the highest number of respondents, academia and public, publish their data. Being mindful of sample size, this further highlights that there is work required to facilitate mobilisation of data that is currently stored personally and not mobilised anywhere. In particular by the academic community, but also by public sector organisations whose data should be publicly available under open licence, such as the Open Government Licence (OGL) or Creative Commons (CC), in a common, machine-readable format so that anyone can access, use and share it. One of the most frequently stated barriers as a reason for lack of data sharing across all sectors was resource (staff time and funding) to correctly format and standardise the data and navigate submitting it to the correct repository for publication. The complexity of the existing MEDIN data guideline spreadsheets and the laborious process involved to submit data were often stated as a key barrier to data providers publishing (and archiving) their data through the MEDIN DAC network.

Figure 5 also shows the differing usage of portals, with more data from the academic sector being submitted directly to international portals and data aggregators, such as ICES, EMODnet (biology, seabed habitats), OBIS (EurOBIS), GBIF and NCBI, rather than into Scottish and UK repositories. There are numerous possible explanations for this, including: the Research Excellence Framework (REF) where it is deemed better to publish internationally rather than regionally; and linked to this the fact that many research projects and/or funding are often national or international in scope. International portals also offer informative signposting and documentation on how academics can publish their datasets, and subsequently cite such datasets in academic papers. If truly representative of the wider marine academic community, there is likely a variety of high-quality datasets collected for specific research projects not currently informing Scottish and/or UK marine policy.

Data access

When asked: *'how do you access marine biodiversity data?'* the majority of respondents said they accessed data from online data portals and directly from government agencies or public bodies (figure 6). The UK Marine Recorder snapshot is unsurprisingly also a key resource used by the public sector to access data and similarly it is unsurprising that the academic sector access data directly from staff / students in academic institutions. Use of EMODNet, EurOBIS and GBIF by the academic sector to search for data is perhaps reflective of the picture in figure 5 where academic institutions publish their data to international portals more than other sectors, with respondents working at an international rather than Scottish / UK scale.

When the use of online data portals is interrogated further it becomes clear that the NBN Atlas and NMPi are key resources used across all sectors (figure 7). Greater proportionate use of NatureScot's SiteLink (protected areas data) and NMPi by the commercial sector in Scotland to access data and evidence is encouraging.

Interestingly, the MEDIN Portal and data archive centres are not being used by commercial sector respondents; caution should be given to the small sample size though and therefore may not be representative of the wider commercial sector use. So little apparent effective use and engagement with MEDIN by the commercial sector for accessing data may be due to the absence of a user-focussed interface and DASSH's mapping tools having offered only limited support for exploring and interacting with the available data. Issues with the MEDIN Portal and DASSH map interfaces were highlighted as a barrier by stakeholders in Data access, collation and use barriers. However, the value chains in public marine data: A UK case study report highlighted that, where public sector marine datasets are accessed by the commercial sector (e.g., the offshore wind and oil and gas industries, marine science), via MEDIN, they are used for a wide range of different purposes, including to analyse risk, inform marine planning decisions and inform operations.
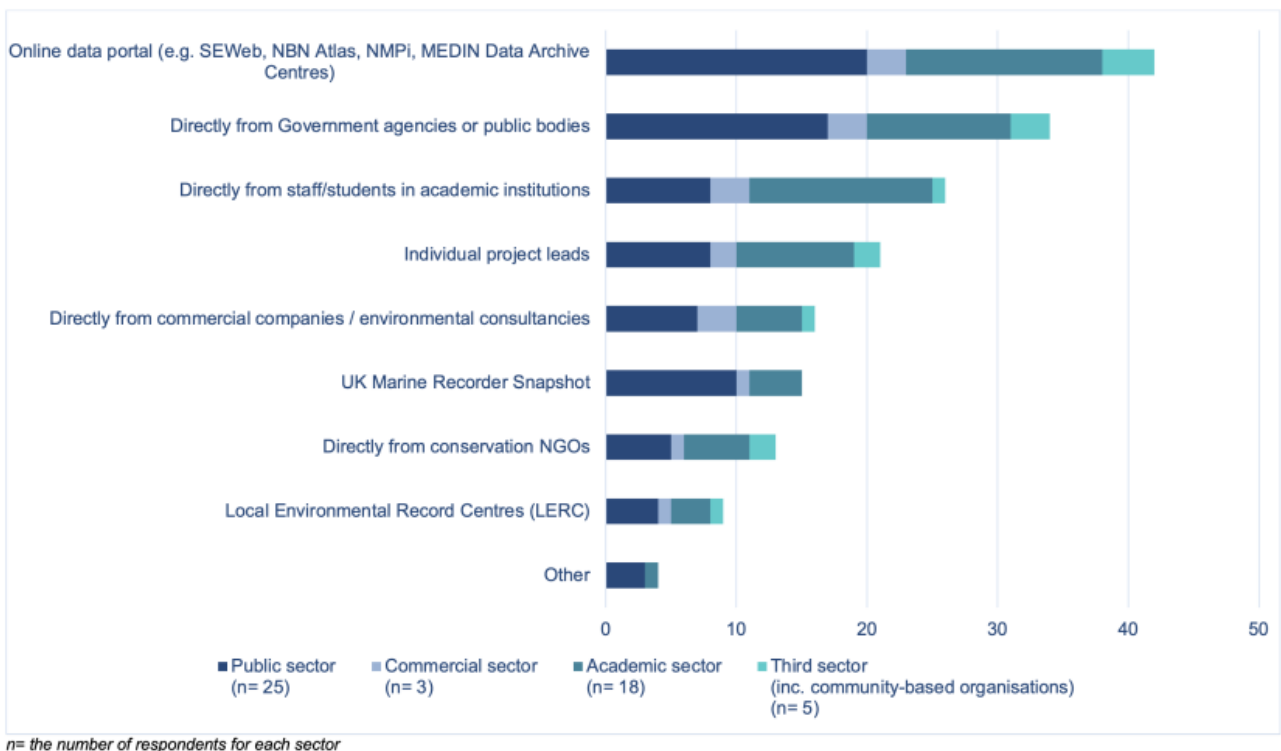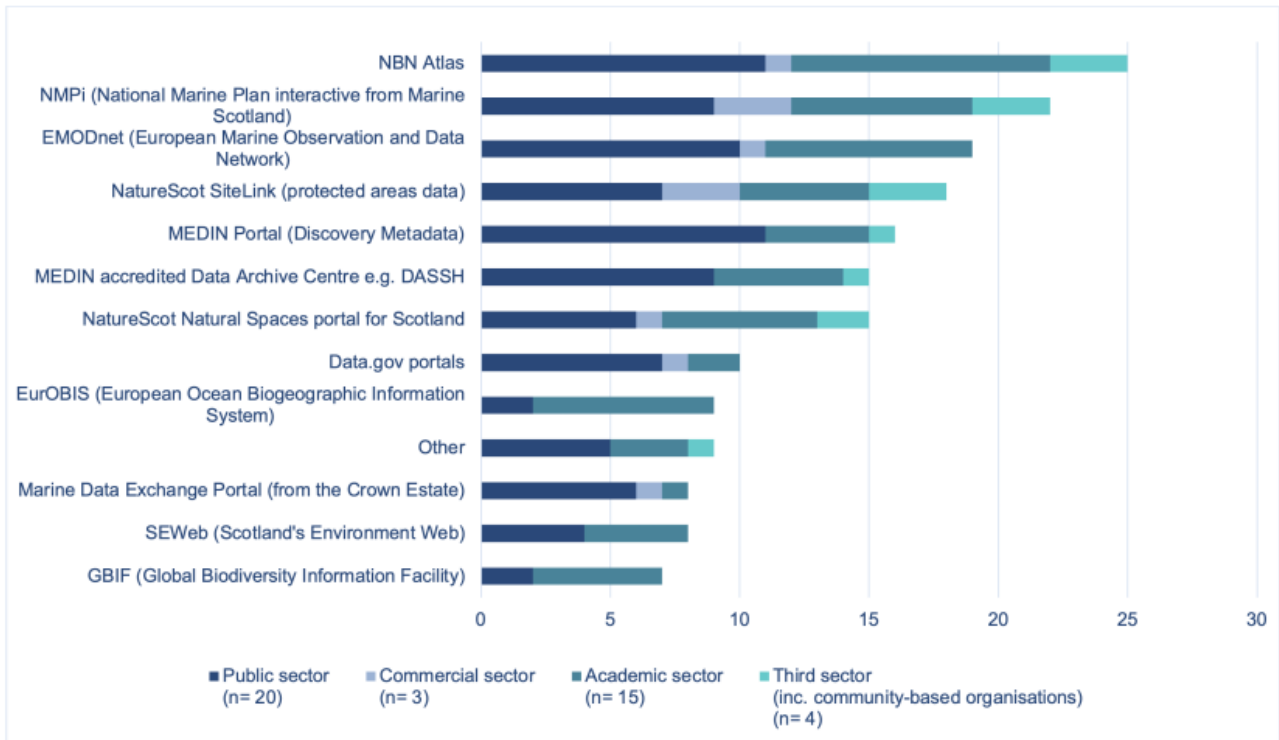

Figure 6: How respondents access marine biodiversity data by sector
Click for a full description

Figure 7: Number of respondents using online data portals to access marine biodiversity data by sector
Click for a full description

Biodiversity receptor data flow mapping

The following sections map out the key existing and developing databases, portals and aggregators that receive and publish Scottish marine species and habitats occurrence records and/or datasets. Marine data providers are faced with a range of data entry and access systems; the choice of which data flow route to use is usually governed by their purpose of data recording, their peer networks, and external stipulating factors such as research funding or planning/development conditions.

Data flow sketches map out the 'as-is' / developing landscape for the following receptors:

- Seabed species and habitats
- Cetaceans
- Seals
- Marine birds (seabirds and waterfowl)
- Fish and elasmobranchs

The data flow diagrams are deliberately presented at a coarse scale omitting funding and technical details, particularly in terms of the infrastructure workflows (i.e., web services, API) and people resource, required to successfully implement and maintain any improvements to the data flows. The data flow mapping identifies data providers, existing data management systems, portals and archive centres. The sketches are intended to help illustrate where the dysfunctional or duplicated workflows exist, missing links in the data network and where data resources are locked away and not yet mobilised for further use. The data flows only represent processed data, not raw data files (e.g., species identifications, observational count data, video and still images, acoustic).

Benthic species and habitats

The existing Marine Recorder has provided a key mechanism for storing and managing marine benthic data in the UK since the early 2000s. It has been predominantly used directly by the SNCBs and National Recording Schemes like Seasearch, although the data it contains are used by a wider range of organisations. Marine

Scotland and SEPA historically have internal data management structures that meet their own operational requirements. Where NatureScot undertake collaborative monitoring with Marine Scotland or SEPA, this data is submitted to Marine Recorder.

The re-developed Marine Recorder database, Marine Recorder Online (MRO), will deliver a modern cloud-based platform for storing and querying benthic occurrence sample data and will significantly streamline the management and mobilisation of data. The system will facilitate storage of processed spatial data from standard benthic survey methodologies (including: data derived from intertidal and subtidal photography or video; cores; grabs; intertidal and subtidal transects; and trawls). Further detail is provided in Annex E.

Benthic records and datasets also flow into the data network via alternative routes, currently by-passing Marine Recorder:

- Ad-hoc records from recreational divers and citizen science data from community groups are submitted to NatureScot and once verified, these are added into the Geodatabase of Marine features adjacent to Scotland (GeMS) collation of marine records of nature conservation importance where they qualify as records of Priority Marine Features or Annex I habitat. There is an aspiration to improve the data flow from citizen science community-led monitoring groups in particular; various avenues are currently being investigated as part of a co-design project led by NatureScot, one of which includes coordinated entry of data into Marine Recorder Online via Seasearch.
- Citizen science records submitted to Seagrass Spotter are mined on an ad-hoc basis for inclusion in GeMS as a feature of conservation importance, where the record meets the definition of a seagrass 'bed' (habitat). Qualifying records are mobilised to NMPi.
- The UHI Shetland undertake survey work on species distribution and habitat mapping with multi-beam and drop-down video as part of the Shetland Islands Regional Marine Plan (SIRMP). Some of this data has been provided to NatureScot along with community project (public records) data for incorporation into the GeMS dataset collation and mobilisation to NMPi.
- Data generated from MarClim surveys, undertaken by the Scottish Association for Marine Science (SAMS) contracted by NatureScot, Marine Scotland and Orkney Islands Council (OIC), are not currently collated into Marine Recorder but are submitted directly to DASSH and published to the NBN Atlas. There is an aspiration to have this data submitted into Marine Recorder Online in future.
- The Shetland Oil Terminal (SOTEAG) rocky shore monitoring data is published directly to DASSH in MEDIN guideline format on an annual basis. Macro-benthos data is also submitted to DASSH by a third party on a bi-annual basis. Both these datasets are under CC-BY licence.

There is an opportunity for some of these data sources to be incorporated through the streamlined workflows and improved functionality associated with Marine Recorder Online.

The existing 'as-is' / developing data flow pathways for benthic species and habitats data are mapped out in Figure 8. An options appraisal (Scottish benthic data flow options appraisal) was undertaken to explore improved data flow scenarios for benthic data, addressing the issues highlighted by the amber diagram arrows.
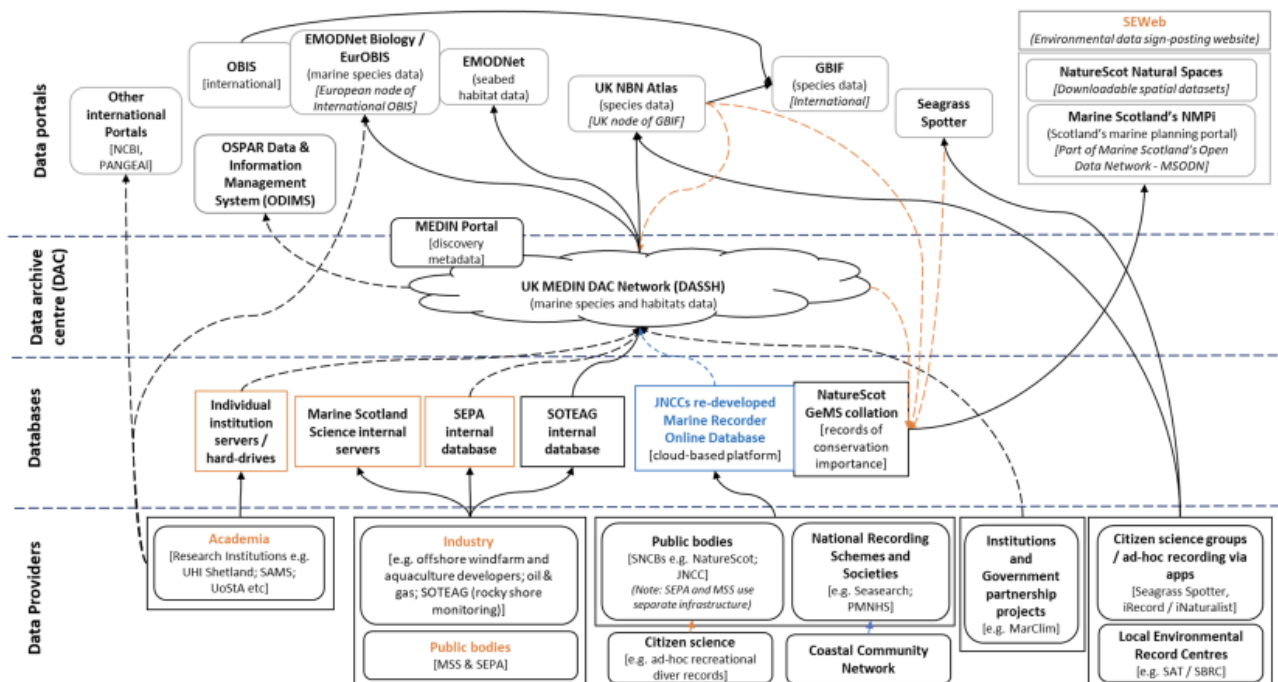
Figure 8: 'As-is' / developing Scottish seabed species and habitat data flow landscape
Click for a full description

Mammals – Cetaceans

Cetacean data has traditionally been stored and managed by the individual organisations who collect the data. This has meant a fragmented picture exists when the data is required for use in reporting, policy and management advice and in scientific research. It also means that collation of available datasets that have been collected using a variety of non-standardised formats and methods is time-consuming and difficult.

The Joint Cetacean Data Programme (JCDP) project was started in 2019 by the JNCC, in collaboration with an extensive steering group, as a 'one-stop shop' vision to house all at-sea cetacean data collected to an agreed standard in one platform for universal access. As currently scoped the JCDP will collate at-sea effort-related vessel and aerial sightings transect data, including SCANS (I-III) datasets, as well as aerial digital data. Point data and citizen science ad-hoc sightings records continue to be submitted into the SeaWatch Foundation national database.

Raw acoustic data is not compatible with the JCDP without further development (although passive acoustic monitoring (PAM) detections data is included in the JCDP from SCANS vessel surveys, along with visual sightings data), but is a potential development option once the initial phase is complete. Huge volumes (terabytes) of passive acoustic monitoring data are generated by several projects in Scotland, e.g. ECOMASS, COMPASS and MarPAMM; the raw data files are currently stored on hard-drives and individual institution's servers because there is currently no infrastructure available in Scotland / UK to collate and hold this volume of data. The Scottish Association for Marine Sciences (SAMS) is investigating the use of the acoustic recordings metadata database Tethys as a solution for storing processed acoustic detections metadata from the COMPASS project.

Cetacean photo-ID catalogues are stored and managed in various repositories within academic institutions and eNGOs. The Whale and Dolphin Conservation (WDC) manage a catalogue for Lewis Risso's dolphin sightings, Hebridean Whale and Dolphin Trust (HWDT) and SAMS manage a photo-ID catalogue for the west-coast of Scotland, Sea Mammal Research Unit (SMRU; University of St Andrews) and the University of Aberdeen (Lighthouse station) manage the East Coast bottlenose dolphin photo-ID catalogue. Photos can also be submitted though the CitizenFins project. There is certainly a degree of photo information sharing between these responsible organisations in relation to species and/or location of sighting, but the data is not currently easily accessible for public download and use.

The existing 'as-is' data flow pathways for cetacean data are mapped out in Figure 9; further text detail is provided in Annex E. Summary of data flow issues identified across receptors summarises the issues identified for mobile species (mammals and birds).

Figure 9: 'As-is' / developing Scottish cetacean data flow landscape
Click for a full description

Mammals – Seals

All statutory seal monitoring in Scotland is carried out by the SMRU, with the exception of land and vessel counts of grey seal pups in Shetland which are undertaken by NatureScot and submitted to SMRU. Seal monitoring data is only mobilised for public access to NMPi and Natural Spaces at a coarse 10km x 10km grid resolution. However, work is in progress to make gridded count data publicly discoverable and accessible through ingestion into DASSH at 5 km x 5km grid resolution. This would in future also facilitate data flow at this resolution into the wider data network (e.g., EurOBIS), however NBN Atlas do not currently support publication at this resolution.

Telemetry (movement tagging) data are also collected through various projects led by SMRU and the University of Aberdeen. The key research questions and funding has varied across the projects but much of the funding in Scotland has come from the Department for Business, Energy and Industrial Strategy (BEIS), Scottish Government and industry (e.g. renewable energy developers). Data from all projects have been combined and, in combination with haul out count data, have been used to generate estimates of at-sea distribution and density. The reports and data products associated with the maps are made available but the location is dependent on the scope of the work and the funding for each iteration (e.g., Marine Scotland (2017); University of St Andrews (2020). Photo-ID datasets of varying extents exist for both seal species (mainly focussed at a few study sites: Isle of May and North Rona for grey seals (SMRU); Loch Fleet (University of Aberdeen), Skye (SMRU), Orkney (SMRU) and Kintyre (SMRU) for harbour seals).

The existing 'as-is' data flow pathways for seal data are mapped out in Figure 10; further text detail is provided in Annex E. Summary of data flow issues identified across receptors summarises the issues identified for mobile species (mammals and birds).
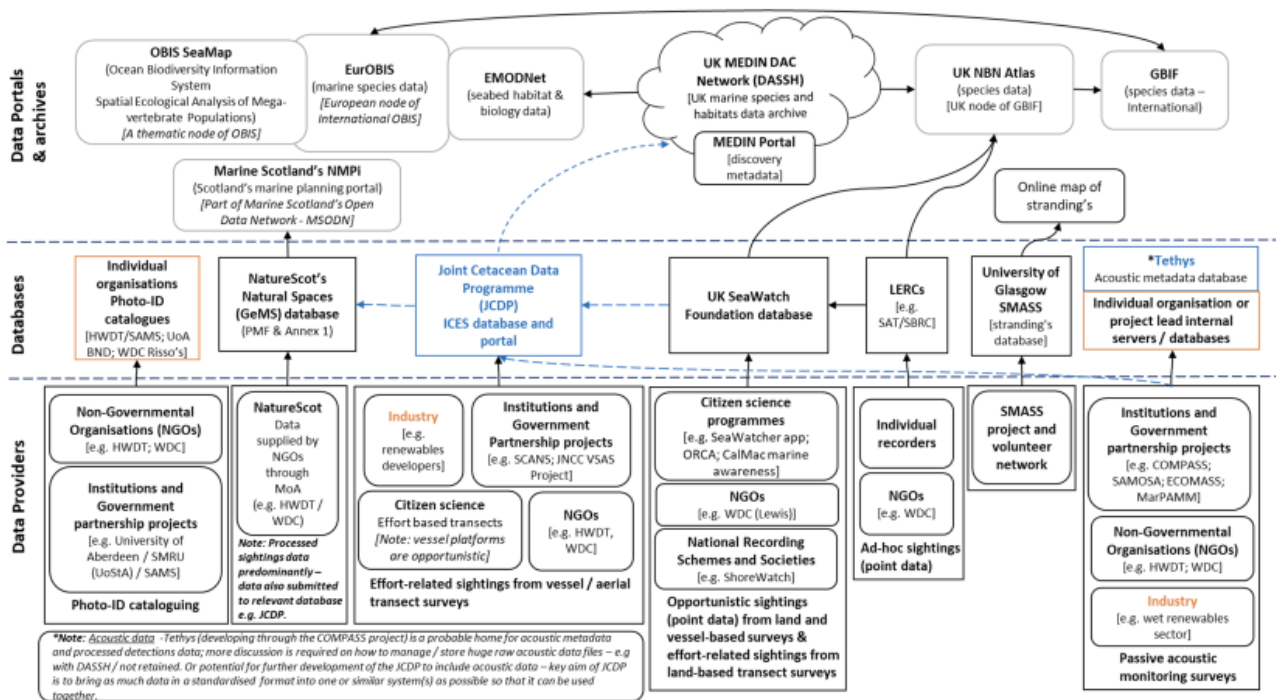
Figure 10: 'As-is' / developing Scottish seals data flow landscape
Click for a full description

Seabirds and waterfowl

The Seabird Monitoring Programme (SMP) database is currently hosted by the BTO and is developed and maintained by JNCC; from April 2022 onwards JNCC, BTO and RSPB will start a new partnership. It is an important database comprising long-term whole-colony count and breeding success data for 25 species of breeding seabirds in Britain and Ireland.

In addition to the annual monitoring scheme, JNCC leads, in association with other SMP partners, on the development and completion of periodic breeding seabird censuses across Britain and Ireland. To date, there have been four breeding seabird censuses completed: Operation Seafarer, Seabird Colony Register, Seabird 2000 and Seabirds Count.

The RSPB have their own internal database and Open Data Portal, but do contribute breeding seabird colony data to the SMP database. The RSPB collect colony data on their reserves (e.g., Shetland and Orkney) and seabird tracking data for selected species from the twin FAME/STAR projects is served up from Marine Scotland servers to Marine Scotland's NMPi portal so that is available publicly; however, this data is not routinely updated. The European Seabirds at Sea (ESAS) database, managed by ICES, holds effort related seabird at sea ship observation data collected by trained professionals and volunteers.

The BTO RAS database, which holds data on seabird survival from a mark-recapture programme, might in future be incorporated into the SMP but this is dependent on the integration of these two databases.

The BTO WeBS (the Wetland Bird Survey) database holds inshore wintering waterfowl data covering inland and coastal regions, collected by thousands of volunteers. Based on WeBS, layers for seasonal mean peak numbers of wildfowl, waders and cormorants and divers are made publicly available through Marine Scotland's NMPi portal; this data however is not regularly updated.

The existing 'as-is' data flow pathways for bird data are mapped out in Figure 11; further text detail is provided in Annex E. Summary of data flow issues identified across receptors summarises the issues identified for mobile species (mammals and birds).
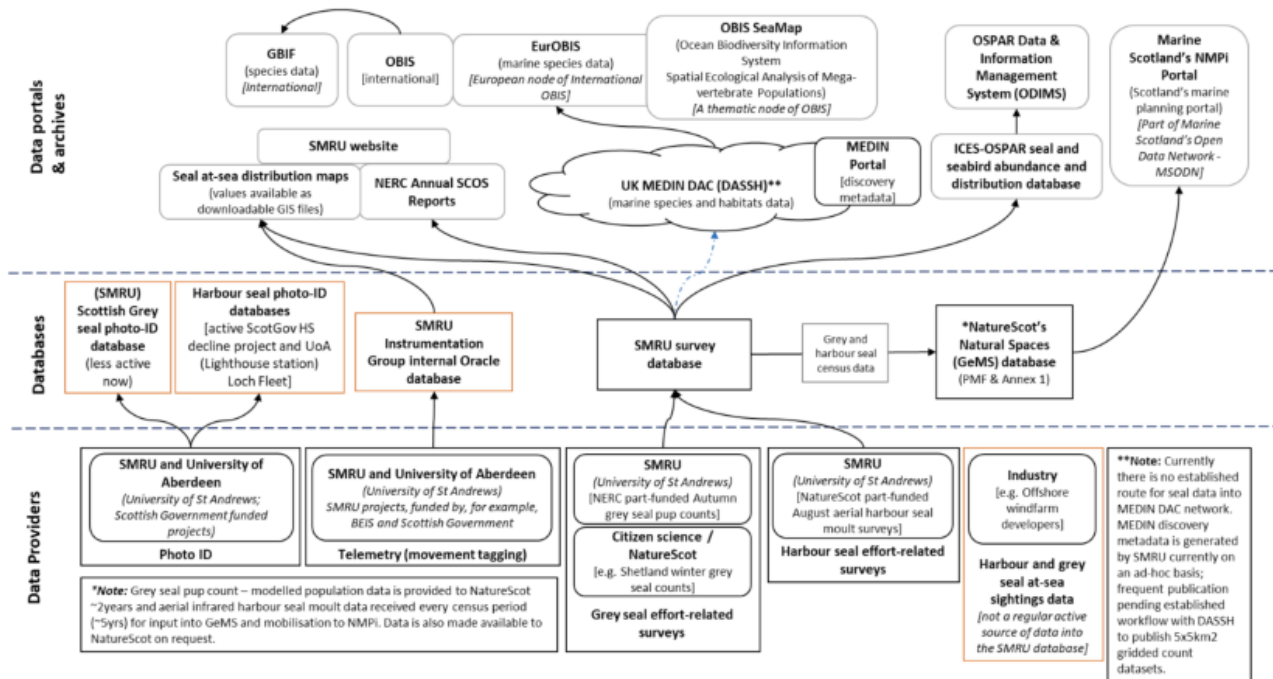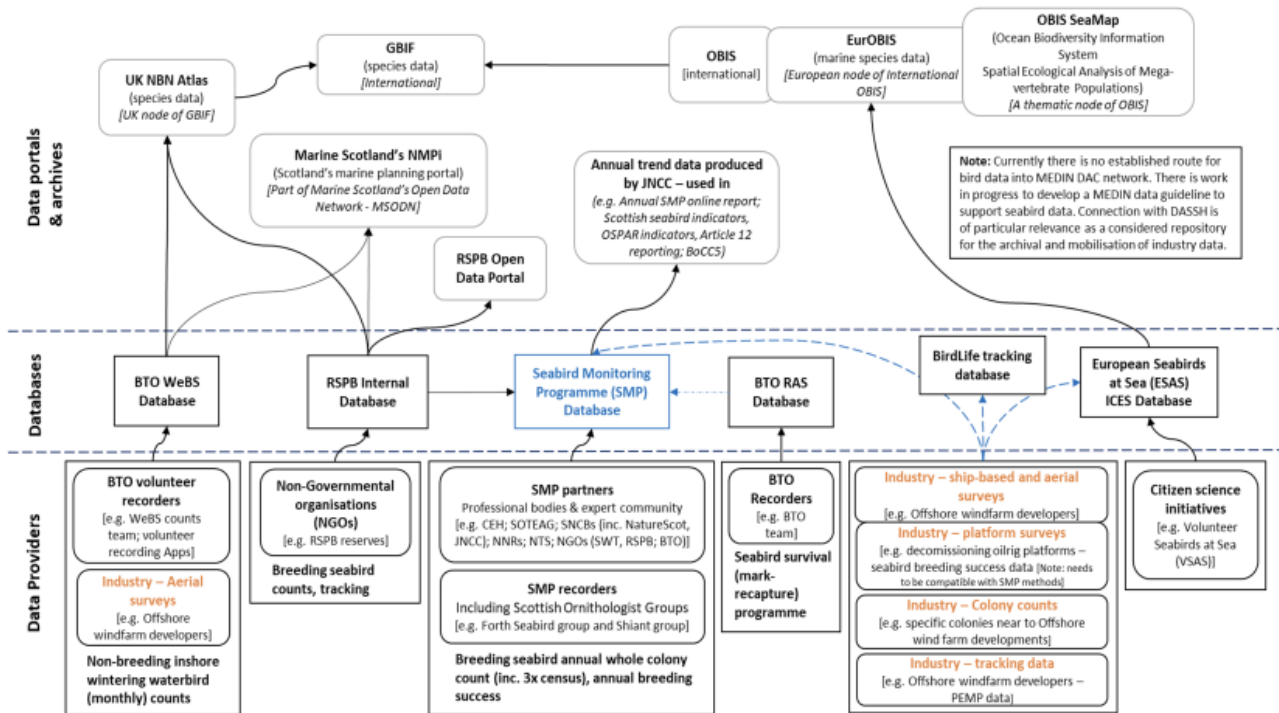
Figure 11: 'As-is' / developing Scottish seabird and waterfowl data flow landscape
Click for a full description

Fish and elasmobranchs

The ICES DATRAS trawl database is the end repository for fish data collected by Marine Scotland Science and by the UHI Shetland.

The Shark Trust database provides a repository for flapper skate eggs (live and case) data; holding data submitted by the Shark Trust itself and also by Seasearch divers and the Orkney Skate Trust. The SkateSpotter database developed and hosted by SAMS-NatureScot holds photo-ID and PITT tagging data collected by skate anglers and through research institute and government partnership projects. The Marine Conservation Society UK Basking Shark Watch database is the repository for basking shark sightings data and is contributed to and used by eNGOs including HWDT, Scottish Wildlife Trust (SWT), RSPB and the Shark Trust.

Skate acoustic tagging data collected through research institute and government partnership projects is currently stored on Marine Scotland internal servers.

The existing 'as-is' data flow pathways for fish and elasmobranch data are mapped out in Figure 12.
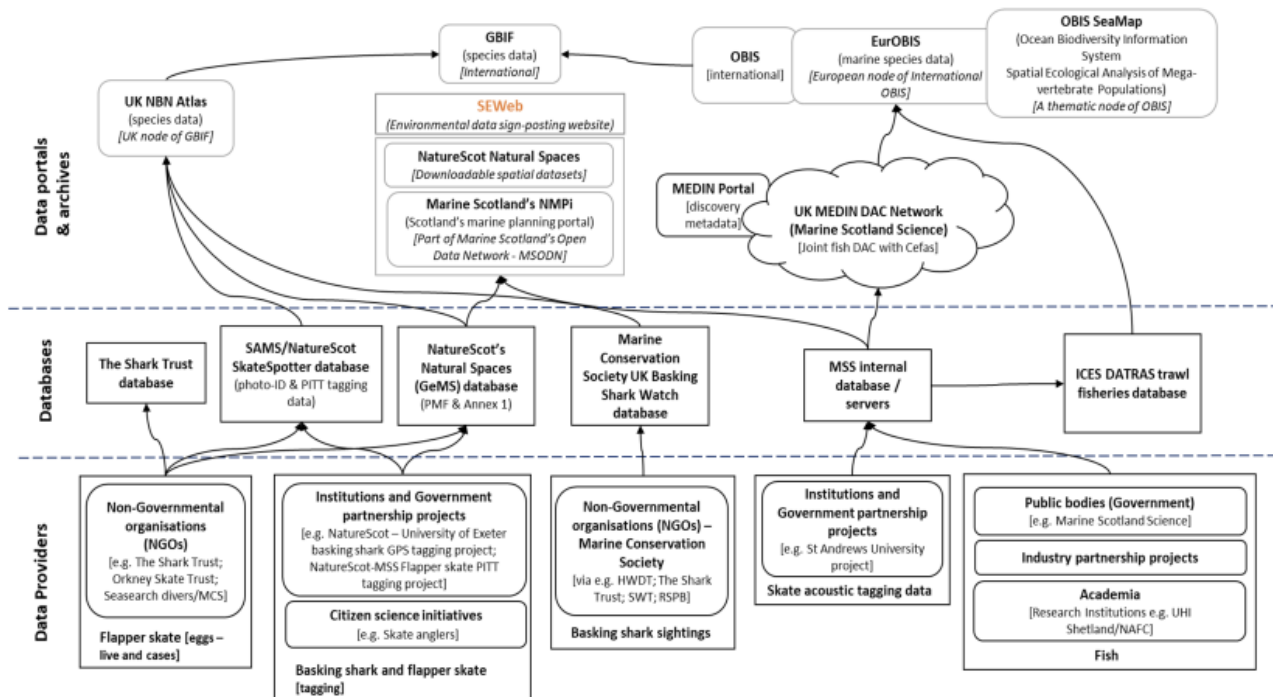
Figure 12: 'As-is' Scottish fish and elasmobranch data flow landscape
Click for a full description

Each receptor data flow has been sketched in isolation; in reality survey datasets and/or data providers may include multiple receptors. It would be beneficial to sketch out a holistic view to show interactions between databases and portals when the receptor data flows are overlaid with one another, to provide a view of the landscape that data providers are faced with when submitting data.

Recommendations 1 – 4

RECOMMENDATION 1 – Undertake a UK-wide marine biodiversity data infrastructure assessment

The UK Monitoring and Assessment Reporting Group (MARG) should expedite a UK-wide marine biodiversity data infrastructure assessment to inform development and agreement of a strategic and integrated technical road map that will simplify the data flow and connectivity of infrastructure.

The five actions below should build on the mapping analysis work done in this current Scottish marine biodiversity data review and work done by the JNCC on UK Marine Strategy indicator data flow mapping to achieve a UK-wide assessment.

1. Provide a clear definition of each of the key components and tools in the marine species and habitats data infrastructure. *This would assist in improving the automated harvesting of records and integration of standalone system data flows;*
2. Clarify the linkages between data repositories, portals, aggregators, DACs etc. *This would ensure that data providers are secure in the knowledge of where their records will be available for use and where they can aid decision-making;*
3. Endorse the key roles each portal and repository fulfils. *This would maximise inter-operability, coordination and ease and speed of data flow into and from the MEDIN DACs and from EU / international portals directly receiving data. It would also promote the endorsed portals and repositories as those supported by the wider marine community and help to encourage more extensive uptake and use;*
4. A single central directory of all Scottish / UK affiliated data submission routes should be developed and maintained by MEDIN. *This would facilitate streamlined data submission into the MEDIN data archive centre network, via affiliated routes (e.g., repositories). A link from SEWeb (or its re-developed form) could also be made to sign-post the directory in future.*

5. Map out the current and future capabilities of organisations with an interest in the data flow pathway. *This would ensure that: the organisations involved have appropriate workforce skills, resource and funding to undertake recommended improvements; the data infrastructure / data flow is joining up; and that system integration is being improved.*

RECOMMENDATION 2 – Scottish (and UK) Government recognise and resource key skills and infrastructure across the full data lifecycle

The key components of the Scottish marine data flow landscape should be recognised and resourced as:

- Core/central database management systems (e.g., Marine Recorder Online; JCDP);
- MEDIN data archive centres (e.g., DASSH);
- Scottish specific and UK-wide marine data portals (e.g., NBN Atlas, NMPi Portal).

Each component should have a clearly defined role, enabling them to work together as a collaborative, connected network.

This links to: a task under RECOMMENDATION 1 – a clear definition of each of the key components and tools in the marine species and habitats data infrastructure; and RECOMMENDATION 3 – primacy of affiliated data submission routes.

RECOMMENDATION 3 – Adopt primacy of affiliated data submission routes

Marine species and/or habitats records should be submitted into the appropriate established database where the database remit permits (e.g. benthic species and habitat occurrence data into Marine Recorder Online, cetacean at-sea effort-related vessel and aerial sightings transect data into the JCDP), as the recognised data entry point to the data flow network, and channelled to a MEDIN DAC (e.g. DASSH) via standard, affiliated workflows for onward dissemination to the NBN Atlas and other data aggregators (e.g. EurOBIS/EMODNet).

This will help avoid duplication of effort and indirect data flows. It would also reduce the complexity of collating records for individual/organisation purposes and help prevent version control and/or record duplication issues.

Links with a task under RECOMMENDATION 1 – a directory of all affiliated Scottish / UK data submission routes maintained by MEDIN.

RECOMMENDATION 4 – Map out marine data flows holistically

A holistic picture of the Scottish data flow landscape (e.g., seabed, mammal, fish and bird data) should be mapped out.

The mapping should build on the individual receptor data flows mapped for Scotland in this current analysis review. This would clearly outline the infrastructure that a data provider would be faced with when deciding where to submit their dataset into the data network to the relevant receptor database or repository.

This links with RECOMMENDATION 21 – development of guidance on optimum data submission routes.

Summary of data flow issues identified across receptors

This section summarises the issues identified through the data flow mapping exercise in Biodiversity receptor data flow mapping, specific to each biodiversity receptor. The issues identified below require further assessment and collaborative stakeholder work to resolve and make step-change improvements to the workflows for sharing, managing, aggregating and navigating data.

- Benthic: see Scottish benthic data flow options appraisal below.

- Mobile species (mammals and birds): Further discussion with stakeholders is required to agree possible solutions and infrastructures for where cetacean passive acoustic monitoring raw data files, seal telemetry data and raw data files from digital aerial surveys sit best within the existing / developing infrastructure.
    - Solutions for standardising and archiving datasets from acoustic and digital aerial surveys are under investigation through the renewables data archiving project led by Marine Scotland (Commercial data – a case study example).
    - The British Oceanographic Data Centre (BODC), has been working towards a technical solution to provide archive and access to passive acoustic data (e.g., raw recordings) but this is not yet operational.
- Mobile species (mammals and birds): Further discussion with stakeholders (in particular the Universities and eNGO's who collect a large amount of megafauna data) is required to ensure that the sharing and flow of data is improved and that data available is up-to-date.
- Photo ID: There is a need to improve the public accessibility of mammal photo identification databases. Further discussion, in particular with the Universities and eNGOs that generate and manage the catalogues, is required to agree a joined-up catalogue with accessible user interface and linked to sightings records, where this data exists.
- Dataset resolution: There is a need to improve the flexibility of grid resolution accepted by repositories and portals to support publication of data at various resolutions (e.g., 5km x 5km; 20km x 20km) and definition of grids using different projections. For example, development of the NBN Atlas to meet the grid resolution requirements for sensitive feature policy (Native Oyster) and OSPAR drivers.
- Data mobilisation: There is an opportunity and a need for Marine Scotland and NatureScot to review the technology and workflow currently used to mobilise Scottish data to Marine Scotland's NMPi portal; the WMS is not performant and the service takes a long time to draw. Consideration given to aligning the healthy and biologically biodiverse layers section to the themed structure of Scotland's Marine Assessment 2020 portal would also provide consistency. To facilitate collation of Scottish species and habitats data for mobilisation to NMPi, DASSH need to develop an API so that NatureScot can efficiently harvest 'loose' species and habitat records that are not managed through Marine Recorder into GeMS.

Scottish benthic data flow options appraisal

A workshop was held with government bodies (NatureScot, Marine Scotland, JNCC and SEPA) who are reliant on access to marine biodiversity data for their statutory and regulatory functions and the MEDIN Data Archive Centre DASSH to discuss results from the benthic data flow options appraisal. The aim of the workshop was to explore potential improvements to the existing data flows for Scottish benthic data. The decision to focus on benthic data was stimulated by the development of Marine Recorder Online and the opportunities that presents for improvement to associated data flows.

The options appraisal and workshop were important in helping to identify a preferred option and form clear recommendations for improvement to the existing infrastructure and/or identify further analysis work required to address the issues and barriers that persist in the existing benthic data flow pathway described and mapped in Benthic species and habitats.

In summary, the issues identified for benthic data flow are:

- Lack of persistent identifiers; resulting in a risk of record duplication within systems / collated datasets.
- Lack of an established and agreed mechanism for collating records submitted directly to the NBN Atlas and DASSH; resulting in data backflow and additional resource intensive non-automatic workflows.
- Lack of infrastructure to facilitate efficient viewing and download of habitat records from DASSH and the NBN Atlas.
- Dysfunctional data flows that prevent data from reaching established data repositories; data are often kept in siloes on individual organisation's servers.
- Data collected by industry and academia in Scotland is not collated and made available for wider use in Scotland (this is largely also applicable to other receptors too).

Key benthic data flow options discussed

1. Use of Marine Recorder Online as a storage and management solution for benthic data generated by SEPA and MSS (building on existing NatureScot and JNCC use); presenting a cohesive approach to marine data management for government body data in Scotland.
2. The role of Marine Recorder Online in managing, collating and mobilising benthic data; including the directive flow of benthic occurrence data through Marine Recorder Online (MRO), where appropriate, to DASSH.
3. Options for most efficiently managing and archiving benthic data coming from industry developers; to DASSH via MRO or directly to DASSH.
4. The existing role and future niche of DASSH for data archiving and dissemination. Including further development of APIs to enable harvesting and collation of both species and habitat data. Distinction of roles between DASSH and the NBN Atlas user-interfaces – online mapper and use of API by end users to access and extract data.
5. Publication of data to NBN Atlas (species and ideally habitats subject to further development) - either via DASSH publication to NBN Atlas or SNCBs responsible for publication directly from their own MRO tenancy.
6. Publication of data to NMPi and the organisation responsible for tagging of records with conservation status - highlighting a likely ongoing requirement for the GeMS collation.
7. Workflow for collation of 'loose' records from other sources (including: ad hoc citizen science records; academic data from EU / International database).

The overarching aim of exploring these options was to work towards achieving the objective of a more streamlined data flow pathway for Scottish benthic data sharing, management and archiving (figure 13). This in turn would help to remove the existing barriers to data access, collation and use identified in Data access, collation and use barriers. Specific recommendations have been made for improvement to Scottish benthic data submission, collation and publication by responsible organisations; see recommendations 5 - 11 and the proposed benthic data flow pathway (figure 14).

From a Scottish perspective, one of the greatest advantages of more benthic data flowing through Marine Recorder Online is that more data are collated and stored in one location. This greatly reduces the requirement for non-automatic workflows and data scientist resource required to collate 'loose' records separately into GeMS from various repositories (including DASSH and the NBN Atlas as is currently necessary). Instead, a larger volume of data can be efficiently managed and curated within a central database management system, Marine Recorder Online, for publication to Marine Scotland's National Marine Plan interactive (NMPi) Portal. Streamlining of data flows like this would also free up time to invest in developing new, and maintaining existing, workflows.

Figure 13 highlights the directional flow of Scottish data from submission into core databases, so that the data is collated and accessible for wider use in science, policy and conservation management advice, with affiliation to MEDIN DACs for archiving and dissemination to portals and other international data aggregators.
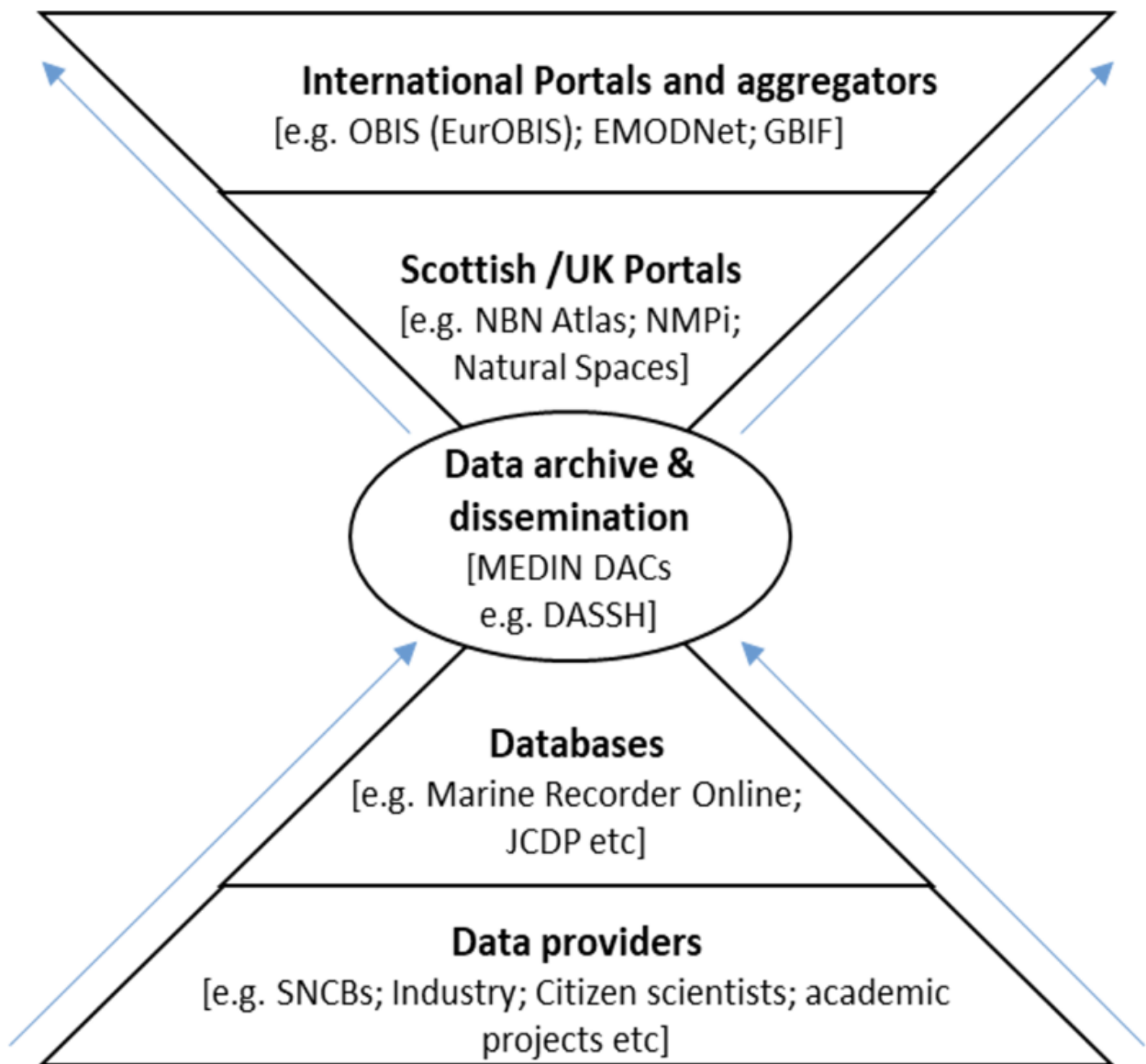
Figure 13: Idealised data flow for Scottish biodiversity data from data providers through the data network to end users in data portals and aggregators for viewing and download
Click for a full description
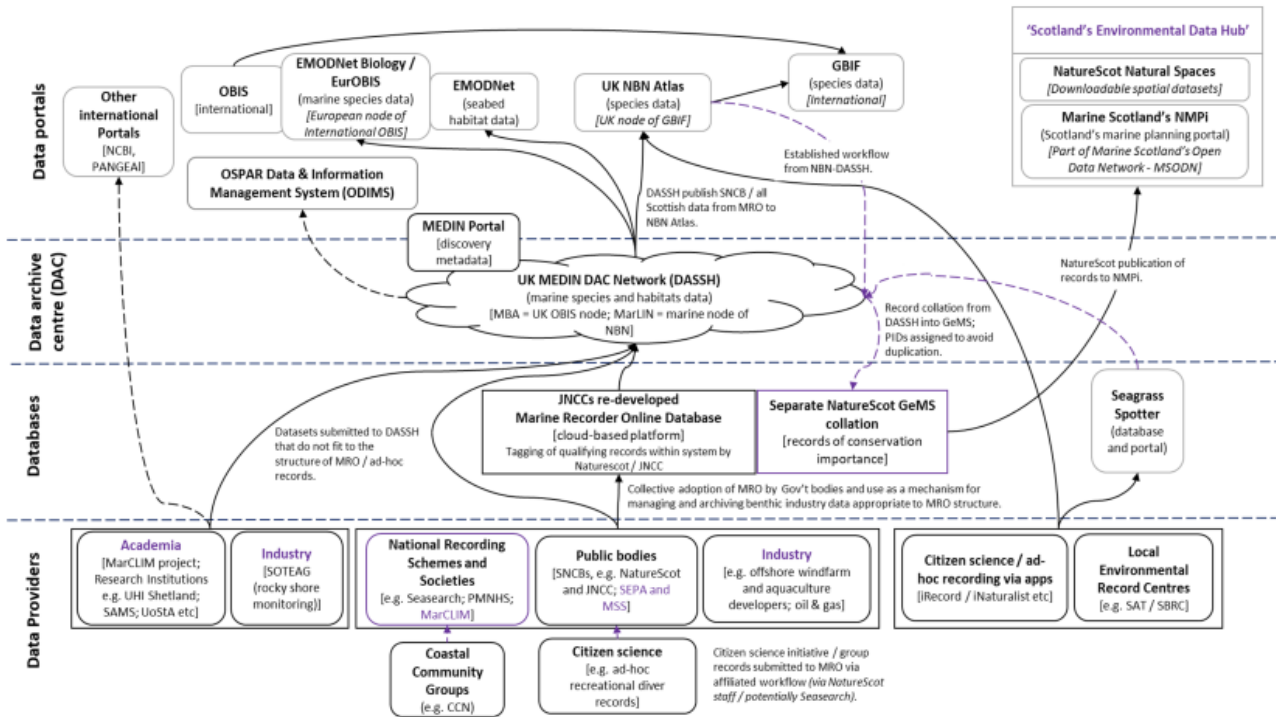
Figure 14: Proposed streamlined Scottish benthic data flow
Click for a full description

Recommendations 5 – 11

RECOMMENDATION 5 – Adopt primacy of Marine Recorder Online

Government bodies in Scotland (NatureScot, JNCC, Marine Scotland and SEPA) should adopt Marine Recorder Online, once it is available in 2022, as the data management and storage solution for benthic species and habitats data.

This links with RECOMMENDATION 3 – primacy of affiliated data submission routes.

RECOMMENDATION 6 – Clarify responsibility for tagging of records of conservation importance

Responsibility for tagging records of conservation status (Priority Marine Features and Annex 1 habitats) in Scotland should remain with JNCC / NatureScot. Marine Recorder Online should be used as the mechanism to do this for benthic data.

This would streamline the dissemination of records of conservation importance to Marine Scotland's NMPi Portal to inform marine planning and management decisions.

RECOMMENDATION 7 – Agree a single, central route for casual records

DASSH should be recognised as the single, central route for the submission of casual Scottish marine biodiversity records/datasets that are not submitted directly to NBN Atlas via apps such as iNaturalistUK and iRecord.

This links with RECOMMENDATION 9 – formalise the data flow between DASSH and the NBN Atlas.

RECOMMENDATION 8 – Each record submitted to have a persistent identifier (PID) to prevent duplication

Disciplined implementation of PIDs should be adopted (i.e., not altered or prefixed by different systems throughout its lifetime) by each data entry point system. PIDs should be allocated at every level of the survey hierarchy by data repositories at the point of data submission by recorders to prevent record duplication in data collations by enabling easy linking/identification of the same record shared to aggregators from different organisations.

The allocation of PID's (inc DOIs) would help with Findability and Interoperability in terms of dataset versioning and also contribute to data provenance to ensure it is fully traceable (Reusable).

RECOMMENDATION 9 – Formalise data flows between DASSH and the NBN Atlas

DASSH and NBN Trust should maintain the established workflow of records from DASSH to the NBN Atlas, and formalise the existing ad hoc workflow from the NBN Atlas into DASSH into an automated workflow, to create an efficient two-way exchange of records.

This would facilitate collation, mobilisation and archiving of marine species records (and habitats in due course) that are submitted directly to the NBN Atlas by recorders, e.g., via iNaturalistUK, iRecord.

This links to, and relies on successful implementation of, RECOMMENDATION 8 – persistent identifiers (PID).

RECOMMENDATION 10 – Develop infrastructure to support viewing and download of habitat records

Resource should be prioritised by (or additional provided to) MEDIN and the NBN Trust, respectively, to:

- Develop the DASSH species mapper infrastructure so that it is capable of also supporting habitats data, with the ability to access both species and habitats data through an API.
- Develop the NBN Atlas infrastructure so that it supports both species and habitats occurrence records, and ensuring that API access covers both species and habitat data.

This would help to deliver the infrastructure required to enable end users of data to efficiently navigate, browse, find and download available species and habitat data resources (complete datasets). There is a need to clearly define the niche and purpose of each system to streamline data flow and avoid duplication.

Links to RECOMMENDATION 9 – formalise the dataflow between DASSH and the NBN Atlas; and RECOMMENDATION 23 - Develop existing portal infrastructure to support efficient searching, data display and dataset collation.

RECOMMENDATION 11 – Clarify workflow responsibilities for mobilising benthic records to the NBN Atlas

DASSH should become the responsible organisation for mobilising Scottish benthic species (and habitats in due course) records to the NBN Atlas, on behalf all Marine Recorder Online custodians with records relating to Scotland's seas (e.g., NatureScot, JNCC, Seasearch); utilising the [developing] automated workflow from Marine Recorder Online to DASSH.

This arrangement should supersede the existing arrangement whereby Marine Recorder data custodians (e.g., NatureScot, JNCC, Seasearch) are responsible for publication of their own species data to the NBN Atlas, simplifying and streamlining the data workflow.

This links with: RECOMMENDATION 3 – primacy of affiliated data submission routes; and RECOMMENDATION 5 - primacy of Marine Recorder Online.

Issues and challenges identified by stakeholders

This section provides an overview of some of the main issues and challenges associated with the accessibility and availability of marine species and habitats data in Scotland, identified by stakeholders. The opinions expressed in stakeholder 1-2-1 informal discussions and through a stakeholder questionnaire carried out for this study align, at least to some extent, with conclusions of other studies, including the SBIF Review (Wilson *et al*. 2018), the Cabinet Office Geospatial Commission Review of species data flows in England (eftec, 2021) and the review carried out by JNCC of biodiversity terrestrial and marine data use in the Country Nature Conservation Bodies (Hassall *et al.,* 2020), but there are some significant differences.

Limitations to (predominantly terrestrial) biodiversity data flows identified by the other studies mentioned above include:

- **Data access**. "The lack of a stable, inclusive, central data repository".
- **Data flows**. "The lack of clear data flows and feedback, and loss of access controls" There are too many submission portals and routes, causing confusion around "who is collecting what from where". Data flows are subject to significant time lags. Complicated and incomplete data flows lead to uncertainty around the proportion of available data being accessed through one portal, and users are not sure if a gap in data coverage is a genuine absence of data.

- **Data coverage**. There are important taxonomic and spatial gaps in data, especially outside protected areas due to current funding limiting survey effort and focus; reliance on citizen science and national recording scheme data outside of protected areas.
- **Data quality**. Some potentially useful data (for example data collected through citizen science projects) need verification to improve confidence in their use in decision-making.
- **Data format and consistency**. Inconsistencies in data collection formats increase the need for pre-processing and limit the collation and application of datasets.
- **Data availability**. A low awareness of what data exists, where data are stored, any caveats around datasets and how to access limits on the datasets use.

Stakeholders in this project echoed these issues and concerns and also amplified some aspects. Some differences in opinion or experiences expressed by the marine community were expected due to the differing nature and status of developments relating to the marine biodiversity data infrastructure; an overview of the current situation and the barriers that exist for sharing, accessing and using Scottish marine biodiversity data is provided in the following sections.

Current situation

The marine biodiversity data pathways have grown organically over many decades in response to a plethora of government, eNGO and private sector influences and requirements and more recently volunteer enthusiasm for protecting their local marine environment through citizen science initiatives and community recording groups.

This diversity has produced successes, as expressed by the quantity and quality of marine biodiversity data aggregated and published through the UK MEDIN DACs, the NBN Atlas and Marine Scotland's NMPi. However, there is significant frustration at infrastructure deficiencies in terms of ease of access to and collation of the datasets that are made available and concerns for capacity to deliver escalating marine biodiversity data demands.

Availability of data

When asked: '*How satisfied are you with the biodiversity records, added-value datasets or derived data products that are available for your use through the existing marine data infrastructure?*', the level of satisfaction was highest for the range and quality of datasets available and lowest for the ease of discovery, accessibility and currency of datasets (figure 15). This picture likely reflects that respondents are aware of the range of high-quality datasets being collected through various monitoring and research programmes, but finding and gaining access to these datasets and knowing how up-to-date the data are is difficult.

One of the key challenges highlighted in Scotland's Marine Assessment 2020 (SMA2020), an assessment of the state of Scotland's seas based on the latest available published data, was availability of long-term and robust datasets for assessing status and trends. This supports the picture provided by stakeholders in this review; emphasising the strong need to '*improve the availability of data*' and also highlighting the importance of making the spatial data used to underpin reporting and policy decisions publicly available to provide transparency and engender confidence. Data sharing barriers explores some of the barriers to data sharing.

What's working well and what's working less well

Figure 16 provides a comparative measure of what is working well and less well grouped into broad themes mentioned by respondents. The respondents were asked '*What is working well and what makes the access to and collation of marine biodiversity data effective?*' and '*What is working less well and how is it problematic to you?*' to build on the quantitative picture of overall contentment presented in figure 15.

The slightly more negative picture being presented in figure 16 is likely a reflection of respondents focussing on providing more detail on the things that are working less well for them so that the issues or barriers can be improved (note that the themes mentioned in figure 16 are broader than those explored in figure 15). For example, the level of satisfaction for data quality scored relatively highly in figure 15 but was highlighted in figure 16 in relation to data quality/QA and verification/QC procedures working less well; Data verification explores some of the verification issues that persist with marine data.
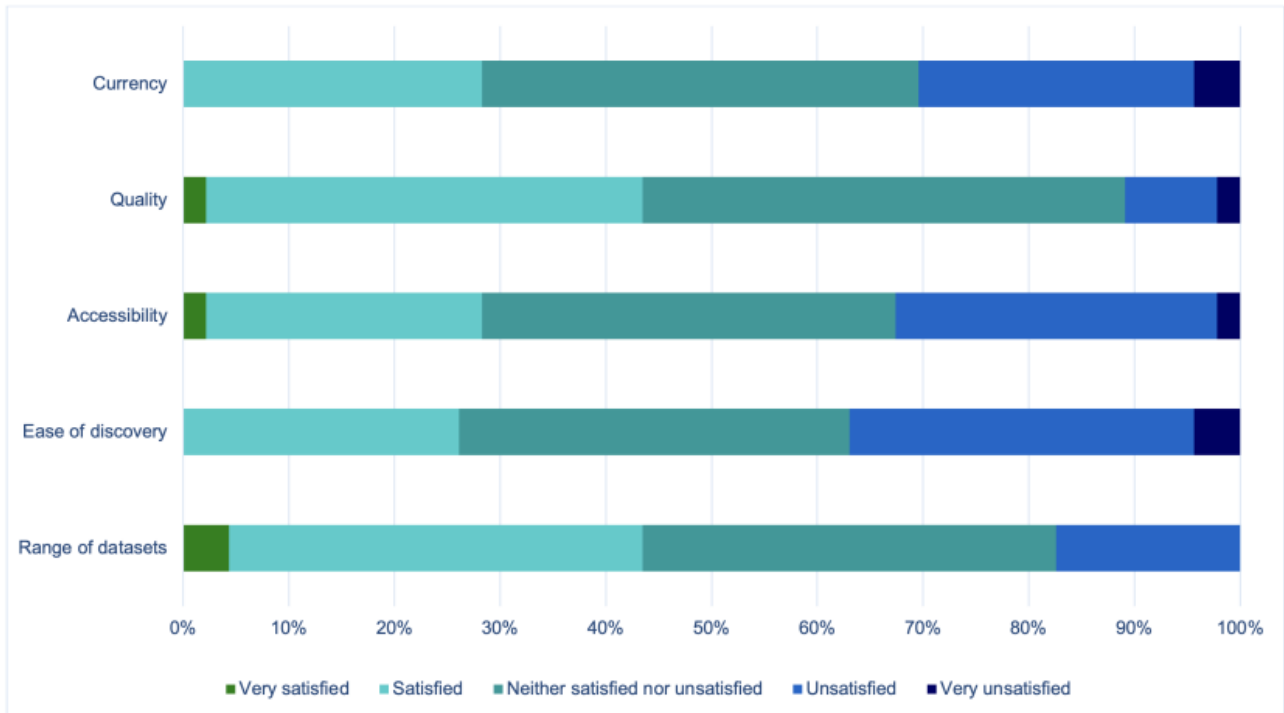
Figure 15: Respondents level of satisfaction with the currency, quality, accessibility, discovery and range available marine biodiversity datasets
Click for a full description
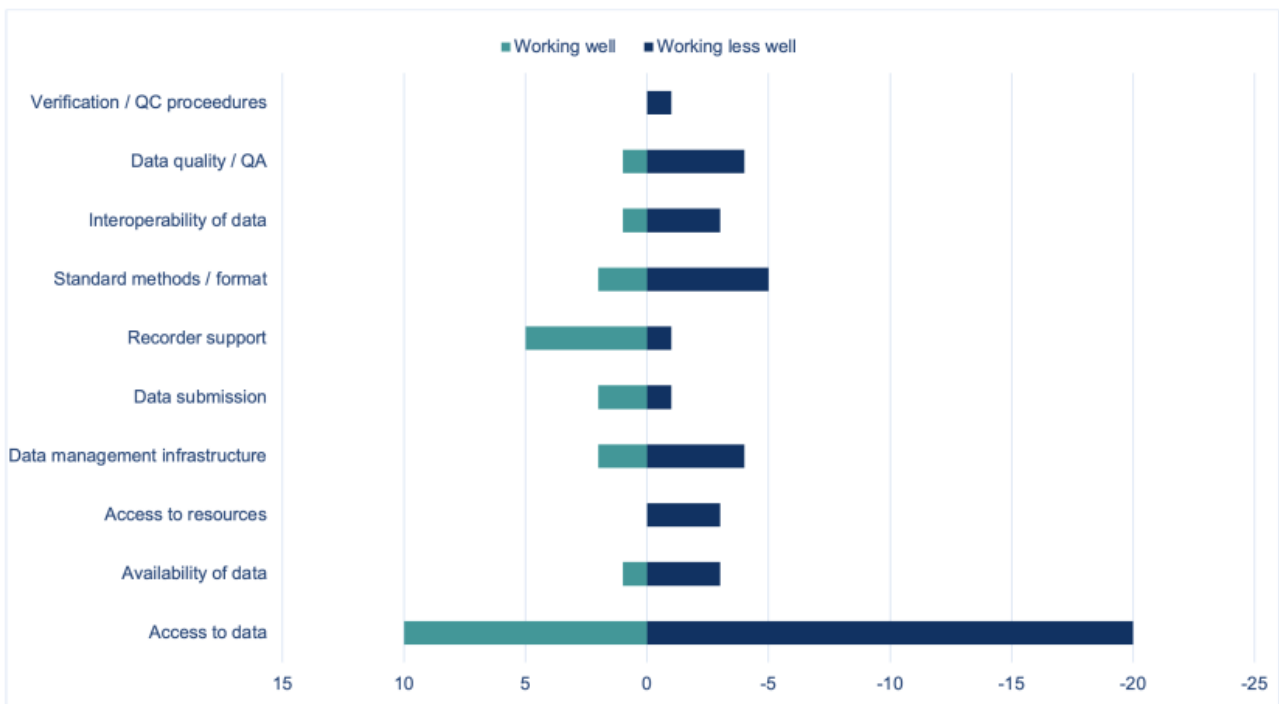
Figure 16: Number of mentions of 'what is working well' and 'less well' for stakeholders accessing, using and managing marine biodiversity data in Scotland
Click for a full description

Data verification

Verification, in combination with data validation, are important steps in the data lifecycle that increase the accuracy and overall quality of data. Marine record verification procedures and status largely do not exist for marine biodiversity data, particularly for the large volumes of data that are collected by government bodies, academia, and the commercial sector. Contrastingly, there are strict verification stages imposed for most terrestrial species data, largely as a result of the data being collected and submitted by citizen scientists

through recording apps (e.g., iRecord and other indicia-based systems). These records are sent for verification by verifiers with taxonomic expertise from recording schemes (including those affiliated with LERCs) using a 2-tier approach, or to the NBN Atlas where identification of the verification status (ranging from accepted to unconfirmed) of each occurrence record is required. Similarly, where citizen science marine records are submitted via specific recording schemes, iSpot, iRecord and Sealife Survey, they are verified by the Marine Biological Association and the Conchological Society.

The NE Atlantic Marine Biological Analytical Quality Control (NMBAQC) scheme provides a source of external quality assurance (QA) for laboratories that produce marine biological data; the processing and analysis of benthic macrofauna from grab and core sediment samples led to the initial focus of the scheme on infaunal invertebrate species. The new Marine Recorder Online system includes a field for data custodians to attribute data that has been subject to NMBAQC, for example where taxonomic determinations have been made by NMBAQC accredited laboratories; this enables data end users to filter records that have been subject to NMBAQC compliance. However, there is increasing recognition that the effective interpretation of underwater video and still image data for biodiversity is growing in importance for marine conservation and management; concerns raised in the margins of this review by statutory and regulatory bodies in Scotland relating to verification status and quality control of video derived taxa data further support this need for an equivalent published protocol for benthic imagery.

The UK Benthic Imagery Action Plan provides a strategic framework to carry out necessary improvements to a wide range of imagery analysis standards; the NMBAQC together with JNCC have developed Best Practice Guides for the operational and interpretation aspects of epibiota monitoring to help standardise epifaunal imagery data. However, verification procedure and status of imagery data is not yet widely attributed or implemented within analysis protocols.

Recommendation 12

RECOMMENDATION 12 – Progress a verification protocol for imagery derived data that complements the existing NMBAQC scheme component for grab and core sediment derived data

This supports the NMBAQC's existing commitment to develop a component of the scheme for epibiota via implementation of the UK Benthic Imagery Action Plan and JNCC's Big Picture work.

This links to an action within RECOMMENDATION 13 – develop a verification protocol for citizen science stakeholders.

Citizen science recording

The Scottish MPA Monitoring Strategy recognises the significant contribution of existing citizen science initiatives (including numerous long-term studies of marine birds and mammals) noting that '*supporting current and future citizen science programmes will be essential to maximise the information available for assessment and reporting'* (Marine Scotland, 2017).

Funding and support for community-based recording initiatives has massively increased, and this has naturally increased the volume of data recording undertaken by the third sector. This has implications for the infrastructure required to successfully store and manage the data to ensure that it can be made available for use; efficient verification and data processing were identified as two of the biggest barriers through this review to the efficient submission and publication of volunteer collected biodiversity data.

NatureScot have worked with communities and groups, to develop Scotland's first "how to" guide for community-led marine biodiversity survey and monitoring. The handbook includes an introduction to marine surveying in Scotland, including survey methods and survey data forms to record marine life. However more support to establish efficient data workflows and data management procedures to harness this data and make it available for collation and use with other marine datasets is still required.

Recommendation 13

RECOMMENDATION 13 – Provide infrastructure and management support for citizen science marine biodiversity recording

A targeted piece of analysis should be undertaken to fully understand the priorities for investment and/or infrastructure necessary to better support the flow of citizen science data into the marine evidence base.

This should include:

1. Assessing the need for provision and update of data management protocols, technical guidance and clear sign-posting of data submission routes available to citizen scientists.
2. Identifying the key data 'types', species groups, methods of data capture (apps, web forms etc), and spatial data visualisation tools to inform the priorities for investment.
3. Development of a verification protocol(s) with key citizen science stakeholders in the verification process, which aligns to current and future verification requirements and technologies. The protocol(s) need to cover the broad range of citizen science data collection methods and expertise. The resources required to support implementation of the protocol and capacity building should also be identified.

This recommendation provides the opportunity to explore funding options, including via the Scottish Marine Environmental Enhancement Fund (SMEEF). The recommendation also has synergy with the SBIF Better Biodiversity Data (BBD) Project* proposed to improve the management and long-term sustainability of LERC citizen science data.

*Currently terrestrial and freshwater species focussed.

This links to: RECOMMENDATION 12 – Progress a NMBAQC scheme component for imagery derived data; RECOMMENDATION 14 – Simplify the requirements for submitting data into DASSH; RECOMMENDATION 17 – develop simplified user interfaces onto repositories; and RECOMMENDATION 18 – guidance development on optimal data submission pathways.

Data sharing barriers

Technical and cultural barriers to data sharing in particular impact on the availability, quality, and accessibility of data for use by others. The following section describes some of the key barriers associated with data sharing that exist and/or persist, contributing to some of the challenges that are experienced by end users with marine data access and use outlined in Data access, collation and use barriers.

The review found that relatively large quantities of data were still stored locally and not fully incorporated into the data flow network. The proliferation of online tools has led to some confusion amongst the recording community on where data should be submitted and apprehensions about the push for records to be published as open data, meaning that the volume and integrity of data that are easily accessible and available for collation and use are often compromised.

Although not well represented as sector respondents to the questionnaire, the lack of accessibility to industry data and the availability of eNGO collected biodiversity data at fine-scale spatial resolution was regularly cited within the stakeholder responses as being a key issue. For example, Environmental Impact Assessment (EIA) and post-consent monitoring datasets are a potentially valuable resource for multiple policy applications. There is an appetite to mobilise these data, but this would involve a change in policy and/or legislation to establish data sharing as a pre-requisite for marine licences and consents, and overcoming the encountered and perceived barriers to data sharing (Commercial data – a case study example); there is a potential role for Scotland's new Natural Environment Bill (2023/2024) in helping to achieve this.

The key barriers to sharing marine species and habitats data identified in this review relate to:

Cultural and behavioural barriers, which range from reticence to share data for commercial reasons to barriers resulting from concerns over how the data might be used or misused / misinterpreted; some data providers requiring embargoes on releasing data, concerns about not receiving proper credit, and lack of incentives for sharing data.

Practical barriers, which range from a lack of data integration (linked to the adoption of standards and consistent data formats) which can cause bottlenecks in the process to not understanding how to make data available in meaningful ways, due to:

- Lack of technical expertise and/or capacity to deliver;
- Lack of understanding of the existing infrastructure and processes;
- Insufficient knowledge of how to adopt standardised classification systems and/or produce sufficient metadata linked to datasets;
- The variety of metadata and data standards, vocabularies, and ontologies that exist for marine biodiversity data, means that choosing the one(s) that best fits the data, methodology, and data management goals can be a time-consuming process.

Inadequate strategies and resources that result in data often being made available in an opportunistic manner rather than being focused on need and frequently without sufficient resources:

- Underestimation of the costs of managing data, particularly maintenance of 'live' information, can lead to abandonment or slow decay of data collected through discrete projects.
- Lack of time and resources to devote to learning new standards and performing data transformation is also a recurring and long-standing issue.

Recommendations 14 – 19

RECOMMENDATION 14 – Simplify the requirements for submitting data into DASSH whilst maintaining data quality

DASSH should simplify the existing requirements that need to be met by recorders in order to submit their datasets, whilst maintaining the quality of data submitted. This should be through provision of support to users of the formal data guidelines to translate and produce practical step-by-step guidance for their peers to facilitate submission of new data.

This will help:

- *Maintain data quality by ensuring that MEDIN requirements are met, but make it easier for users to share their data;*
- *Encourage more organisations and individuals to submit their data currently stored on publicly inaccessible hard-drives or servers into the data network;*
- *Increase the volumes of data made available in standard formats;*
- *Wider application of FAIR data principles and facilitate the integration potential of marine data from different disciplines and sectors (including private sector).*

RECOMMENDATION 15 – Plan for and fund the management and sharing of all new data being collected

Funding providers should stipulate that a requirement of funding will be the development and execution of a data management plan that assures datasets are provided in accordance with FAIR data principles and shared within a timely manner, following embargo periods [e.g., a requirement for research projects receiving public funds to share data that they generate with MEDIN, via affiliated data flows, to contribute to the Scottish / UK marine evidence base].

Further collaboration and discussion with the organisations that fund the collection of data will be fundamental to achieving this.

This links with: RECOMMENDATION 3 – affiliated data submission routes; and RECOMMENDATION 18 – developing guidance on optimal data submission pathways.

RECOMMENDATION 16 – Develop proactive engagement with data custodian stakeholders who weren't fully involved in the review

Further targeted engagement should be undertaken with stakeholders (including eNGOs, commercial sector) in an endeavour to increase the flow of biodiversity records into the marine data infrastructure.

This further engagement would support and facilitate the wider cultural step-change required to increase data sharing and availability.

RECOMMENDATION 17 – Develop simplified user interfaces onto repositories to support wider data submission

The development of simplified user interfaces onto repositories should be encouraged to support the submission of data by citizen science initiatives.

This links with RECOMMENDATION 2 – simplifying the requirements for submitting data into DASSH

RECOMMENDATION 18 – Develop guidance on optimal data submission pathways

Guidance should be developed with stakeholders to clarify the optimal pathways for submitting biodiversity records into Scottish / UK marine data repositories, in accordance with FAIR data principles.

For example, developing guidance with academic researchers would aim to provide reassurance to the academic sector that their data submitted at a Scottish / UK level would flow to the appropriate EU / international portals (with likely timelines); data flow into UK infrastructure in the first instance would make research data more readily available for use in a Scottish/UK policy context.

This links to RECOMMENDATION 13 – infrastructure to support citizen science record submission.

RECOMMENDATION 19 – Invest in data engineers and allocate resource for system decommissioning

Data scientists should be funded to input 'loose' data stored in file storage on networked drives into systems; complimented by short-term resource (monetary and/or effort) made available to government organisations to enable legacy data management system decommissioning so that and the benefits of cloud-based system technology can be fully adopted.

This would reduce technical debt and longer-term data management staff resource requirements associated with non-automated workflows and duplication of effort, by facilitating full adoption of automated workflows and the benefits of new cloud-based system technology.

Commercial data – a case study example

Offshore windfarm developers are required to collect marine environmental data to inform a Marine Licence application in order to assess the potential impacts of the proposed licenced activity on the environment. Developers are subsequently required to provide the survey data to Scottish Government (Marine Scotland) as part of their licence conditions and/or to Crown Estate Scotland (CES) as part of the lease agreement. However, data from Scottish offshore developments are not routinely mobilised by industry for public access or archived in DACs. There is little to no surveillance on what data is made available by developers - much of the data currently requested is held internally by Marine Scotland (figure 8) and has not been quality checked or formatted for sending to suitable DACs – therefore the data is currently not FAIR.

This routinely collected biodiversity data is needed for sustainable management of marine ecosystems, including cumulative impact studies on species or habitats, marine spatial planning and scientific research. Therefore, these data should be mobilised and archived by industry in such a way as to make them accessible and suitable for future re-use.

A cross-sector data strategy for offshore energy is also being created – the Offshore Energy Digital and Data Strategy – through taskforce engagement with participants from across the sector (including CES and RenewableUK) to identify common digital and data challenges; particularly where data sharing and collaboration efforts could be further developed. The principal aim is to maximise the value of this industry data by creating a framework for improving data management and coordination; making data available, visible and mobile.

Discussions with science and policy staff in Marine Scotland and Defra suggest that marine industry sectors on the whole are positive about sharing and allowing the reuse of marine environmental data which they collect. However, there are barriers that still need to be overcome. The primary barriers (encountered and perceived) to data sharing in open access repositories by industry (Murray *et al*, 2018; ABPmer, 2015) are:

- Commercial confidentiality (of certain data sets);
- Reuse of data (and concerns of how it will be used);
- Format of data provision (the time and cost implications of this);
- Industry motivation to supply data (as they cannot see any direct benefits).

Concerns over data ownership, accreditations, confidentiality and whether there is potential liability where 3rd parties have used the data are largely resolvable and mechanisms to handle these issues already exist, e.g., anonymisation and embargo methods. Serious reconsideration of what biodiversity data really is commercially sensitive and requires embargoed status rather than immediate publication needs further thought. Concerns over the misinterpretation or misuse of data released remain more difficult to address. For example, secondary data analysis needs to be interpreted with the original survey aims in mind, but data owners cannot control or influence interpretation once the data have been made open access. This is similarly a barrier to data sharing by the eNGO sector.

The ABPmer (2015) study also concluded that industries are only weakly motivated to supply data, as they do not perceive any direct benefits of doing so. Cited in both the above-mentioned studies, a primary barrier to data sharing by industry is the accessibility ("open access") of other datasets to them. Use of third-party biodiversity data is often free of charge, although under licence, to non-profit users (e.g., academics, government agencies and eNGOs), but is unavailable for commercial use (generation of income or profit), dictated by the CC-BY-NC licence terms. If industry is expected to submit data free of charge and cannot benefit from the data contributed by other sectors, it may be difficult to incentivise data sharing. However, there is a need for change in culture and perceptions; FAIR data sharing ought to become an accepted 'cost' associated with deriving commercial gain from the marine environment. Improved knowledge and understanding of the benefits of making marine biodiversity datasets FAIR and the socio-economic value gained when shared, is needed.

It was also identified in these two reviews that industry often lacks knowledge about the data platform infrastructures; and the time and cost required to adapt data formats to make them useful to other users is seen as a key barrier, as the marine industry operates to standards that are different from those used by government data systems. There is an opportunity, with the announcement of ScotWind round 2 offshore windfarm development leases, to establish best practices and to develop tools and protocols to facilitate standardised, low-cost data collection and management to provide consistent, standardised datasets endorsed by operators, scientists and regulators.

Marine Scotland have contracted a 'Developer Data Archiving' project that will develop and produce guidance for offshore wind and marine renewables developers in order to instruct developers where to mobilise and archive their raw and processed environmental monitoring data within appropriate database repositories and MEDIN archive centres (e.g., DASSH). The aim is that monitoring data from ScotWind developments (and other developers) will be publicly available to inform future marine ecosystem assessments, planning and policy development. The following are being explored for the biodiversity data elements of developer surveys:

1. Submission direct to DASSH (the MEDIN DAC) in MEDIN data guideline format;
2. Submission to database repositories (receptor specific) and DASSH harvest data automatically from these databases for publication to UK and international portals and other data aggregators;
3. Production of data standards for data collection methods where these are lacking, namely marine bird and cetacean digital aerial survey data and passive acoustic monitoring data.

Recommendation 20

RECOMMENDATION 20 – Provision of biodiversity records collected under licence or for consent into the MEDIN data archive centre network

It should be a statutory requirement for records collected by commercial developers through the licensing and consenting system to be provided, via affiliated data submission routes (i.e. established databases), into the UK MEDIN archive DASSH. This would enable onward publication to the NBN Atlas and other international portals.

This links to RECOMMENDATION 3 – primacy of affiliated data submission routes.

Data access, collation and use barriers

While there have been efforts to make marine data freely available, via portals and other mechanisms, there are still substantial barriers. Most data collection, by both private and public organisations in Scotland working across a range of disciplines, is carried out for a single, specific purpose and using heterogeneous survey

methods, often in isolation from each other. The result is that much of the data are scattered throughout unconnected databases and repositories. Even when data are shared, they are often not compatible, making the collation of available data particularly challenging.

The key barriers to efficiently access, collate and use marine species and habitats data identified in this review relate to four key themes: "data infrastructure, technology and skills"; "data availability"; "data accessibility"; and "data quality". Selected stakeholder responses are quoted under each theme below to illustrate the issues experienced:

Data infrastructure, technology and skills: Emerging technology can play a pivotal role in tackling many of the critical issues facing management and accessibility of Scottish marine data, but maximising the impact of these technologies requires addressing several significant barriers. These barriers include lack of awareness of technologies and tools, prohibitive cost, degree of adoption and transferability across systems and/or scales, and lack of technical expertise.

- *"Infrastructure exists but it is complex and needs to be better organised"* - the technologies that underpin automated data dissemination and collation currently exist across a range of maturity levels. In many cases, data are often still forced into traditional relational database systems, or worse, into unstructured files — even though that is not the optimal place for data use and analysis. This makes using this data difficult and less effective. Investment in the skills required for use of cloud platforms configured into a modern data architecture makes data more readily available and more cost effective. A key advance is the development of Application Programming Interfaces (APIs). However, the development and use of APIs is not yet widespread across portal and aggregator interfaces, and where they have been developed the skills capacity for utilising APIs varies across end user groups.
- *"There are so many different online portals that it's difficult to know where to go to and know when new data has been published"* - there is confusion among end users as to where data ends up; it is difficult to know which datasets feed into which portals. This is largely because submission of data to one database or repository does not automatically guarantee data is made available on Scottish and UK portals or data aggregators due to lack of connected and automated data flows between systems along the data pathway.

Data availability: The apparent gaps in data availability identified are largely a result of either data not being properly shared or organised (Data sharing barriers) or due to resource driven workflow time-lags; figure 4 (Data publication) highlighted large quantities of data being stored locally and not fully incorporated into the data flow network.

- *"Limited capacity to process data and submit to make available. Often not seen as a priority"*
- *"Lack of sufficient technological knowledge, staff resource and/or expertise restricts the ability to engage sufficiently in data management (particularly in short-term or species-specific projects)"*
- *"Some datasets are made available only at reduced resolution, which reduces their utility when trying to compare them with other datasets"*

Data accessibility: There is a lack of knowledge about where data exist and how to access data that is known to exist. Data discovery is difficult when data sources are unknown, metadata quality is poor, or when there are data silos and compliance restrictions. Duplicated data, lengthy and/or not well-defined approval processes for accessing data, and a general lack of understanding around what data is available also results in negative end user experience when trying to access data.

- *"It's difficult to find data – there's no easy way to collate data from multiple systems and know whether relevant datasets have been missed"*
- *"Getting definitive and up to date information is difficult; data is stored in lots of different places leading to a fragmented picture (i.e., it's difficult to know which one portal, aggregator or database is the best to use to gain maximum access to the data available)"*
- *"Portals don't contain enough information about what the data mean and it can be difficult to access and understand files; difficult to view and browse records quickly"* –lack of or poor metadata is core to this issue; when data are available as machine readable and fully-described, it removes the need for end users to download data from a portal or repository to know what data exists and where it resides.

- *"Access to industry data can be difficult"* – largely attributable to the barriers discussed in <u>Commercial data – a case study example</u>.

Data quality: Quality data typically results from the application of community best practices and adherence to standards across the data lifecycle.

- *"Data duplication across systems"; "Unclear how duplicates are dealt with by data aggregators"* - ambiguity caused by multiple repositories and third-party hosts having different versions of data is an issue. If the data are to be used in decision making then users need to be sure they have the definitive version. When copies of data are re-exposed to the web via third parties there is a long-term overhead in ensuring that the most pertinent version of data is maintained.
- *"Reliability and accuracy of records vary greatly"; "Difficult to know who to contact about data issues"* – the usability of the data is compromised without well-structured metadata and descriptions of provenance. Ensuring the integrity of the data (i.e., avoiding data corruption) is especially important for data that are to be stored in perpetuity and intended for future reuse.

Recommendations 21 – 23

RECOMMENDATION 21 – Maintain data version control through encouraging active custodianship

Data custodians should perform checks to determine whether the version of data in portals is true to source, and ensure that portals harvest updated data, in addition to re-archiving.

Re-harvesting of data by Data Archive Centres (e.g., DASSH from MRO), either periodically or on request following active management of data, would help ensure that data are up-to-date and robust throughout the data network.

RECOMMENDATION 22 – Optimise re-use of data through adherence with FAIR Data Principles

All organisations should champion open data and FAIR data principles: use of Open Government Licensing for all data commissioned by public bodies and Creative Commons (by attribution) (CC-BY) licences for industry, eNGO volunteer recording and academia should be encouraged; clear licensing conditions; and easy to access descriptions of the dataset (metadata).

The generation, management, collation and sharing of data should be based on FAIR Data Principles to make marine species and habitat data in Scotland Findable, Accessible, Interoperable and Reusable (FAIR) throughout the data flow network.

RECOMMENDATION 23 – Develop existing portal infrastructure to support efficient searching, data display and dataset collation

DASSH, Marine Scotland and the NBN Trust should prioritise investigation and requirements gathering to fully understand DASSH's species mapper, NMPi portal and the NBN Atlas's existing and future customer needs against the current and planned work to respective portal interfaces; i.e., what stakeholders need access to and how, and where the highest value lies for each customer*.

*There are a wide range of user groups with different needs / expectations.

This could include:

- A discovery phase, prior to undertaking user needs research, to clearly understand and articulate each system's niche and its purpose within the data network to avoid duplication. See <u>Annex D</u> for description of existing key system purposes relevant to Scottish data.
- Understanding what is working well and what can be improved in the existing user interfaces of these platforms, the mapping tools functionality, download services and use of APIs. This would help to deliver the infrastructure required to facilitate efficient searching, harvesting and collation of records. This links to <u>RECOMMENDATION 10</u> – infrastructure to support habitat records.
- Enabling the querying, visualisation and download of multi-disciplinary datasets for use in end-user systems, via cross-DAC re-aggregation of data.

Summary of stakeholder needs and requirements

In response to being asked '*What ideas do you have for specific or general improvements that could help resolve any issues that you have with the data that are available?*' 51 questionnaire respondents made 118 suggestions. 1-2-1 discussions with stakeholders (data providers, data managers and data end users) throughout the data flow mapping exercise (The existing Scottish biodiversity data landscape) also contributed to the formation of these needs and requirements.

The needs and requirements are split out into functional and non-functional, under 4 headings: infrastructure, technology and skills; data availability; data accessibility; and data quality. There is no particular order of the needs and requirements identified across sectors and biodiversity receptors listed under each heading.

Functional requirements

Functional: defines what the system should do; usually defined as a process (i.e., system features and user requirements).

Infrastructure, technology and skills

- *Data managers and end users require* a persistent identifier (PID) which stays with the data throughout its lifetime so that duplicate records can be easily identified in collations of data from multiple repositories.
- *Data providers require* simple and efficient workflows and tools for publishing data.
- *Data providers, managers and end users require* that databases and portals are unified and/or rationalised, with greater clarity on how different portals and databases feed into and relate to each other.
- *Data providers require* that data flows are streamlined so that data only needs to be submitted into the network once (i.e., into established databases) and gets distributed outwards to portals, data aggregators and archives; removing duplicated effort of submitting to multiple repositories.
- *Data end users require* a user-friendly (intuitive structure) online portal enabling access to all databases (i.e., 'one-stop-shop' to go to search, browse and find data for downloading) that's kept up to date with a 'live' feed showing what and when data has been added.
- *Data managers require* a defined and efficient workflow for coordinating and mobilising data collected through citizen science initiatives into the wider data network.
- *Data managers and end users require* that systems (repositories, databases etc) are linked to provide a holistic view of datasets available with clear links retained between processed data and raw data stored separately.
- *Data end users require* efficient querying tools in terms of time and output.
- *Data managers and end users require* an established and agreed workflow and mechanism for collating records submitted directly to the NBN Atlas and DASSH.
- *Data providers require* an efficient and simple interface or tools for submitting datasets to databases and/or MEDIN data archive centres.

Data availability

- *Data end users require* both marine species and habitat records to be made available on the NBN Atlas and via DASSH through mapping interfaces and/or an API for viewing and download.
- *Data end users require* to be able to reliably spatially display data before downloading it.
- *Data end users require* that data collected by the industry (commercial developers) and academic sectors is mobilised routinely into established databases and/or Scottish / UK repositories.
- *Data end users require* that public-funded data are made openly available within established databases for dissemination and re-use.

Data accessibility

- *Data end users require* simplified data access routes; clearer policies for data re-use with a reduction of artificial and unnecessary barriers in data usage agreements and licensing (e.g., all data attributed with at least CC-BY).
- *Data end users require* download of complete datasets at fine-scale granularity that can be looked at in different kinds of software packages.
- *Data end users require* coded datasets to enable easy merging or conversion.

Data quality

> *Data end users require* unduplicated data that can be easily sourced for use in collations

Non-functional requirements

Non-functional: defines how the system should do it; everything that makes the process happen (i.e., system properties and user expectations)

Infrastructure, technology and skills

- *Data mangers require* skilled people resource and funding (re-)prioritised for infrastructure maintenance and development and dataflow management.
- *Data managers require* dedicated resource to migrate data from old systems into new systems, to take advantage of advances in technology that facilitate more efficient data workflows and active upskilling.
- *Data providers require* clear sign-posting and best practice guidance for data submission and the associated dissemination opportunities.
- *Data end users require* standardised data structures within data types, with more cooperation in format and presentation, to enable straightforward collated use of data from a variety of collectors.

Data availability

> *Data end users require* data absence gaps to be addressed and re-fresh of out-of-date information.

Data accessibility

> *Data providers require* clear sign-posting to which portals and databases to use to access data, based on data type.

Data quality

- *Data managers and end users require* nominated long-term active data custodians who are responsible for error checking and correcting.
- *Data managers and end users require* the establishment of a more streamlined, robust and well-documented quality assurance and quality control process (clear audit / labelling of data quality) to engender confidence in available data.
- *Data providers require* guidance on how to publish data which retains appropriate acknowledgment of the data creator and/or owner.
- *Data managers require* adoption of a consistent approach to data recording to streamline the dataflow process and avoid record duplication.

Applicability of the SBIF Review recommendations

While the already established marine data infrastructure is likely to remain substantially separate from terrestrial and freshwater systems, a level of interoperability around the coastal zone and species that use both environments, in particular, would be beneficial to all parties.

The SBIF Review made 24 recommendations grouped by five outcomes. Work is underway through the SBIF Better Biodiversity Data (BBD) project to progress and implement some of the prioritised SBIF Review recommendations, for an improved biodiversity recording infrastructure by 2025. While thinking has since evolved within the SBIF for some of the recommendations, there are 7 recommendations from the SBIF Review report that are aligned to some degree with the 25 recommendations being made through this review for Scottish marine biodiversity data. The relevant SBIF Review recommendations are listed below:

- **SBIF Review Recommendation 2:** AFFILIATION OF DATA SUBMISSION ROUTES: All biological records should be submitted online and channelled to the NBN Atlas via standard, affiliated routes.
- **SBIF Review Recommendation 3:** SINGLE, CENTRAL ROUTE FOR CASUAL RECORDS: iRecord should be the single, central affiliated channel through which to submit 'ad hoc' records for verification, inclusion in relevant National Recording Schemes and dissemination via the NBN Atlas.

- **SBIF Review Recommendation 4:** PRIMACY OF AFFILIATED DATA SUBMISSION ROUTES: Biological records for a specific National Recording Scheme, recording group, project or organisation should be submitted via their affiliated route.
- **SBIF Review Recommendation 5:** PROVISION OF RECORDS COLLECTED UNDER LICENCE OR FOR CONSENT/STATUS: Biological records collected with public funding, under licence, for Environmental Impact Assessment or planning consent, or for an academic or professional qualification, should be provided to the NBN Atlas as a matter of good practice.
- **SBIF Review Recommendation 6:** RECOGNITION & RESOURCING OF A CENTRAL DATA MANAGEMENT PORTAL: Recorder 6 and Marine Recorder should evolve to become the common, central data management portal for data custodians to collate, view and manage their own biological records and datasets (unless a suitable internal business system is used).
- **SBIF Review Recommendation 8:** SYSTEM SIMPLIFICATION: The systems and tools available for collecting, curating, aggregating and disseminating biological records across all environments (terrestrial, freshwater and marine) and sectors should be rationalised.
- **SBIF Review Recommendation 23:** COMMUNITY FUNDS TO SUPPORT VERIFIERS, RECORDERS & OUTREACH: A Community Fund should be established to facilitate the scaling up of public participation in biological recording to ease current pressure points and to encourage participation and equal access for all.

One of the aims of this marine review was to identify where mutual opportunities exist between the marine and terrestrial sectors and create synergy and/or sharing of resources where appropriate. Shared opportunities for collaboration and coordination are considered in the bullets below in relation to establishment of a National Biodiversity Data Hub for Scotland.

- **Hub infrastructure development:** There is an opportunity for the SBIF BBD project to 'learn' from all the work that has gone into understanding and gathering requirements for the development of Marine Recorder Online (MRO), based on a user needs approach, to identify stakeholder needs and technical requirements: system functionality, schema, technology (APIs etc), to properly define what the scope and functionality of 'a single online system' for terrestrial data will look like. This has enabled an agile build approach and minimum viable product to be developed, with subsequent phases based on the prioritisation of user requirements.
- **Hub interoperability with marine data systems:** In the SBIF Review Recommendation 6, Marine Recorder (now the re-developed MRO system) was identified as the "central data management portal" for collating, viewing and managing marine benthic data. There is an opportunity to include marine data within the proposed Scottish national and regional LERC Hub model of self-sufficient income generation, through value added services to public and private sector organisations, with the following dependencies:
  - Incorporation of marine biodiversity data relies on accessibility to, and interoperability with, existing and developing marine databases (e.g., MRO, JCDP) and on the Hub staff having a knowledge of marine environmental data structures and ecology.
  - This could be achieved through, for example, the National Hub holding a MRO tenancy or through DASSH harvesting marine data (e.g., via API) from the Hub infrastructure (the nature of which is still to be determined).
- There is also significant opportunity for collaboration, within the implementation stages, to form a part of the solution to RECOMMENDATION 13 – infrastructure and management support for citizen science marine biodiversity recording.
  - The LERCs in coastal and island locations, e.g., Shetland, receive marine data from citizen science recording. Infrastructure to better support the flow, and/or collation, of this data into existing marine data repositories would help support the aspiration of community groups to have their data used in policy and decision making by Government to protect their local environment.
  - Any marine data submitted into the LERC infrastructure would be required to flow into the wider marine data network through, for example a tenancy within MRO and/or via an established workflow into the MEDIN DAC network (e.g., to DASSH).

Recommendation 24

RECOMMENDATION 24 – Embed marine expertise in, and interoperability of, the national and regional (LERC) hubs infrastructure in Scotland

- The NBN Trust should require, where possible, marine ecological expertise and/or marine data management expertise in at least one of the role holders recruited for the SBIF Better Biodiversity Data (BBD) Project, i.e., for one of the roles to be located in the National Hub for Scotland being established by that project.
- The infrastructure of National and Regional Hubs (i.e., the Scottish LERC infrastructure) should be scoped and developed to enable interoperability with the existing and developing marine data infrastructure; including Marine Recorder Online, the JCDP, and the MEDIN biodiversity data archive centre DASSH, so that data are made openly available for others to reuse under licence terms. This is in alignment with MEDIN's ethos of FAIR data.
- There is a need to further tease out the relevance to the LERC model (which is based on charging developers for 'value added services') to marine biodiversity data; if marine data submitted to LERCs flows efficiently into MEDIN DACs there would likely be little call on LERCs for 'value added services' as most enquiries about marine data would likely end up with DASSH (coastal data is possibly an exception).

This would enable incorporation of marine species and habitat data products into the 'value-added service' available, e.g., to industry and local authorities, through the Scottish Hub for use in coastal (terrestrial / freshwater-marine interface) planning and development. It would also facilitate the integration of existing marine data management infrastructure with the future Hub infrastructure to ensure that any marine data submitted into the LERCs flows into the wider marine data network. This links with RECOMMENDATION 10 – infrastructure to support habitat records.

Prioritisation of recommendations made in this marine review

High-level recommendations summary

The **25 recommendations** summarised in table 1 address the issues and barriers associated with the existing marine data infrastructure, technology and skills, data availability, data accessibility and data quality; brigaded under the six themes and colour coded based on their priority (value gained – investment required). Prioritisation and dependencies below describes how this prioritisation was done.

Annex F contains a summary of the full recommendations with associated action points.

Prioritisation and dependencies

A prioritisation matrix (figure 17) was used to score the high-level recommendations, firstly based on their impact or value (reward) and secondly on the effort or investment (time, money) needed to complete them. This exercise revealed that:

- 11 out of the 25 recommendations made in this review are considered 'quick wins'; where low investment is required but a high value is gained.
- 12 out of the 25 recommendations are considered 'major projects'; where in order to gain the high value, a larger investment (time and/or funding) is initially required.

LOW INVESTMENT

Medium priority
"do later"

High priority
"quick win"

LOW
VALUE

HIGH
VALUE

Low priority

Do next
"major project"

HIGH INVESTMENT
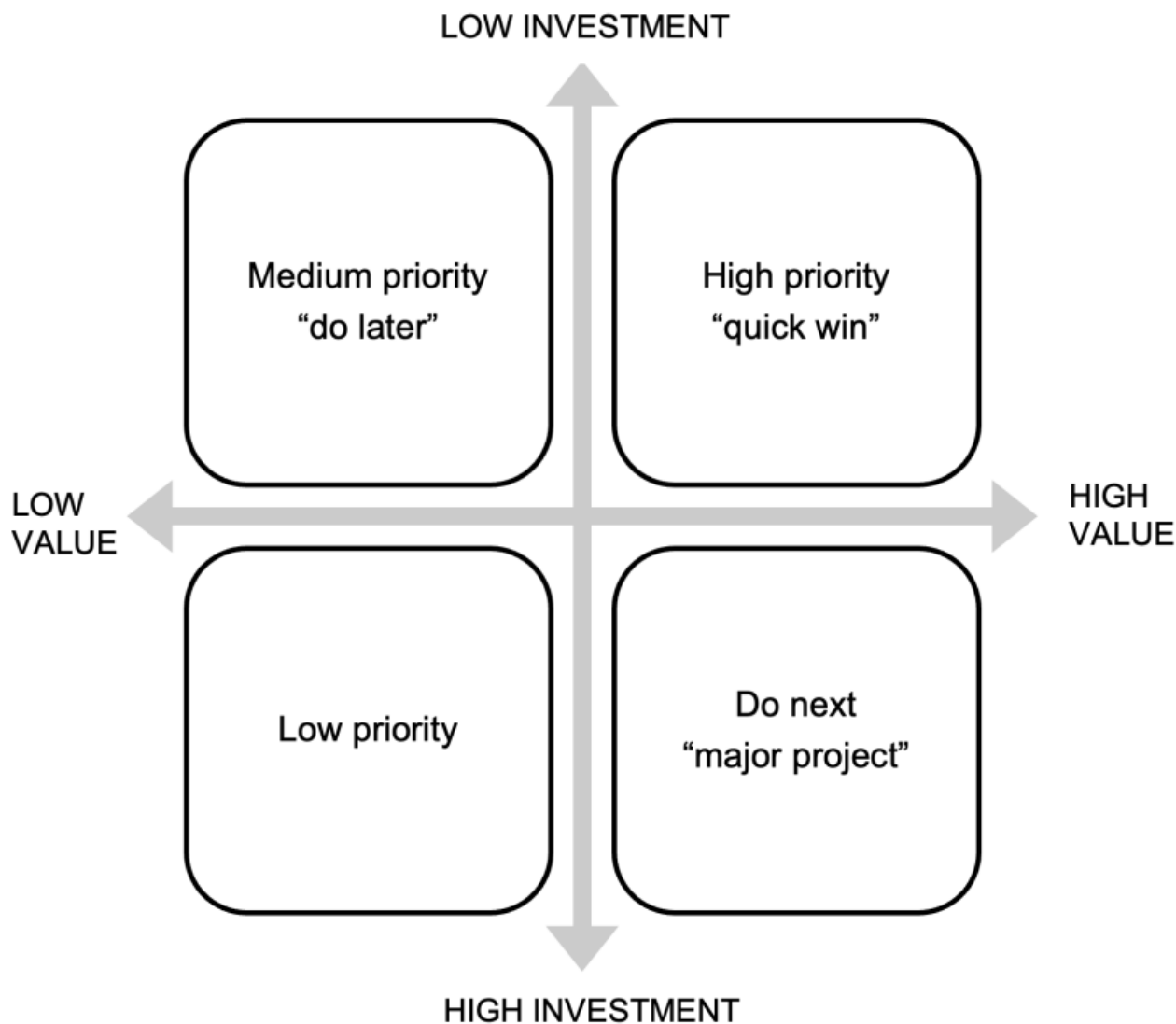
Figure 17: Prioritisation matrix used to score the high-level recommendations
Click for a full description

The recommendations dependency matrix in Annex G conveys the recommendations from this project that are dependent on the output or implementation of other recommendations. The matrix highlights that few of the quick wins rely on implementation of the bigger projects and/or huge resource input. Focus, initially, should therefore be directed towards implementing the quick wins whilst sufficient resource is allocated to supporting and progressing solutions to tackle the projects that require more effort to complete.

Some dependencies, which will require navigation within the next steps of this project (Future management and next steps), are as follows:

- RECOMMENDATION 2: UK government and devolved administrations recognise and resource key skills and infrastructure across the full data life cycle *[high investment, high value]*;
- RECOMMENDATION 3: Adopt primacy of affiliated data submission routes *[low investment, high value]*.
- RECOMMENDATION 8: Each record submitted to have a persistent identifier (PID) to prevent duplication *[low investment, high value]*;

Further analysis and work are required to fully understand the details of the changes required and the benefits of these changes related to each recommendation; this could be done through developing a benefits dependency network diagram as part of an Implementation Plan to action the findings from this review.

Recommendation 25

RECOMMENDATION 25 – Ensure future governance of marine data management in Scotland

A Scottish / UK advisory group should be formed to facilitate continued cross-sector stakeholder engagement and collaboration and guide implementation of the recommendations.

The group's role should involve:

- Guiding the development of an 'Implementation Plan'; this should follow an agile approach, focussing on priority areas and areas of highest value/benefit first.
- Within the 'Implementation Plan', develop a benefits dependency network diagram for the marine community, identifying the case for change:
  - Drivers of change
  - Change objectives
  - Benefits of change
  - Changes needed
- Oversee / monitor and provide leadership to progress, find solutions to, and implement the Review's recommendations.
- Ongoing and iterative collaboration between stakeholders.

Table 1. 25 high-level recommendations

The recommendations in this table are listed according to theme and scored to indicate their priority.

THEME 1: Continued engagement with key stakeholders

Recommendation	Prioritisation score
RECOMMENDATION 16: Develop proactive engagement with data custodian stakeholders who weren't fully involved in the review	Do next "major project"
RECOMMENDATION 25: Ensure future governance of marine data management in Scotland	Do next "major project"

THEME 2: Clarifying and streamlining data flows

Recommendation	Prioritisation score
RECOMMENDATION 3: Adopt primacy of affiliated data submission routes	High "quick win"
RECOMMENDATION 4: Map out marine data flows holistically	Low priority
RECOMMENDATION 5: Adopt primacy of Marine Recorder Online	High "quick win"
RECOMMENDATION 7: Agree a single, central route for casual records	High "quick win"
RECOMMENDATION 9: Formalise data flows between DASSH and the NBN Atlas	Do next "major project"

Recommendation	Prioritisation score
RECOMMENDATION 11: Clarify workflow responsibilities for mobilising benthic records to NBN Atlas	High "quick win"
RECOMMENDATION 20: Provision of biodiversity records collected under licence or for consent into the MEDIN Data Archive Centre network	High "quick win"

THEME 3: Improving the quality of existing data management

Recommendation	Prioritisation score
RECOMMENDATION 6: Clarify responsibility for tagging of records of conservation importance	High "quick win"
RECOMMENDATION 8: Each record submitted to have a persistent identifier (PID) to prevent duplication	High "quick win"
RECOMMENDATION 12: Progress a verification protocol for imagery derived data that complements the existing NMBAQC scheme component for grab and core sediment derived data	Do next "major project"
RECOMMENDATION 15: Plan for and fund the management and sharing of all new data being collected	High "quick win"
RECOMMENDATION 21: Maintain data version control through encouraging active custodianship	High "quick win"
RECOMMENDATION 22: Optimise re-use of data through adherence with FAIR Data Principles	High "quick win"

THEME 4: Investing in infrastructure and resource (people skills and funding)

Recommendation	Prioritisation score
RECOMMENDATION 2: Scottish (and UK) Government recognise and resource key skills and infrastructure across the full data lifecycle	Do next "major project"
RECOMMENDATION 19: Invest in data engineers and allocate resource for system decommissioning	Do next "major project"

THEME 5: Improving existing and creating new data infrastructure

Recommendation	Prioritisation score
RECOMMENDATION 1: Undertake a UK-wide marine biodiversity data infrastructure assessment	Do next "major project"
RECOMMENDATION 10: Develop infrastructure to support viewing and download of habitat records	Do next "major project"
RECOMMENDATION 13: Provide infrastructure and data management support for citizen science marine biodiversity recording	Do next "major project"
RECOMMENDATION 17: Develop simplified user interfaces onto repositories to support wider data submission	Do next "major project"
RECOMMENDATION 23: Develop existing portal infrastructure to support efficient searching, data display and dataset collation	Do next "major project"
RECOMMENDATION 24: Embed marine expertise in, and interoperability of, the National and Regional (LERC) hubs infrastructure in Scotland	Medium "Do later"

THEME 6: Simplifying existing and creating new guidance

Recommendation	Prioritisation score
RECOMMENDATION 14: Simplify the requirements for submitting data into DASSH whilst maintaining data quality	Medium "Do later"
RECOMMENDATION 18: Develop guidance on optimal data submission pathways	High "quick win"

Conclusions and next steps

Conclusions

Rapid developments in technology, together with novel data capture methods, are creating opportunities to accelerate the rates of marine biodiversity data recording and data sharing. While significant advances have occurred to improve marine data interoperability and transparency, the effect has been largely incremental. Many datasets are still not shared, are hard to find, and cannot be efficiently accessed. To support decision-making processes and scientific research, data should be directly and easily accessible and useable.

This report has covered a broad range of themes exploring the existing landscape of Scottish biodiversity data for seabed species and habitats, mammals (cetaceans and seals), birds (seabirds and waterfowl) and fish (including elasmobranchs). It also investigated the breadth of existing issues and barriers that hinder efficient access, collation and use the data being collected by a wide range of organisations and individuals.

Key findings

A large amount of Scottish marine data flows into the Scottish and wider UK infrastructures (e.g., databases, repositories, portals, and the Marine Environmental Data and Information Network (MEDIN)) at varying levels of efficiency; ranging from well-established automated workflows to ad-hoc or non-automatic workflows, depending on the biodiversity receptor (mammals; benthic; birds; fish) in question and the organisations contributing the data. Difficulties in identifying and accessing marine biodiversity data persist; it is widely recognised that current data flows could be simplified and that there are still barriers to be overcome with data sharing, spatial resolution and coverage. The existing framework and mechanisms to mobilise and access the wide range of existing marine biodiversity datasets can also be labour intensive and inefficient.

A key strength of the established data flow and systems is the ability to support the large volume of species and habitat data that are recorded and shared by Government bodies, but there is a widespread lack of clarity regarding roles, responsibilities and processes. Historical under-funding is a contributory factor, which has also limited capacity to capitalise on new infrastructure and advances in technology (e.g., cloud-based systems and widespread use of Application Programming Interfaces (APIs)). The lack of dedicated resource and skills within the existing infrastructure for efficient provision and management of commercial and third sector (NGO and citizen science) data is an issue, combined with cultural and behavioural barriers to data sharing, and prevents this data being easily and fully incorporated into the marine evidence base.

In compiling this report, it was found that the barriers and issues faced by stakeholders fell broadly into four categories: "infrastructure, technology and skills"; "data availability"; "data accessibility"; and "data quality", all governed, at least in part, by both technical and cultural aspects of openly sharing data. The report has discussed these categories and used them to guide its recommendations, aimed at making the management and use of marine species and habitat data more consistent, joined up and accessible, brigaded under the following six themes:

1. Continued Engagement with Key Stakeholders
2. Clarifying and Streamlining Data Flows
3. Improving the Quality of Existing Data Management
4. Investment in Infrastructure and Resource (skills and funding)
5. Improving Existing and Creating New Data Infrastructure
6. Simplifying Existing and Creating New Guidance

The **25 prioritised high-level recommendations** are set out under these six themes; 11 out of the 25 recommendations made in this review are considered 'quick wins'; where low investment is required but a high value is gained, these primarily sit under: Theme 2: Clarifying and Streamlining Data Flows; and Theme 3: Improving the Quality of Existing Data Management. Committed investment into people skills resource and data infrastructures and technology, now, will lead to long-term impact and savings (time and money) in the future; by creating more streamlined data workflows and more accessible data.

Some of the recommendations are specific to Scottish data flow, however the majority could or would have UK-wide implications and/or benefits. Many of the recommendations are centred on continued negotiation and discussions with stakeholders, cultural change and behavioural adaptation, building on/improving existing systems and workflows, and better sign-posting; rather than huge infrastructure change. Continued collaboration and coordination with key stakeholders; data providers, data managers, and data users, is critical to successfully implementing solutions.

Recommendations 14, 18, 20 and 22, in particular, made in this marine data review offer the opportunity for collaboration with the work that the SBIF are undertaking to progress recommendations identified in the 2018 SBIF Review for terrestrial and freshwater data.

Priority actions

The resulting **top priorities** to better coordinate and streamline the flow of Scottish biodiversity data, identified through this analysis of stakeholder needs, are:

- **Governance and stakeholder engagement:** Ensure future governance of marine data management in Scotland;
- **Data sharing:** Unification and/or rationalisation of databases and portals with clear guidance on where to submit datasets and the dissemination opportunities; a culture of data sharing and good management needs to be fostered, invested in and sustainably resourced;
- **Data accessibility:** Clear sign-posting to data resources, with consistent use of persistent identifiers to ensure that data are easily accessible for use in collations without duplication;
- **Data availability:** Address gaps in data availability through increased data sharing and publication and removal of data flow bottlenecks; improved access to industry survey data and ensure that academic research data are available within the Scottish / UK data landscape as well as internationally;
- **Collaboration and commitment:** Widespread adoption of Findable, Accessible, Interoperable and Reusable (FAIR) data principles to improve current practices (e.g., on data verification, adoption of standards, culture of sharing and 'open-ness' of data).

Key messages for achieving more integrated and accessible data

This section summarises the core requirements, in no particular order, to help make data more integrated, accessible and usable in Scotland:

1. Build knowledge of the main data stakeholders and their respective roles and competencies to provide a foundation for improving management, sharing, and wider use of data;
2. The purpose and benefits of any future infrastructure change for wider environmental outcomes are clearly articulated to achieve cultural change at the same pace.
3. Adherence across all sectors to data flow guidance to reduce unnecessary duplication of data and embed the critical verification stages to instil user confidence in data quality;
4. Data must be usable by machines, not just humans;

- Providing full metadata in a machine-readable form means that data discovery can be done via automated harvesting rather than manual searches.
- Full metadata will enable data to be repurposed, adapted, and applied for multiple functions, with appropriate attribution – 'collect once, use many times'.

5. Use Digital Object Identifiers (DOIs) to uniquely identify the source of data (datasets, models and data products), ensure provenance is clearly defined (including identifying the definitive version of a data set), and embed quality control using documented best practice systems (e.g., the NMBAQC scheme);
6. Organisational structures, relationships and governance, people resources and skills, and funding should all aim to support the FAIR data principles;
7. Making infrastructure simple and intuitive for non-data scientists, alongside the widespread use of technology and tools for automated workflows, will stimulate cultural change and enhance stakeholders understanding of the benefits of making data FAIR.

- Many data sources, federated from a number of providers, need to be delivered through a small number of portals and aggregators via web services and APIs to a range of users. *[There has been tendency for individual organisations to manage their own data collections and systems; in order to provide access to up-to-date, robust and fine-scale granularity datasets, the number of repositories and portals needs to be reduced because these are expensive to maintain]*.
- By simplifying the flow of data between organisations, costs for the management and dissemination of data are reduced, and the integration of unrelated datasets simplified.

8. Develop an environment of mutual trust between data providers and end users to promote data sharing.

- Understanding the needs of end users' (e.g., decision makers) is the crucial first step in the delivery of useful and impactful data.
- Feedback and communication from decision makers to data providers, of the decisions taken and impacts achieved, provides context, purpose and incentive for sharing. It can also be an opportunity to give guidance/direction on what data is needed and in what format to be of most use and make it fit-for-purpose.

9. Create full transparency of data use through careful documentation so that data providers are able to readily determine the impact of their open datasets through cited reference searches within the academic literature and in data download statistics and metrics.

Future management and next steps

Future management of the Scottish species and habitats data pathway will continue to be a collaborative activity between a range of stakeholders from the public, private and third sectors. Consulting them deeply enough to lay out any firm plans for this was beyond the scope of this study, making further discussion and partnership working essential.

Following the publication of this analysis report and a period of time for discussion with key stakeholders, an 'Advisory Group' will work to develop an appropriate strategy and Implementation Plan with a view to progressively fulfilling the recommendations. The Implementation Plan will, subject to consultation, develop a benefits roadmap for the public, private and third sectors to encourage each one to actively engage, support, and realise the benefits anticipated. Progress will be tracked through highlight reports and an annual programme review.

References

ABPmer, (2015). A Review of Access to Industry Environmental Data. A report produced by ABP Marine Environmental Research Ltd for Productive Seas Evidence Group, November 2015.

Economics for the Environment Consultancy Ltd (eftec), (2021). Mapping the species data pathway: Connecting species data flows in England.

Hassall, I., Cheffings, C., Robinson, A. & Robinson, P. (2020). Review of biodiversity data use in the Country Nature Conservation Bodies. *JNCC Report No.670*, JNCC, Peterborough, ISSN 0963-9091.

Jolly, C., *et al*. (2021) "Value chains in public marine data: A UK case study", *OECD Science, Technology and Industry Working Papers*, No. 2021/11, OECD Publishing, Paris.

Murray F., *et al*. (2018) Data challenges and opportunities for environmental management of North Sea oil and gas decommissioning in an era of blue growth, *Marine Policy*, Volume 97, Pages 130-138.

Wilson, E., Edwards, L., Judge, J., Johnston, C., Stroud, R., McLeod, C. and Bamforth, L. (2018) A Review of the Biological Recording Infrastructure in Scotland by the Scottish Biodiversity Information Forum: Enabling Scotland to be a global leader for biodiversity. Scottish Biodiversity Information Forum Commissioned. Report No. 1.

Annexes

Annex A – Project Advisory Group members

Name – Organisation:

Angus Jackson - *Seasearch (Marine Conservation Society)*

Annie Breaden - *Crown Estate Scotland (CES)*

Brian Eardley - *NatureScot / Scottish Government (Chair)*

Clare Postlethwaite - *UK Marine Environmental Data and Information Network (MEDIN)*

Clint Blight - *Sea Mammal Research Unit (SMRU), University of St Andrews*

Dan Lear - *Marine Biological Association (MBA)/DASSH (BioDIG lead)*

Gill Dowse - *Scottish Wildlife Trust (SWT) / Scottish Biodiversity Information Forum (SBIF)*

Graeme Duncan - *Joint Nature Conservation Committee (JNCC)*

James Dargie - *NatureScot*

Janet Khan - *Scottish Environment Protection Agency (SEPA)*

Jens Rasmussen - *Marine Scotland (MS)*

Jonathan Willet - *Scottish Wildlife Trust (SWT) / SBIF*

Lea-Anne Henry - *University of Edinburgh / MASTS*

Lisa Chilton - *NBN Trust*

Richard Shelmerdine - *Marine Centre UHI*

Rona Sinclair - *NatureScot*

Sophia Ratcliffe - *NBN Trust*

Annex B – Stakeholder organisations engaged with during the project

Engagement with multiple individuals within each stakeholder organisation (including in the early scoping stages) was common due to the diverse nature of the organisations work or research areas of expertise.

Organisation:

British Trust for Ornithology (BTO)

Cefas *(OneBenthic)*

Crown Estate Scotland (CES)

Data Archive for marine species and habitats data (DASSH)

Flora and Fauna International (FFI)

GeoSpatial Commission *(Natural England species data review)*

Hebridean Whale and Dolphin Trust (HWDT)

Heriot-Watt University (HWU)

Joint Nature Conservation Committee (JNCC) *(various leads for benthic, cetacean and bird data)*

Marine Environmental Data and Information Network (MEDIN)

Marine Management Organisation (MMO)

Marine Scotland Science / MS-LOT

Marine Alliance for Science and Technology Scotland (MASTS)

NatureScot *(internal - various leads for benthic, cetacean and bird data)*

National Biodiversity Network (NBN) Trust

National Museums Scotland (NMS)

Oil and Gas UK

Orkney Harbours Authority (OHA) *(mINNS data)*

Porcupine Marine Natural History Society

Royal Society for the Protection of Birds (RSPB)

ScotLINK

Scottish Association for Marine Science (SAMS)

Scottish Biodiversity Information Forum (SBIF) Advisory Group

Scottish Environmental Protection Agency (SEPA)

Scottish Renewables

Sea Mammal Research Unit (SMRU) University of St Andrews (UoStA) *(lead for seal data)*

Seasearch - Marine Conservation Society (MCS)

SeaWatch Foundation

Shetland Amenity Trust / Shetland Biological Records Centre

Shetland UHI

SOTEAG / University of St Andrews

The Crown Estate (TCE)

University of Aberdeen (& Lighthouse Field Station)

University of Edinburgh (UoE)

Whale and Dolphin Conservation (WDC)

Annex C – Data standards and controlled vocabularies

Darwin Core

The Darwin Core Standard (DwC) is an internationally recognised standard maintained by Biodiversity Information Standards (TDWG). It offers a stable and flexible framework for compiling biodiversity data from varied and variable sources. The standard includes a glossary of terms intended to facilitate the sharing of information about biodiversity by providing identifiers, labels and definitions; playing a fundamental role in the sharing, use and re-use of open access biodiversity data. DwC is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples and related information.

EurOBIS (and OBIS) use the OBIS-ENV data format, based on the Darwin Core Archive (DwC-A) standard for biodiversity, enabling data publishers to share their data using a common terminology.

Species occurrence records on the NBN Atlas use the DwC data standard, making them interoperable with those of other countries and allowing an easier export of data to GBIF.

The UK Marine Environmental Data and Information Network (MEDIN)

The MEDIN Partnership support the DwC standard for marine biodiversity data along with a number of data exchange formats. The MEDIN Data Guidelines provide a framework for capturing the assumed and otherwise potentially unrecorded knowledge necessary to re-use marine data. They provide a list of information that must be collected with marine data to ensure they can be re-used in the future, and have driven improvement in standard formats used for the exchange and storage of data. The guidelines for the collection and curation of marine biodiversity data are tailored to different marine data collection methods (e.g., underwater video or grab) and approaches including benthic and pelagic data collection at sea in addition to guidance for submission of ad hoc data.

Development of the MEDIN Discovery Metadata Standard (GEMINI 2 compliant and also compliant with other international conventions such as INSPIRE and ISO19115) has also improved the accessibility of marine data. It provides a framework for the collection of consistent discovery metadata; including information that accompanies a dataset to allow other people to find out what the dataset contains, where it was collected and how they can get hold of it. Importantly it includes measures of uncertainty in the data which help end users to determine what uses the data is appropriate for.

WoRMS and MSBIAS controlled vocabularies

The World Register of Marine Species (WoRMS) and the UK subset of taxa from WoRMS, known as Marine Species of the British Isles and Adjacent Seas (MSBIAS), is fundamentally a reference list aiming to provide an authoritative and comprehensive list of names of marine organisms, including information on synonymy, for use in biodiversity databases such as Marine Recorder.

Taxonomic information from infauna and epifauna sampling methods are required in the form of "AphiaID" identifiers from the international standard WoRMS system. Newly identified taxa are fed back through to the WoRMS taxonomic curators, to ensure that the system is up to date. MSBIAS is used as the source of taxa present in the Marine Recorder Species Dictionary. This list is actively maintained through the UK MEDIN partnership by the UK OBIS node.

Annex D – Portal and data aggregator descriptions

Portals and data aggregators that receive Scottish marine biodiversity data. The following descriptions have been developed in collaboration with the UK indicator data flow mapping project undertaken by JNCC (2021/2022).

NatureScot's Natural Spaces

Purpose: Natural Spaces is NatureScot's portal for mobilising marine and terrestrial natural heritage geospatial data held by the organisation (this includes marine species and habitats records (i.e., the GeMS collation of marine records of nature conservation importance); protected area boundaries; land forms and geology). The portal provides open access (3-4 stars) to NatureScot's main spatial datasets as downloads; users can browse through the available Scottish datasets on the webpage and access the data in several different GIS formats or via WMS consumption.

Scope: Covers a wider user base than just marine (also covers terrestrial and freshwater datasets). Natural Spaces contains all mobilised publicly available data held by NatureScot. The portal doesn't offer a 'view' of the data. Natural Spaces contains full spatial datasets that users can download as a zipped file or consume via WMS and import to their own GIS systems for their own purposes, providing a resource more tailored to individual analyses rather than for planning, development purposes or general public viewing. The availability of a WMS, which provides a raster image, rather than WFS restricts the usability of some data types (e.g., sightings biodiversity data) directly in detailed analyses.

Marine Scotland's National Marine Plan Interactive (NMPi)

Purpose: NMPi is Scotland's marine planning portal. An interactive tool that enables user access to spatial information relating to the marine environment and activities. Designed to assist in the development of Scotland's national and regional marine planning and to support work of the regional Marine Planning Partnerships. NMPi allows users to view different types of information and, where appropriate, links are provided to the related parts of Scotland's Marine Assessment and to the National Marine Plan.

Scope: Marine and coastal (qualifying PMF and Annex 1 habitats) specific datasets (no terrestrial). Layers include biological datasets, physical data, industry development data, and administrative and boundary data. Data are fed by WMS feed, layers are served up by multiple organisations, including NatureScot on a routine basis in line with UK Marine Recorder snapshot provision and GeMS collation. Users gain public access to a Scottish picture of biodiversity records tagged with conservation status used in policy and management advice by Government, NatureScot and SEPA.

MSODN (Marine Scotland's Open Data Network)

Purpose: The Marine Scotland's Open Data Network platforms are being consolidated into a more integrated platform (2022/23) containing links to datasets and information made publicly available through Marine Scotland. It also signposts users to the wide range of additional, supporting resources that are available online from Marine Scotland and other organisations. Datasets and maps are grouped together by topic/theme and the content is categorised in to three types: Information, Maps and Data.

Scope: Marine focussed. MSODN includes MSI, MS Maps NMPi, MS Data and MS Assessment (the new home for Scotland's Marine Assessments, including Scotland's Marine Assessment 2020).

Marine Scotland Information (MSI): A web portal that provides detailed information and data about the Scottish marine environment. It has been designed to bring together: the information pages that support the spatial layers in Marine Scotland MAPS NMPi, providing metadata and links to related resources; contextual information and descriptions for data resources provided through the Marine Scotland Data Portal; the content previously held on the Marine Scotland interactive (MSi) web pages.

Marine Scotland Data (MS Data): A dedicated portal that allows users to search Marine Scotland's published datasets and reports. Citation information for these datasets is provided through the use of DOIs. The portal provides a single point of access to Marine Scotland's published data, and allows users to explore, download, share and cite those data; provides a user interface for searching datasets as well as machine readable services to locate and retrieve data; describes each data set with standardised metadata, and downloadable resources are described in detail in terms of units, use of vocabularies, etc; groups data sets into broader topics to help exploration, but all content is also searchable right across the portal. Downloadable data are made available as 3-4 star open data and is released under the UK Open Government Licence, where possible; uses persistent identifiers to allow accurate citation and location of datasets.

Marine Scotland Maps NMPi: [*see above re: NMPi*] An online, interactive GIS-based tool allowing users to view different types of information (as layers) at various scales.

Scotland's Environment (SEWeb)

Purpose: SEWeb, historically managed by SEPA, was built as a web platform with the purpose of being a gateway to everything that data users would want to know about Scotland's environment; bringing together environmental data and information into one place so that is easy to search, discover, analyse and interpret.

Scope: The functionality is as a sign-posting website to established portals such as NBN and NMPi etc. Not marine focussed.

NBN Atlas

Purpose: NBN Atlas is a UK portal that collates species records from various organisations including LERCs into a national picture on an online web portal for users to browse and download, according to the reuse licence. The NBN Atlas data is open access by default, except in relation to sensitive species which is restricted, and licenced as freely available (e.g., under OGL; CC-BY) or as CC-BY-NC (i.e., not consented for commercial use without prior agreement from the data owner). The NBN Atlas combines multiple sources of information about species occurrence in the UK, with the ability for users to interrogate, combine, and analyse these data in a single location.

Scope: provides a UK picture of both marine and terrestrial species data together. It is not a data management system, but rather a discovery point for users to find datasets; it allows users to view species records together with other environmental information and geographical boundaries and to download and export maps and reports or summaries for your own use. The NBN Atlas is the UK node of GBIF and so it also provides a mechanism for disseminating species data internationally.

UK MEDIN (Marine Environmental Data and Information Network)

Purpose: MEDIN is a UK portal established in 2008 as the hub for UK marine data and provides the framework for the management and harmonisation of marine data in the UK. Its primary objectives are to improve access to and management of UK marine environmental data and information. The MEDIN model operates using a series of standards and specifications, underpinned by a network of 7 thematic Data Archive Centres (DACs). MEDIN provides a mechanism for the Governments and their agencies to meet their obligations under the EU INSPIRE Directive within the marine sector. MEDIN also provides measurable benefits to the UK economy by providing efficient access to marine data for government, industry, academia and eNGOs and by supporting better decisions via a more comprehensive evidence base.

Scope: The MEDIN Portal provides users with a tool for searching and discovering UK datasets collected or managed by over 600 different organisations. Most of these datasets are accessible from the MEDIN DAC network   Data end users can search for data online through the metadata discovery portal or through searches on DAC portals. MEDIN provides a service that brings together data resources in a thematic way, into one place, so that users can find biodiversity data and related environmental survey data – and then be directed to the specialist DAC or data custodian for access to the relevant dataset(s). Search results can be exported, and data can be downloaded where available and many datasets are also made accessible using web mapping services, increasingly API's and with Digital Object Identifiers (DOIs) for citation purposes, for example in peer review publication.

DASSH (Archive for marine species and habitats data)

Purpose: DASSH operates as the UK MEDIN archive for marine biodiversity data. DASSH also operates internationally as the UK node of the Ocean Biodiversity Information System (OBIS) and is a partner in EMODnet Biology. DASSH provides tools and services for the long-term curation, management and publication of marine species and habitats data, within the UK and internationally. DASSH are a key provider of marine data to the NBN Atlas. As an archiving centre, DASSH also provide a server infrastructure for storing raw digital data files such as video and still images collected on surveys. DASSH species data holdings can be browsed and downloaded via the DASSH Data Mapper online tool and can also be accessed in the DASSH database via using WFS (Web Feature Service), OGC (Open Geospatial Consortium) WMS (Web Map Service) or via API.

Scope: DASSH has well established links between UK and International marine data systems, which other UK databases and portals, such as NBN, do not have. DASSH also archives and publishes fully attributed data. DASSH supports both marine species and habitats data. DASSH, as a DAC, has a flexible database structure and is able to receive data from many different sources and in multiple formats. DASSH fulfils the niche well as a data archive and data disseminator.

EMODnet Biology / EurOBIS

(European Marine Observation and Data Network Biology; European node of the Ocean Biodiversity Information System (OBIS))

Purpose: EurOBIS and EMODnet Biology act as an international data aggregator / portal. EurOBIS is the European Node of the international Ocean Biodiversity Information System (OBIS). EurOBIS aims to centralise biogeographic data on marine species collected by European Institutions inside or outside Europe. Data published through EurOBIS is freely available through EurOBIS, EMODnet Biology, OBIS, GBIF. The EMODnet Biology portal provides free access to data on temporal and spatial distribution of marine species and species traits from all European regional seas. It is built upon the World Register of Marine Species and EurOBIS; EurOBIS is the data system that underpins EMODnet Biology. The overarching aim of the network is to convert Europe's otherwise fragmented marine data landscape into an interoperable data sharing framework, adopting the "*collect once, use many times*" data philosophy. MEDIN is part of the EMODNet network.

Scope: EMODnet Biology / EurOBIS focuses on marine species data across Europe. It is an online marine biogeographic database compiling data on all living marine creatures. The database focuses on taxonomy and occurrence records in space and time. When data are added to EurOBIS, the data are immediately available

through the EurOBIS and EMODnet Biology Portal. On a regular basis, all the EurOBIS data are also sent to OBIS, which in turn sends its data to GBIF. It also provides a number of additional data services to browse, visualise and retrieve data layers from various disciplines and themes simultaneously.

EMODnet Seabed Habitat

Purpose: EMODnet Seabed Habitats is an international aggregator / portal that provides access to seabed habitat data across Europe. This includes EMODnet broad-scale seabed habitat map for Europe (EUSeaMap). It is part of the European Marine Observation and Data Network (EMODnet) and continues the work started by MESH and MESH Atlantic projects in collating and making available European seabed habitat maps from surveys through the map viewer.

Scope: EMODnet Seabed Habitats focuses solely on modelling and categorizing seabed habitats in European waters.

ICES

Purpose: The ICES data portal is separated into several thematic portals focused on the marine environment including benthic and pelagic biota as well as oceanographic and pressure data. Data in the ICES data portal are collected for the purpose of aiding assessments of expert groups and regional sea conventions. The ICES data portal has a web-based user-interface which provides a suite of tools which help visualise and calculate data products. Data held in ICES data portal contributes to OSPAR CEMP, ICES stock assessments and AMAP contamination assessments.

Scope: The ICES data portal focuses on the ICES regions and providing data for specific assessments

OSPAR ODIMS

Purpose: The OSPAR Data and Information System (ODIMS) is an online tool providing a single point of access to all the data and information gathered through OSPAR's Joint Assessment and Monitoring Programme across the different thematic work areas of the Convention. It will help ensure that data is readily accessible for OSPAR assessments, but also help a broad range of users to find data held by OSPAR, to facilitate access to it and make use of it.

Scope: ODIMS is focused on the OSPAR regions and includes data from different aspects related to Ocean health which include information on benthic species but also on offshore industry, hazardous Substances, environmental impact of human activity etc. It is specifically designed to hold data for OSPAR assessments.

Annex E – Biodiversity receptor data flow mapping – descriptive text

Benthic data

JNCC leads on collating and sharing a UK dataset of marine benthic data and it is through this network that the integrity of the UK picture has been maintained and shared. Through this arrangement NatureScot gain access to relevant data collected by others in Scotland's seas i.e., Marine Recorder does not just contain data collected or commissioned by NatureScot.

NatureScot uses Marine Recorder to capture and manage data from individual marine seabed surveys. Biodiversity data products from the JNCC UK 'snapshot' are made available internally to staff and mobilised externally as a collated subset of marine species and habitat records tagged with conservation status (GeMS) to:

- Natural Spaces as a GIS download
- Marine Scotland NMPi via NatureScot web mapping service (WMS)
- National Biodiversity Network (NBN) Atlas Scotland

The data support a range of uses including the establishment of Marine Protected Areas (MPAs), PA monitoring, development management casework and marine planning.

The new Marine Recorder Online application supports many key improvements including:

- Ability to handle complex spatial geometries.
- The streamlined mobilisation of data - allowing direct harvesting by the MEDIN data archive centre network.
- Compatibility with widely accepted UK marine data standards, and direct linking to associated discovery metadata.
- Allow the tagging of records with conservation status like Scottish Priority Marine Feature or Habitats Directive Annex I habitat making record identification for conservation designation and management purposes easier.
- Functionality to enable organisations that have a MRO tenancy to have long-term data custodianship responsibility but delegate access to contractors or other organisations, e.g., commercial developers, to allow data entry and editing.
- Data custodian control over when products stored in MRO are published, meaning that the temporary embargos or tiered access on datasets (although there is a preference for data to be released immediately) can be accommodated.
- An ongoing relationship with the developer to provide user support to facilitate further development and maximisation of the applications' life-span.

Cetacean data

The JCDP platform (database and associated portal infrastructure) is an important step towards addressing the patchiness in time, space and scale of datasets; aiming to maximise the value of cetacean monitoring data through promotion and facilitation of standardised data protocols and submission requirements. The platform is hosted by the International Council for the Exploration of the Sea (ICES) Data Centre. A MEDIN-compliant cetacean data guideline standard for UK dataset collection is being co-developed through the JCDP; enabling streamlined upload to the JCDP, archiving with DASSH and provision of a 'UK picture' of cetacean data. Importantly, there will be two tiers in the JCDP for data release (open and restricted available on request) and resolution, dependent on the user's affiliation.

Sightings data:

NGOs, including the Hebridean Whale and Dolphin Trust (HWDT) and Whale and Dolphin Conservation (WDC) in Scotland, currently store and publish data within their own internal data management systems. Processed data and raw data are treated separately; raw data are not released openly at capture resolution and will likely continue to be restricted in the JCDP; however, the data are made available on request. This enables the NGOs oversee how and what the data are being used for, and retain an important feedback loop for the charities volunteers and stakeholders to report back the value of the data that they collect. Processed sightings data (e.g., SPUE and DPUE) from HWDT and WDC will be submitted into the JCDP once the system is live (Spring 2022). Sightings data collected and stored in the UK ORCA database will also flow into the JCDP and there is an aspiration for data collected by commercial developers as part of the consenting and licensing process, currently stored with Marine Scotland, to also flow into MEDIN archives via the JCDP.

The SeaWatch Foundation database currently provides a central archive for cetacean land-based effort and sightings point data collected from all around the UK. Citizen scientist volunteers (e.g., CalMac Awareness; WDC ShoreWatch etc) can submit cetacean sightings in recording forms or via the 'SeaWatcher' recording app, which then get verified/validated and integrated to the main database. There are reported difficulties in accessing and downloading non-transect sighting data in this current infrastructure. There is the potential to consider effort-related land-based sightings data to be included in the JCDP in future.

Acoustic data:

The Scottish Association for Marine Science (SAMS) is investigating the use of the acoustic recordings metadata database Tethys for storing processed acoustic detections metadata (including information about the detectors, detection thresholds, data cleaning etc) from the COMPASS project, making the data accessible for modelling work/meta-analyses and for use to compliment effort-related sightings data. There is potential to

integrate these types of data into the JCDP in future development. A feasible long-term solution for safe storage and archiving of the raw data files is also under discussion with MEDIN and its network of data archiving centres.

Photo-ID data:

The WDC manage a catalogue for Lewis Risso's dolphin sightings, HWDT and SAMS manage a photo-ID catalogue for the west-coast of Scotland, SMRU (University of St Andrews), and the University of Aberdeen (Lighthouse station) manage the bottlenose dolphin (Moray Firth) photo-ID catalogue. There is certainly a degree of photo information sharing between these responsible organisations in relation to species and/or location of sighting, but the data is not currently easily accessible for public download and use.

Seal data

All statutory seal monitoring in Scotland is carried out by the Sea Mammal Research Unit (SMRU), with the exception of land and vessel counts of grey seal pup in Shetland which are undertaken by NatureScot and submitted to SMRU. Three types of aerial survey are undertaken by SMRU: August helicopter surveys (mainly for harbour seals during their moult but grey seals seen are also counted) aim to cover the entire Scottish coastline every 5 years, using a multi-sensor gimbal with a thermal imaging video camera and high resolution photographs. The inner Moray Firth and the Firth of Tay and Eden Estuary SAC are surveyed annually in August, using oblique photography from a small fixed-wing aircraft. Fixed wing aircraft surveys of grey pups at major colonies in Scotland now take place biennially using vertical photography. Telemetry (movement tagging) data are also collected through SMRU (University of St Andrews) and Lighthouse Field Station (University of Aberdeen) projects and used in combination with count data to generate estimates of at-sea distribution/use.

SMRU have their own spreadsheets, databases and networked drives (jointly managed by staff from SMRU and the University's central IT department). SMRU provide collated high-resolution August harbour seal moult count data for every census period (~5yrs) and autumn grey seal pup modelled population data (~2yrs) to NatureScot for use in policy and advice. This data is published to Marine Scotland's NMPi via NatureScot's GeMS collation and is available via WMS (with key) for users to interact with, but is not queryable or downloadable. Seal monitoring data is not widely mobilised for public access, e.g., to the NBN Atlas, however work is in progress to make gridded count data publicly discoverable and accessible through DASSH at 5x5km2 resolution which would in future also facilitate data flow into the wider network to EurOBIS etc. The OSPAR Data & Information Management (ODIMS) database also has a mix of publicly available maps and datasets (e.g., from indicator assessments), but this data is not easy to find. Data within the ICES seal database is available for authorised users only. A yearly NERC Special Committee on Seals (SCOS) scientific report is published with the latest counts, and estimates of population and trends at varying spatial resolution including regional Seal Monitoring Units (SMUs; Scottish SMUs: Southwest Scotland, West Scotland, Western Isles, North Coast and Orkney, Shetland, Moray Firth and East Scotland). This report includes the committee's scientific advice on matters related to the management of seal populations in response to questions from Scottish Government as well as Defra and Natural Resources Wales.

The collection of the telemetry data from Scottish seals uses electronic tags (for the most part produced by SMRU Instrumentation) and has over the years has been funded by various organisations, including BEIS, Scottish Government and industry. The datasets from such tags are received, processed and then stored in the group's Oracle database by SMRU Instrumentation. Data ownership is variable, depending on each individual project's funding and contractual terms. Such telemetry datasets are therefore securely archived within the University of St Andrews data centres but those versions are not always easily accessible and available for general use.

Photo-ID datasets exist for both seal species; although ongoing data collection is now largely limited to the work being carried out by SMRU under the Scottish Government funded harbour seal decline project as well as the long term photo ID study of Loch Fleet harbour seals carried out by the University of Aberdeen.

Bird data

Seabirds:

The Seabird Monitoring Programme (SMP) database is currently hosted by the BTO and is developed and maintained by JNCC; from April 2022 onwards JNCC, BTO and RSPB will start a new partnership. The SMP is an important database comprising long-term whole-colony count and breeding success data for 25 species of breeding seabirds in Britain and Ireland. Data are inputted to this database from the annual seabird monitoring programme, led and coordinated by the JNCC in partnership with 18 other UK organisations, a network of volunteers and professional bodies. Representative samples are collected and analysed, and robust country abundance and breeding success trends are generated annually through the SMP online report.

In addition to the annual monitoring scheme, JNCC leads, in association with other SMP partners, on the development and completion of periodic breeding seabird censuses across Britain and Ireland. To date, there have been four breeding seabird censuses completed: Operation Seafarer, Seabird Colony Register, Seabird 2000 and Seabirds Count.

The SMP monitoring handbook is a single online resource compiling the standard methods for surveying and monitoring seabird colonies; providing species-specific step-by-step procedures so that the data entered into the SMP database is collected in a consistent and systematic manner by the volunteer recorders. Raw data is easy to download directly from the SMP database and end users can use it for what they want; bespoke datasets can be requested from JNCC (e.g., for PhD research). The SMP datasets are well managed and freely accessible. The user interface is being developed to enable data queries on SMP samples and census data and graphs can be generated in the future. The data resources are generally used to produce abundance and breeding success trends and national and international indicators and help researchers and conservation organisations understand seabird populations and assist in their conservation.

Since 2017, JNCC and the Statutory Nature Conservation Bodies (SNCB) have funded additional development of the SMP database via a hosting agreement with the British Trust for Ornithology (BTO). In April 2020, JNCC launched the new SMP database which included an improved interface and many new, user-friendly, features to enter and view seabird monitoring data.

The BTO RAS database, which holds also data on seabird survival from a mark-recapture programme, might in future be incorporated into the SMP database but this is dependent on funding.

The RSPB have their own internal database and Open Data Portal, but contribute breeding seabird colony data to the SMP 'in kind'. The RSPB collect colony data on their reserves (e.g., Shetland and Orkney) and make publicly available seabird tracking data for selected species from the twin FAME/STAR projects to Marine Scotland's NMPi portal. The ICES European Seabirds at Sea (ESAS) database, managed by ICES, holds effort related seabird at sea ship observation data collected by trained professionals and volunteers.

Waterbirds:

The WeBS (the Wetland Bird Survey) run by the BTO monitors non-breeding waterbirds (wildfowl (ducks, geese and swans), waders, rails, divers, grebes, cormorants and herons) in the UK. The WeBS database holds inshore wintering waterfowl data covering inland and coastal regions, collected by thousands of volunteers. The data is not as freely available as SMP data however, a variety of data outputs are available and if a user requests the data, then it is provided. Bespoke commercial data requests are also provided at a charge. Based on WeBS, layers for seasonal mean peak numbers of wildfowl, waders and cormorants and divers are made publicly available through Marine Scotland's NMPi portal.

Seabird and waterfowl monitoring data collected by the Scottish offshore industry (e.g. offshore wind developers and decommissioning of oil rig platforms) through the consenting and licensing process is not currently input to any of the above databases or shared publicly; however there is an aspiration to include aerial survey data, platform survey data and tracking data into the appropriate databases with linking up to the MEDIN data archive centre DASSH and development of a MEDIN data guideline to support submission of bird data.

Annex F – Full recommendations summary

RECOMMENDATION 1 – **Undertake a UK-wide marine biodiversity data infrastructure assessment**

The UK Monitoring and Assessment Reporting Group (MARG) should expedite a UK-wide marine biodiversity data infrastructure assessment to inform development and agreement of a strategic and integrated technical road map that will simplify the data flow and connectivity of infrastructure.

The five actions below should build on the mapping analysis work done in this current Scottish marine biodiversity data review and work done by the JNCC on UK Marine Strategy indicator data flow mapping to achieve a UK-wide assessment.

1. Provide a clear definition of each of the key components and tools in the marine species and habitats data infrastructure. *This would assist in improving the automated harvesting of records and integration of standalone system data flows;*
2. Clarify the linkages between data repositories, portals, aggregators, DACs etc. *This would ensure that data providers are secure in the knowledge of where their records will be available for use and where they can aid decision-making;*
3. Endorse the key roles each portal and repository fulfils. *This would maximise inter-operability, coordination and ease and speed of data flow into and from the MEDIN DACs and from EU / international portals directly receiving data. It would also promote the endorsed portals and repositories as those supported by the wider marine community and help to encourage more extensive uptake and use;*
4. A single central directory of all Scottish / UK affiliated data submission routes should be developed and maintained by MEDIN. *This would facilitate streamlined data submission into the MEDIN data archive centre network, via affiliated routes (e.g., repositories). A link from SEWeb (or its re-developed form) could also be made to sign-post the directory in future.*
5. Map out the current and future capabilities of organisations with an interest in the data flow pathway. *This would ensure that: the organisations involved have appropriate workforce skills, resource and funding to undertake recommended improvements; the data infrastructure / data flow is joining up; and that system integration is being improved.*

RECOMMENDATION 2 – **Scottish (and UK) Government recognise and resource key skills and infrastructure across the full data lifecycle**

The key components of the Scottish marine data flow landscape should be recognised and resourced as:

- Core/central database management systems (e.g., Marine Recorder Online; JCDP);
- MEDIN data archive centres (e.g., DASSH);
- Scottish specific and UK-wide marine data portals (e.g., NBN Atlas, NMPi Portal).

Each component should have a clearly defined role, enabling them to work together as a collaborative, connected network.

This links to: a task under RECOMMENDATION 1 – a clear definition of each of the key components and tools in the marine species and habitats data infrastructure; and RECOMMENDATION 3 – primacy of affiliated data submission routes.

RECOMMENDATION 3 – **Adopt primacy of affiliated data submission routes**

Marine species and/or habitats records should be submitted into the appropriate established database where the database remit permits (e.g. benthic species and habitat occurrence data into Marine Recorder Online, cetacean at-sea effort-related vessel and aerial sightings transect data into the JCDP), as the recognised data entry point to the data flow network, and channelled to a MEDIN DAC (e.g. DASSH) via standard, affiliated workflows for onward dissemination to the NBN Atlas and other data aggregators (e.g. EurOBIS/EMODNet).

This will help avoid duplication of effort and indirect data flows. It would also reduce the complexity of collating records for individual/organisation purposes and help prevent version control and/or record duplication issues.

Links with a task under RECOMMENDATION 1 – a directory of all affiliated Scottish / UK data submission routes maintained by MEDIN.

RECOMMENDATION 4 – **Map out marine data flows holistically**

A holistic picture of the Scottish data flow landscape (e.g., seabed, mammal, fish and bird data) should be mapped out.

The mapping should build on the individual receptor data flows mapped for Scotland in this current analysis review. This would clearly outline the infrastructure that a data provider would be faced with when deciding where to submit their dataset into the data network to the relevant receptor database or repository.

This links with RECOMMENDATION 21 – development of guidance on optimum data submission routes.

RECOMMENDATION 5 – Adopt primacy of Marine Recorder Online

Government bodies in Scotland (NatureScot, JNCC, Marine Scotland and SEPA) should adopt Marine Recorder Online, once it is available in 2022, as the data management and storage solution for benthic species and habitats data.

This links with RECOMMENDATION 3 – primacy of affiliated data submission routes.

RECOMMENDATION 6 – Clarify responsibility for tagging of records of conservation importance

Responsibility for tagging records of conservation status (Priority Marine Features and Annex 1 habitats) in Scotland should remain with JNCC / NatureScot. Marine Recorder Online should be used as the mechanism to do this for benthic data.

This would streamline the dissemination of records of conservation importance to Marine Scotland's NMPi Portal to inform marine planning and management decisions.

RECOMMENDATION 7 – Agree a single, central route for casual records

DASSH should be recognised as the single, central route for the submission of casual Scottish marine biodiversity records/datasets that are not submitted directly to NBN Atlas via apps such as iNaturalistUK and iRecord.

This links with RECOMMENDATION 9 – formalise the data flow between DASSH and the NBN Atlas.

RECOMMENDATION 8 – Each record submitted to have a persistent identifier (PID) to prevent duplication

Disciplined implementation of PIDs should be adopted (i.e., not altered or prefixed by different systems throughout its lifetime) by each data entry point system. PIDs should be allocated at every level of the survey hierarchy by data repositories at the point of data submission by recorders to prevent record duplication in data collations by enabling easy linking/identification of the same record shared to aggregators from different organisations.

The allocation of PID's (inc DOIs) would help with Findability and Interoperability in terms of dataset versioning and also contribute to data provenance to ensure it is fully traceable (Reusable).

RECOMMENDATION 9 – Formalise data flows between DASSH and the NBN Atlas

DASSH and NBN Trust should maintain the established workflow of records from DASSH to the NBN Atlas, and formalise the existing ad hoc workflow from the NBN Atlas into DASSH into an automated workflow, to create an efficient two-way exchange of records.

This would facilitate collation, mobilisation and archiving of marine species records (and habitats in due course) that are submitted directly to the NBN Atlas by recorders, e.g., via iNaturalistUK, iRecord.

This links to, and relies on successful implementation of, RECOMMENDATION 8 – persistent identifiers (PID).

RECOMMENDATION 10 – Develop infrastructure to support viewing and download of habitat records

Resource should be prioritised by (or additional provided to) MEDIN and the NBN Trust, respectively, to:

- Develop the DASSH species mapper infrastructure so that it is capable of also supporting habitats data, with the ability to access both species and habitats data through an API.
- Develop the NBN Atlas infrastructure so that it supports both species and habitats occurrence records, and ensuring that API access covers both species and habitat data.

This would help to deliver the infrastructure required to enable end users of data to efficiently navigate, browse, find and download available species and habitat data resources (complete datasets). There is a need to clearly define the niche and purpose of each system to streamline data flow and avoid duplication.

Links to RECOMMENDATION 9 – formalise the dataflow between DASSH and the NBN Atlas; and RECOMMENDATION 23 – Develop existing portal infrastructure to support efficient searching, data display and dataset collation.

RECOMMENDATION 11 – **Clarify workflow responsibilities for mobilising benthic records to the NBN Atlas**

DASSH should become the responsible organisation for mobilising Scottish benthic species (and habitats in due course) records to the NBN Atlas, on behalf all Marine Recorder Online custodians with records relating to Scotland's seas (e.g., NatureScot, JNCC, Seasearch); utilising the [developing] automated workflow from Marine Recorder Online to DASSH.

This arrangement should supersede the existing arrangement whereby Marine Recorder data custodians (e.g., NatureScot, JNCC, Seasearch) are responsible for publication of their own species data to the NBN Atlas, simplifying and streamlining the data workflow.

This links with: RECOMMENDATION 3 – primacy of affiliated data submission routes; and RECOMMENDATION 5 – primacy of Marine Recorder Online.

RECOMMENDATION 12 – **Progress a verification protocol for imagery derived data that complements the existing NMBAQC scheme component for grab and core sediment derived data**

This supports the NMBAQC's existing commitment to develop a component of the scheme for epibiota via implementation of the UK Benthic Imagery Action Plan and JNCC's Big Picture work.

This links to an action within RECOMMENDATION 13 – develop a verification protocol for citizen science stakeholders**.**

RECOMMENDATION 13 – **Provide infrastructure and management support for citizen science marine biodiversity recording**

A targeted piece of analysis should be undertaken to fully understand the priorities for investment and/or infrastructure necessary to better support the flow of citizen science data into the marine evidence base.

This should include:

1. Assessing the need for provision and update of data management protocols, technical guidance and clear sign-posting of data submission routes available to citizen scientists.
2. Identifying the key data 'types', species groups, methods of data capture (apps, web forms etc), and spatial data visualisation tools to inform the priorities for investment.
3. Development of a verification protocol(s) with key citizen science stakeholders in the verification process, which aligns to current and future verification requirements and technologies. The protocol(s) need to cover the broad range of citizen science data collection methods and expertise. The resources required to support implementation of the protocol and capacity building should also be identified.

This recommendation provides the opportunity to explore funding options, including via the Scottish Marine Environmental Enhancement Fund (SMEEF). The recommendation also has synergy with the SBIF Better Biodiversity Data (BBD) Project* proposed to improve the management and long-term sustainability of LERC citizen science data.

Currently terrestrial and freshwater species focussed.

This links to: RECOMMENDATION 12 – Progress a NMBAQC scheme component for imagery derived data;RECOMMENDATION 14 – Simplify the requirements for submitting data into DASSH; RECOMMENDATION 17 – develop simplified user interfaces onto repositories; and RECOMMENDATION 18 – guidance development on optimal data submission pathways.

RECOMMENDATION 14 – **Simplify the requirements for submitting data into DASSH whilst maintaining data quality**

DASSH should simplify the existing requirements that need to be met by recorders in order to submit their datasets, whilst maintaining the quality of data submitted. This should be through provision of support to users of the formal data guidelines to translate and produce practical step-by-step guidance for their peers to facilitate submission of new data.

This will help:

- *Maintain data quality by ensuring that MEDIN requirements are met, but make it easier for users to share their data;*
- *Encourage more organisations and individuals to submit their data currently stored on publicly inaccessible hard-drives or servers into the data network;*
- *Increase the volumes of data made available in standard formats;*
- *Wider application of FAIR data principles and facilitate the integration potential of marine data from different disciplines and sectors (including private sector).*

RECOMMENDATION 15 – **Plan for and fund the management and sharing of all new data being collected**

Funding providers should stipulate that a requirement of funding will be the development and execution of a data management plan that assures datasets are provided in accordance with FAIR data principles and shared within a timely manner, following embargo periods [e.g., a requirement for research projects receiving public funds to share data that they generate with MEDIN, via affiliated data flows, to contribute to the Scottish / UK marine evidence base].

Further collaboration and discussion with the organisations that fund the collection of data will be fundamental to achieving this.

This links with: RECOMMENDATION 3 – affiliated data submission routes; and RECOMMENDATION 18 – developing guidance on optimal data submission pathways.

RECOMMENDATION 16 – **Develop proactive engagement with data custodian stakeholders who weren't fully involved in the review**

Further targeted engagement should be undertaken with stakeholders (including eNGOs, commercial sector) in an endeavour to increase the flow of biodiversity records into the marine data infrastructure.

This further engagement would support and facilitate the wider cultural step-change required to increase data sharing and availability.

RECOMMENDATION 17 – **Develop simplified user interfaces onto repositories to support wider data submission**

The development of simplified user interfaces onto repositories should be encouraged to support the submission of data by citizen science initiatives.

This links with RECOMMENDATION 2 – simplifying the requirements for submitting data into DASSH.

RECOMMENDATION 18 – **Develop guidance on optimal data submission pathways**

Guidance should be developed with stakeholders to clarify the optimal pathways for submitting biodiversity records into Scottish / UK marine data repositories, in accordance with FAIR data principles.

For example, developing guidance with academic researchers would aim to provide reassurance to the academic sector that their data submitted at a Scottish / UK level would flow to the appropriate EU / international portals (with likely timelines); data flow into UK infrastructure in the first instance would make research data more readily available for use in a Scottish/UK policy context.

This links to RECOMMENDATION 13 – infrastructure to support citizen science record submission.

RECOMMENDATION 19 – **Invest in data engineers and allocate resource for system decommissioning**

Data engineers should be funded to input 'loose' data stored in file storage on networked drives into systems; complimented by short-term resource (monetary and/or effort) made available to government organisations to enable legacy data management system decommissioning so that and the benefits of cloud-based system technology can be fully adopted.

This would reduce technical debt and longer-term data management staff resource requirements associated with non-automated workflows and duplication of effort, by facilitating full adoption of automated workflows and the benefits of new cloud-based system technology.

RECOMMENDATION 20 – **Provision of biodiversity records collected under licence or for consent into the MEDIN data archive centre network**

It should be a statutory requirement for records collected by commercial developers through the licensing and consenting system to be provided, via affiliated data submission routes (i.e., established databases), into the UK MEDIN archive DASSH. This would enable onward publication to the NBN Atlas and other international portals.

This links to RECOMMENDATION 3 – primacy of affiliated data submission routes.

RECOMMENDATION 21 – **Maintain data version control through encouraging active custodianship**

Data custodians should perform checks to determine whether the version of data in portals is true to source, and ensure that portals harvest updated data, in addition to re-archiving.

Re-harvesting of data by Data Archive Centres (e.g., DASSH from MRO), either periodically or on request following active management of data, would help ensure that data are up-to-date and robust throughout the data network.

RECOMMENDATION 22 – **Optimise re-use of data through adherence with FAIR Data Principles**

All organisations should champion open data and FAIR data principles: use of Open Government Licensing for all data commissioned by public bodies and Creative Commons (by attribution) (CC-BY) licences for industry, eNGO volunteer recording and academia should be encouraged; clear licensing conditions; and easy to access descriptions of the dataset (metadata).

The generation, management, collation and sharing of data should be based on FAIR Data Principles to make marine species and habitat data in Scotland Findable, Accessible, Interoperable and Reusable (FAIR) throughout the data flow network.

RECOMMENDATION 23 – **Develop existing portal infrastructure to support efficient searching, data display and dataset collation**

DASSH, Marine Scotland and the NBN Trust should prioritise investigation and requirements gathering to fully understand DASSH's species mapper, NMPi portal and the NBN Atlas' existing and future customer needs against the current and planned work to respective portal interfaces; i.e., what stakeholders need access to and how, and where the highest value lies for each customer*.

*There are a wide range of user groups with different needs / expectations.

This could include:

- A discovery phase, prior to undertaking user needs research, to clearly understand and articulate each systems niche and its purpose within the data network to avoid duplication. See <u>Annex D</u> for description of existing key system purposes relevant to Scottish data.
- Understanding what is working well and what can be improved in the existing user interfaces of these platforms, the mapping tools functionality, download services and use of APIs. This would help to deliver the infrastructure required to facilitate efficient searching, harvesting and collation of records. This links to <u>RECOMMENDATION 10</u> – infrastructure to support habitat records.
- Enabling the querying, visualisation and download of multi-disciplinary datasets for use in end-user systems, via cross-DAC re-aggregation of data.

RECOMMENDATION 24 – Embed marine expertise in, and interoperability of, the national and regional (LERC) hubs infrastructure in Scotland

- The NBN Trust should require, where possible, marine ecological expertise and/or marine data management expertise in at least one of the role holders recruited for the SBIF Better Biodiversity Data (BBD) Project, i.e., for one of the roles to be located in the National Hub for Scotland being established by that project.
- The infrastructure of National and Regional Hubs (i.e., the Scottish LERC infrastructure) should be scoped and developed to enable interoperability with the existing and developing marine data infrastructure; including Marine Recorder Online, the JCDP, and the MEDIN biodiversity data archive centre DASSH, so that data are made openly available for others to reuse under licence terms. This is in alignment with MEDIN's ethos of FAIR data.
- There is a need to further tease out the relevance to the LERC model (which is based on charging developers for 'value added services') to marine biodiversity data; if marine data submitted to LERCs flows efficiently into MEDIN DACs there would likely be little call on LERCs for 'value added services' as most enquiries about marine data would likely end up with e.g., DASSH (coastal data is possibly an exception).

This would enable incorporation of marine species and habitat data products into the 'value-added service' available, e.g., to industry and local authorities, through the Scottish Hub for use in coastal (terrestrial / freshwater-marine interface) planning and development. It would also facilitate the integration of existing marine data management infrastructure with the future Hub infrastructure to ensure that any marine data submitted into the LERCs flows into the wider marine data network. This links with <u>RECOMMENDATION 10</u> – infrastructure to support habitat records.

RECOMMENDATION 25 – Ensure future governance of marine data management in Scotland

A Scottish / UK advisory group should be formed to facilitate continued cross-sector stakeholder engagement and collaboration and guide implementation of the recommendations.

The group's role should involve:

- Guiding the development of an 'Implementation Plan'; this should follow an agile approach, focussing on priority areas and areas of highest value/benefit first.
- Within the 'Implementation Plan', develop a benefits dependency network diagram for the marine community, identifying the case for change:
    - Drivers of change
    - Change objectives
    - Benefits of change
    - Business changes needed
- Oversee / monitor and provide leadership to progress, find solutions to, and implement the Review's recommendations.
- Ongoing and iterative collaboration between stakeholders.

Annex G – Recommendations dependency matrix

The number in the matrix corresponds to the recommendation in Annex F. Example of how the matrix works: the success of recommendation 1 is dependent on recommendation 2 and recommendation 19.

-	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	X	-	X	-	-	-	-	-	-	-	-	-	-	X	-	X	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-
14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	-	X	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-

-	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-