JNCC Report
No. 670

Review of biodiversity data use in the Country Nature Conservation Bodies

Hassall, I., Cheffings, C., Robinson, A. & Robinson, P.

October 2020

© JNCC, Peterborough 2020

ISSN 0963 8091

For further information please contact:
Joint Nature Conservation Committee
Monkstone House
City Road
Peterborough PE1 1JY
https://jncc.gov.uk/

This report should be cited as:
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.

Acknowledgments:
JNCC would like to thank all interviewees contacted as part of this project for their invaluable contributions to this report. JNCC would like to thank the Department of Agriculture, Environment and Rural Affairs, Natural England, Natural Resources Wales, NatureScot, National Museums Northern Ireland and Northern Ireland Environment Agency.

EQA:
This report is compliant with JNCC's Evidence Quality Assurance Policy
https://jncc.gov.uk/about-jncc/corporate-information/evidence-quality-assurance/

Summary

Background and rationale

The UK possesses a wealth of long-term, taxonomically diverse biological data spanning the entire nation. These data provide an invaluable and powerful tool for nature conservation. Access to biological data enables protection of species and habitats, effective land management and national trend assessments to ensure the sustainability of the UK's natural resources. Importantly, biological data allow land managers to assess the effectiveness of conservation schemes, promoting efficient use of resources. Applications of biological data have become increasingly important as the world faces the current biodiversity crisis and global climate change. Biological data allows researchers to understand environmental response to pressures, including anthropogenic stressors and underlying climate change, and underpins significant policy decisions to avoid irreversible damage.

The UK Country Nature Conservation Bodies (CNCBs) – Natural England (NE), Natural Resources Wales (NRW), NatureScot, and Northern Ireland Environment Agency (NIEA) – aspire to be "evidence-led" in environmental management, aligning policies and decisions with a strong evidence base. The CNCBs rely heavily on access to high-quality biological data to fulfil their key statutory functions. Recent changes in biodiversity data workflows have created barriers to data access and confusion among the recording community. To alleviate these difficulties and secure the continued application of biological data in the UK, it is crucial to build an understanding of current usage, data requirements and potential future uses.

This report provides a comprehensive understanding of current data use cases across the CNCBs, gathering information from a wide cross-section of data users, from data collectors through to policy advisors.

Aims

This project aims to:

- Identify current data uses, including the information gained, data sources, process of using data and data requirements;
- Identify limitations to current workflows;
- Outline suggested improvements to data workflows;
- Identify future research priorities across the CNCBs

Methods

A total of 61 interviews were conducted across CNCB staff, giving a wide cross-section of data users. The breakdown of interviewees according to the scale of their work is summarised in Table 1. Interview responses were summarised, and emerging themes were extracted. Responses were coded according to themes to provide rankings of limitations, improvements and future priorities.

Table 1. Breakdown of interviewees across NatureScot, NIEA, NE and NRW. Numbers of interviewees working on national and local scales and within terrestrial and marine teams are given.

	Total	National	Local	Terrestrial	Marine
NatureScot	19	16	3	17	2
NE	29	16	13	25	4
NIEA	6	6		3	3
NRW	7	7		6	1

Main findings

The main current use cases identified are:

1. **Casework**, including responding to planning applications, informing planning policy and providing advice to remediate site damage. The main data source is Environmental Impact Assessment (EIA) reports submitted with planning applications. Local teams and species/habitat specialists provide comments on these reports to ensure conclusions are valid and provide advice on appropriate mitigation measures.
2. **Designating and assessing the condition of protected sites**. CNCBs are required to assess the condition of designated features on protected sites on a continuous basis. The main data source is specific internal surveys focusing on designated features at each site. Surveyors assess the condition of each feature and these data are stored internally. Reports are generated for site managers and summary statistics of site condition are provided to JNCC to feed into Habitats Directive reporting and UK Indicators.
3. **Reporting**, both international and domestic. Examples of international biodiversity reporting for the UK currently include the EU Habitats Directive, EU Birds Directive, and the Convention on Biological Diversity (CBD). The CNCBs collate all available country level data concerning the conservation status of habitats and species listed under the Habitats Directive and provide these status assessments to JNCC to compile a UK-wide Article 17 report. This has been a major focus for all CNCBs in recent years as a main application of biological data. Domestic reports include NRW's State of Natural Resources Report and the State of Nature reports.
4. **Agri-environment scheme development and assessment**. Biological data are applied when designing schemes, providing agreement level assessments, and evaluating scheme impacts on a national scale. All available data are incorporated when designing schemes, including the best available habitat data, site-specific data and information from discussions with land managers. When assessing scheme impact, specific survey data are compared with existing national monitoring scheme data to provide counterfactual information. National evaluation projects incorporate data from a wide range of sources, including internal survey data, national monitoring schemes, data from partner organisations and verbal communication with land managers. Analysis for agreement level and national assessments is contracted out and summary reports are produced to inform internal agri-environment scheme strategies.

Detailed case studies cover a wide range of applications, including priority habitat mapping, identifying areas for conservation and developing a new approach to protected species licencing.

The five main identified limitations to current workflows include:

1. **Data access**. Interviewees estimated that only 50% of all available data are currently accessible. Limited data sharing to the National Biodiversity Network (NBN) Atlas and other repositories is due to concerns around commercial exploitation of data and other political issues. Staff feel CNCB data applications are not harnessing as much data as possible, and that EIA and historic datasets should be mobilised for use.
2. **Data flows**. Current data flows have too many submission portals and routes, causing confusion around "who is collecting what from where". Data are not reaching their ultimate destinations and flows are subject to significant time lags.
3. **Data coverage**. Important taxonomic and spatial gaps were identified, including areas outside protected sites boundaries.

4. **Data quality**. Interviewees highlighted the ambition to capture more citizen science data, but emphasised the importance of verification to improve their confidence using these data in decision-making.
5. **Internal data management**. A lack of clarity on where to store data internally, low awareness of existing datasets and complicated data access systems were raised as issues. Interviewees unanimously felt that internal data were not being used in as many applications as possible.

The five main suggested improvements include:

1. **Connected, simple workflows**. Data could flow through a central hub or there could be a central metadata catalogue to flag the existence of datasets, signpost data locations and provide a point of contact. Interviewees emphasised improved workflows may not consist of a single repository, but that several platforms could be employed to fulfil different purposes. Each repository and data management system would form a node within the framework and would need to connect and communicate with all other nodes.
2. **Improved internal data management**. The need for a searchable metadata catalogue was highlighted to ensure maximum data usage. Web server hosting was suggested to improve data sharing and storage. A main priority would be to centrally manage local data to improve access in area teams.
3. **Improved data access**. Building an understanding of the barriers to data sharing, reforming the Local Environmental Records Centre (LERC) framework and improving relationships, providing access to EIA and historic data and sharing data across Defra were some of the possible solutions to data access issues.
4. **Communication across recording community**. Maintaining volunteer engagement through innovative approaches, such as creating a single hub to provide information about recording schemes, and the provision of feedback and training opportunities would help to sustain citizen science data collection. Continued and transparent communication across the recording community is required to foster trust and build relationships with recording groups and individuals.
5. **Harness additional data**. Mobilising EIA and historic data and capturing more citizen science data were identified as priorities.

The main future data use cases were:

1. **Remote sensing**. Applications included high-resolution habitat mapping, abrupt change detection and wildfire/muirburn time-series data analysis. The requirement for improved collection of robust ground-truth data was highlighted.
2. **eDNA**. Currently widely used in freshwater monitoring, but ambition to expand to terrestrial and marine environments.
3. **Themes linked to Government environment policies**. The main themes mentioned were Nature Recovery Networks, Biodiversity Net Gain and applying the Natural Capital Approach. Other areas of interest included peatland restoration, ecosystem connectivity, ecosystem resilience, monitoring invasive non-native species and identifying suitable areas for tree planting.
4. **Increased data collection** to fill taxonomic and spatial gaps and expand applications to examine broad impacts, such as air pollution.
5. **Artificial Intelligence (AI)**. Applications include automatic species identification and counting large populations of seals and seabirds from high-resolution drone or satellite imagery.

Implications

The report provides a thorough understanding of current data usage across the CNCBs and offers clarity on data requirements and limitations to current workflows. Possible improvements and changes required to facilitate research priorities could inform future development in biological data infrastructure. The comprehensive interview process provides detailed descriptions of data access and analysis processes, and raises problems encountered at every stage of data use. The results of this report will help with planning future investment in biological recording.

Contents

1 Introduction

1.1 Biological recording in the UK

Naturalists have been recording wildlife in the UK for centuries. Environmental legislation protecting species of interest dates back to the late 19th century and has become increasingly important since the establishment of the Nature Conservancy in 1949 (Department of the Environment, 1995). The UK is fortunate in holding a wealth of current and historic biological records covering a wide range of taxa on a national scale. The longevity of these datasets is remarkable. The water beetle recording scheme for Britain and Ireland is over 100 years old, for example (Foster, 2015). Biological recording has continued to grow in the UK, and it is now estimated that over 70,000 people are involved annually (Pocock et al., 2015). The contributions of volunteer recorders are incredibly valuable. The Department of the Environment (1995) estimated that 70% of the recording community are volunteers. There have been several descriptions of past and present biological recording in the UK, see (Department of the Environment, 1995; Pescott et al., 2015; Pocock et al., 2015; Roy et al., 2012). Technological developments have increased data volumes and transformed biological recording opportunities (August et al., 2015).

The UK's vast quantities of biodiversity data present many opportunities for nature conservation. Biological data are invaluable in reporting trends, informing land management decisions and protecting wildlife. The UK Country Nature Conservation Bodies (CNCBs) aim to be "evidence-led" in all policymaking and research, which highlights the need for robust, accurate datasets.

1.2 Background and rationale

Biological recording infrastructure and data flows have undergone significant changes over the last decade, which has led to confusion within the recording community. There is the appetite for discussions around improving the current framework of data use, from collection and submission, to access and application. It is widely recognised that current data flows could be simplified and that there are hurdles to overcome with data sharing, spatial resolution and coverage. There are also concerns about the sustainability of UK biological recording in light of diminishing resources. The 2018 Scottish Biodiversity Information Forum (SBIF) "Review of the Biological Recording Infrastructure in Scotland" investigated current limitations of the data workflow in Scotland, and culminated in recommendations to achieve five main outcomes, including transformed data flows, service provision, governance and funding by 2025 (Wilson et al. 2018). This review demonstrated many of the problems arising from current data workflows, and highlighted the need for further discussion around improvements to biological recording infrastructure across the rest of the UK.

It is widely accepted that the complications arising from current data workflows are hindering efficient use of the large volume of biological data that the UK possesses. In order to determine the best way forward, it is crucial to investigate current biological data use across the UK CNCBs and potential future applications. This report outlines common biological data use cases across the CNCBs, looking in greater detail at a number of case studies, and describes the limitations and possible improvements to the current data workflows identified by CNCB representatives. The results of this report will highlight current and future research priorities across the UK and will identify possible measures to streamline and improve data workflows.

1.3 Scope

JNCC have focussed on the CNCBs – Natural England (NE), Natural Resources Wales (NRW), NatureScot, and Northern Ireland Environment Agency (NIEA) – for the scope of this report. JNCC have prioritised gaining a representative view of biological data use across the CNCBs and have included a wide cross-section of data users, from data collectors and analysts through to specialist advisors and the users of interpreted products. It was also important to include members of staff working at both local and national scales and in both terrestrial and marine environments. JNCC collected information on the use of biological data from all sources to gain a full understanding of how data are currently accessed.

1.4 Project objectives

This project aims to gain a clear understanding of current biodiversity data usage across the CNCBs, the reasons for these applications, and explore future opportunities. Objectives include:

1. Identify current data use cases;
2. Discuss limitations to current workflows;
3. Outline possible improvements and opportunities for different ways of working;
4. Explore future research priorities.

2 Methods

2.1 Interviews

Representatives from each CNCB provided contact details for members of staff across each organisation. A wide cross-section of biological data users were contacted, including data managers, to give an understanding of internal data flows, and staff working at both a national and local scale were involved to ensure a representative view was attained. Note that interviewees representing Northern Ireland comprised NIEA, Department of Agriculture, Environment and Rural Affairs (DAERA) and the Centre for Environmental Data and Recording (CEDaR)[1], which is the LERC for Northern Ireland, managed by National Museums NI.

A total of 61 interview responses were collated throughout July-August 2020. Table 1 shows the breakdown of responses by agency and scale and Table 2 provides a breakdown by role type.

Table 1. Breakdown of interviewees across NatureScot, NIEA, NE and NRW. Numbers of interviewees working on national and local scales and within terrestrial and marine teams are given.

	Total	National	Local	Terrestrial	Marine
NatureScot	19	16	3	17	2
NE	29	16	13	25	4
NIEA	6	6		3	3
NRW	7	7		6	1

Table 2. Breakdown of interviewees according to role type.

	Data Managers	Data Collectors	Analysts	Advisors	Strategists	Project Managers
NatureScot	2		2	7	4	4
NE	4	6	2	15	2	4
NIEA	3	1	3			
NRW	3		1	1	2	

2.2 Analysis

Notes were taken during interview discussions and were summarised to identify key themes. Data use case examples and case studies were examined in depth, detailing the data source, scale, resolution, update frequency, access, analysis and outputs for each case. Identified limitations, suggestions for improvements and future data applications were collated from all interviewees. Interview responses to a selection of topics were coded according to the main themes in Microsoft Excel to establish which ideas were raised most frequently.

[1] https://www.nmni.com/CEDaR/CEDaR-Centre-for-Environmental-Data-and-Recording.aspx

3 Results

3.1 Information gained from biological data

Interview responses identified the main areas of information gained from biological data applications. These areas are listed in order of how often each was mentioned:

1. Casework, including responding to planning applications and advice for site remediation work;
2. Designating and assessing the condition of protected sites;
3. Reporting, including international requirements, such as the European Commission Habitats Directive[2] and Water Framework Directive[3], and domestic reports, such as NRW's State of Natural Resources Report (SoNaRR)[4];
4. Monitoring species and habitat status and changes over time;
5. Creating habitat maps;
6. Assessing management scheme effectiveness, including agri-environment schemes;
7. Informing land management decisions;
8. Identifying areas of opportunity for conservation and restoration;
9. Specific species or habitat guidance and advice, including notification guidelines.

These areas of information allow the CNCBs to fulfil statutory duties, such as reporting and condition assessment requirements, and reflect the current focus on protected sites across the agencies. When asked about information ideally gained from biological data, many mentioned that a higher coverage of data outside protected site boundaries would enable users to draw conclusions about the entire landscape, rather than selected priority areas. Other suggestions included (in no particular order):

- Filling taxonomic gaps, particularly fungi, invertebrates, lichens and bryophytes, in data coverage;
- Filling spatial gaps in coverage, including upland areas and the intertidal and shallow sub-tidal zones. For example, Ulster Wildlife Trust launched a citizen science project to gather data in the intertidal area;
- Collecting more data on common species to better understand national trends. For example, sparrows are a more effective indicator species to assess climate change impacts than many rare species;
- The ability to link species and habitat data to abiotic and socioeconomic contextual data layers, such as water quality, atmospheric deposition, climatic variables, wellbeing, access and recreation, was also identified as a priority;
- Many felt the information they would ideally like to access was already being collected, and that mobilising historic and EIA data would enhance current data applications;
- Conducting a national survey of habitat extent and condition;
- Collecting population information in addition to presence/absence records, which would provide a better understanding of species status;
- Broadening the scope of data collection to include more mobile species and species functionally linked to other protected species;
- Assessing species in different ways, for example using species as an indication of ecosystem function and resilience or as a representation of functional groups;

[2] https://ec.europa.eu/environment/nature/legislation/habitatsdirective/index_en.htm
[3] https://ec.europa.eu/environment/water/water-framework/index_en.html
[4] https://naturalresources.wales/evidence-and-data/research-and-reports/the-state-of-natural-resources-report-assessment-of-the-sustainable-management-of-natural-resources/?lang=en

- Achieve almost real-time monitoring programmes by utilising EO methods and continuous field observations. For example, monitoring peat bog "breathing" using satellite images and overlaying results with species data would confirm the return of rare species as a result of successful peat restoration;
- Exploring novel techniques to assess genetic diversity and link results to underlying habitat information.

3.2 Common use cases

The following common use cases were identified by interviewees from all agencies. The data sources, access, scale and resolution and process of the applications are described in each case. The use cases and data sources are not in any order, but if there is a main data source this is clarified in the text.

3.2.1 Reporting

'Reporting' was a wide-ranging theme, including statutory and non-statutory, international and domestic requirements, as well as more local reports (covered under some of the other use cases below). Examples of key international biodiversity reporting commitments for the UK currently include the reporting under the EU Habitats Directive, EU Birds Directive[5], and the Convention on Biological Diversity (CBD). The CNCBs have a key role in collating country level data concerning the conservation status of habitats and species listed under the Habitats Directive, and providing these status assessments to JNCC to compile a UK-wide Article 17 report. This has been a major focus for all CNCBs in recent years as a main application of biological data. In support of the CBD and country level biodiversity strategies, some countries produce country-level biodiversity indicators (which are parallel to the UK Biodiversity Indicators hosted by JNCC), for example the England Biodiversity indicators[6], which require regular data inputs to update. In Wales, a key reporting requirement is the State of Natural Resources Reporting (SoNaRR) and associated indicators being developed. An example of non-statutory reporting in partnership with a coalition of organisations (predominantly NGOs), is CNCBs involvement in the State of Nature report which regularly reports trends on a very wide range of taxa.

Reasons behind data use:
The wide range of reporting requirements leads to different types and levels of data required. Even under a single reporting requirement there may be data required for different purposes. For example, Habitats Directive reporting requires species distribution data as well as population trend data. Data are accessed from a variety of sources in different formats It was noted that data requirements are often not prescriptive, but CNCBs generally use the best data available, for example using the most up-to-date data. Generally for mapping species distributions CNCBs are keen to use the highest resolution data available to make reports more accurate, but note some reports are more focussed on national species trends, where repeat sampling is of more relevance than sample resolution.

Data sources:
- Internal survey data, including protected site condition monitoring data;
- Third-party data provided by partner organisations, such as Forest Research (FR), Scottish Environmental Protection Agency (SEPA), Environment Agency (EA);
- Recording scheme data provided by NGOs and research bodies, such as the British Trust for Ornithology (BTO), and Bat Conservation Trust (BCT) and UK Centre for

[5] It is likely that these reporting requirements will still be required in some form even now the UK has left the EU.
[6] https://www.gov.uk/government/statistics/england-biodiversity-indicators

Ecology and Hydrology (UKCEH) often on behalf of a JNCC partnership monitoring scheme. Examples of datasets include the BTO/RSPB/JNCC Wetland Bird Survey (WeBS)[7], the BTO/JNCC/RSPB Breeding Bird Survey (BBS)[8] and the UK Butterfly Monitoring Scheme[9].

- National Biodiversity Network (NBN) Atlas data. The extent of the use of NBN Atlas data varied across roles and CNCBs;
- Local Environmental Record Centre (LERC) data;
- Local biological recording groups, such as badger and bat recording groups;
- Survey data completed for specific projects;
- Results from agri-environment scheme monitoring. For example, data collected under the Glastir Monitoring and Evaluation Programme (GMEP) feeds into NRW's SoNaRR reports, see Alison et al. (2020).

Frequency: Dependent on the report. Every 6 years for Habitats Directive reporting, domestic reports vary in frequency, for example SoNaRR reports are produced every 4 years, and the State of Nature reports every 3 years. Some biodiversity indicators are updated annually.

Scale: National

Resolution: Not prescriptive, reports tend to use the highest resolution data available

Access:

- NBN Atlas openly accessible;
- LERC data accessed either through agreement or by purchasing the data by request;
- National and local recording schemes approached directly for survey data;
- Internal survey data stored on internal systems.

Process:

- For Article 17 reporting, all data are gathered by agency staff, assessed and supplied to JNCC, who compile the UK-wide report;
- For many domestic reports, data are gathered and assessed by species and habitat specialists, who focus on individual topics.
- High level synthesis summarises the information and incorporates other broad-scale factors, for instance cultural and socioeconomic factors and the estimated global impact of national natural resource exploitation.
- Surveyors (including contractors) often analyse their own data and provide interpreted information to the CNCBs, which feeds into indicator reports

Outputs:

- International reports such as the Article 17 reports periodically submitted to the EU are openly available online;
- Domestic reports are openly available online.

3.2.2 Casework and responding to planning applications

Responding to proposed development applications, influencing planning policy and carrying out assessments on damaged sites were mentioned by local and national staff as a main data use case. Casework chiefly involved providing feedback and suggesting mitigation

[7] https://www.bto.org/our-science/projects/wetland-bird-survey
[8] https://www.bto.org/our-science/projects/bbs
[9] https://www.ukbms.org/

measures in response to supporting information submitted with planning applications, such as EIA surveys conducted by contractors. In some cases additional data from various sources were collated to assess the validity of conclusions and make recommendations to avoid damage to species and habitats.

Reasons behind data use:
Site-specific and high-resolution data are required to provide informed land management decisions. The vast majority of casework involves providing comments on ecological reports, including EIA survey data, submitted as part of a planning application. It was recognised that these datasets meet the requirements for providing advice to developers and land managers, although there were concerns that this workflow relies on consultant information rather than independent data sources. NBN Atlas data were often used to examine previous records at a site, but it was noted that these data are often unsuitable for casework, due to the coarse resolution. Interviewees also mentioned that the time lag involved in updating NBN Atlas data, estimated to be at least 6 months, rendered these records unsuitable for responding to reactive casework, as these applications require current, fine resolution data. Data with a resolution of $10m^2$ was cited as ideal, but it was conceded that this is not always available and resolutions $<100m^2$ were acceptable.

Data sources:
- The majority of casework was based solely on information provided with applications. In most cases this consists of ecological reports and information, such as species and Phase 1 habitat reports, but some larger projects require EIA survey data. Surveys were conducted by consultants contracted by the developer;
- Internal survey data, including protected site notification and condition assessment surveys, agri-environment scheme data, priority habitats inventory;
- NGO and JNCC partnership data, for example BTO/RSPB/JNCC Wetland Bird Survey (WeBS) dataset and Wildlife Trust survey data;
- NBN Atlas data used in background searches to show species present at site;
- LERC data;
- Local recording groups, such as badger and bat recording groups.

Frequency: ad-hoc

Scale: Local, site level

Resolution: As fine as possible, as some sites only 0.5ha. A resolution of $<100m^2$ is required for site-level assessments, while $10m^2$ was cited as an ideal resolution.

Access:
- Supporting ecological information and EIA survey data provided with planning application;
- Data requests to partners, local recording groups, LERCs, NGOs. Local area teams often accessed data through relationships with providers rather than official agreements;
- Internal protected site survey data accessible through internal systems;
- Internal project specific survey data accessed opportunistically, as staff are reliant on knowledge of the survey being conducted. The existence of survey data was uncovered through suggestions from other members of staff

Process:
- Supporting information are submitted as part of the planning application. Local area teams, and in some cases national species/habitat specialists, assess the impacts of development to species/habitats based on information collated by contractors,

including EIA data in some cases, and additional sources. Submitted data are assessed for deficiencies and the validity of conclusions is checked;
- Assessments are based on “finding a narrative” in the available data, rather than specific analyses;
- Area staff incorporate comments based on a range of sources, including input from species and habitat specialist staff, to create an official agency response.
- In some cases, feedback on planning applications and EIA findings goes to the contractors who then take the response on board before local planning authorities (LPAs) make the final decision;
- CNCBs also liaise directly with LPAs to respond to planning frameworks and act as statutory consultees for planning applications that impact protected sites. CNCBs advise LPAs and developers on EIA requirements and provide formal responses to planning applications directly to LPAs;
- When assessing damage to a site, local area teams are reliant on “tip-offs” from contacts or members of the public about site damage. Local teams then investigate the claims and advise the land manager accordingly.
- Internal survey data gathered as part of a casework assessment are shared internally.

Outputs:
- Advice on mitigation, any concerns or areas that need further consideration are provided as comments on planning application reports. Specific reports or Habitats Regulations Appraisal (HRA) forms are only produced if the case is under scrutiny, or development will affect a Natura site;
- An example would be Coul Links golf course, which affected a biodiverse site of national importance, whereby NatureScot local teams were able to advise development;
- Internal survey data gathered as part of casework are available internally for future use and via LERCs.

3.2.3 Condition assessment of protected sites

Each CNCB has commitments to monitor the status of designated features within protected site boundaries. These surveys are done in-house or contracted out on a rotational basis. The aim of these surveys is to assess the condition of features and assemblages on protected sites, to identify any changes over time and determine the effectiveness of management practices. Surveys also discover new notifiable assemblages, which can then be added as a designated feature. These data can inform land management decisions, identify suitable areas for restoration work and provide estimates for the feasibility of such schemes.

Reasons behind data use:
Assessing the condition of protected sites requires a bespoke monitoring programme, as each site has different designated features. The monitoring must adhere to Common Standards Monitoring protocols. Assessments are made on a continuous rotational basis to inform site management practices and ensure protected habitats and species remain in favourable condition. Additional datasets are used in assessments to provide historic data at certain sites, but the amount of historic data available is variable and availability depends on previous recorders and curating. For example, NE ecologists used a data archive dating back to the 1700s to identify the change trajectory of species and habitats on a Ramsar site in the West Midlands. The NBN Atlas is also used to look at historic records and “get a feel” for the site. Records on the NBN Atlas put site distributions into a national context and help to guide survey effort. Internal survey data are collected using standard methodologies and

are at high capture resolution, while data from additional sources is accessed at the highest possible resolution.

Data sources:

- Internal site condition monitoring surveys;
- Other internal data covering the site boundary, including scanned paper files, historic archives, previous research projects, best available habitat data;
- National recording scheme data;
- LERCs data;
- NBN Atlas data;
- Botanical Society of Britain and Ireland (BSBI) data;
- Verbal contact with local county recorders/naturalists.

Frequency: Countries have aimed for new data for sites every several years. However, the exact frequency is based on a risk assessment and varies according to the vulnerability of features and sites.

Scale: Local site level

Resolution: Dependent on the feature in question, size of the site and mobility and range of the species. For small sites, 25m^2 resolution is needed, while for larger sites and more mobile species, such as invertebrates and mammals a resolution of 1km^2 is suitable.

Access:

- Internal data such as specific condition assessment survey data and project-based survey data are accessible through internal processes;
- Historical data, for example internal paper files or previous surveys conducted by partner organisations, are accessed by links or directions from other members of staff. Finding relevant data is dependent on staff knowledge and details from other staff members. Gathering all data therefore takes a long time, up to 12 months in some cases;
- NBN Atlas data accessed online;
- BSBI Distribution Database accessed through agreement;
- LERC and third-party data provided either by agreement or by specific request.

Process:

- Prior to conducting a site assessment, agency staff gathered all internal data, including historic records and previous survey data, and any data from external sources to assess previous site condition;
- Survey protocols varied slightly across the CNCBs, but all mentioned adhering to Common Standards Monitoring protocols. Data are collected in different formats and using different habitat classification systems. Field surveys are either conducted by agency staff or by contractors, and data are converted to the correct format in-house;
- Data are collected to cover a list of notified features (species, habitat, geological, historic) and staff assess the condition of the feature based on these records;
- Pantheon[10] software is often used to analyse invertebrate data and provide an assessment of the status of invertebrate assemblages;
- Internal data management varies across CNCBs, but follows the same pathway of pre-processing survey data to required format and uploading into a central repository.

[10] https://www.brc.ac.uk/pantheon/

Outputs: Condition assessments on internal systems, more detailed reports to site managers. Summary statistics on protected site condition are reported to JNCC to feed into UK Habitats Directive reporting and producing the protected sites biodiversity indicator.

Future: It was recognised that the framework of condition assessment needed to be adapted to suit current biodiversity aims and meet requirements for wider data applications, such as Earth Observation (EO) ground-truthing.

3.2.4 Agri-environment schemes

Designing a framework of agri-environment schemes is becoming increasingly important as the UK leaves the EU. Biological data are used at the design and implementation stages and to assess the effectiveness of schemes over time.

Some agri-environment schemes are part of national evaluation projects assessing the success of schemes in terms of achieving specific UK-wide objectives, and these programmes incorporate data from a range of sources. Project aims include assessing the use of EO and eDNA in monitoring, landscape scale monitoring of a certain species/habitat, impact of schemes on specific species/habitat (for example lowland wetland birds), impact of schemes on air quality, scheme impact on historic/geological features, land sparing vs sharing approach and scheme impact on connectivity.

Reasons behind data use:
When designing agri-environment schemes all available data are incorporated, including the best available habitat data, existing survey data covering the site and qualitative information gathered from discussions with land managers. These data inform decisions around scheme eligibility and the most appropriate management options for each site. When assessing individual scheme impacts data are collected for areas within agri-environment schemes and compared to non-scheme areas, or counterfactual sites, using existing datasets. The changes attributed to agri-environment schemes are slow, small and subtle, and scheme assessments require a multi-objective approach. Therefore, it was noted that existing national datasets are often not suitable to assess individual scheme outcomes, and specific surveys are carried out at a site or land parcel level. In non-scheme areas (counterfactual sites), national monitoring scheme data can be used to provide comparisons. The required data resolution is high to allow fine-scale changes to be detected at a land parcel level National evaluation projects assessing the effectiveness of agri-environment schemes collectively utilise all available data, including scheme-specific survey data for all sites, national monitoring scheme data, existing internal data and LERC datasets. Methodologies and data requirements vary according to the specific evaluation focus.

Data sources:

- Scheme design:
    - Best available habitat data, such as the Habitat Map of Scotland[11], Priority Habitat Inventory[12], Centre for Ecology and Hydrology (CEH) Land Cover Map[13] and Phase 1 Habitat Map[14];
    - Existing survey data covering the site, including internal datasets and data provided by partners;

[11] https://www.environment.gov.scot/our-environment/habitats-and-species/habitat-map-of-scotland/
[12] https://data.gov.uk/dataset/4b6ddab7-6c0f-4407-946e-d6499f19fcde/priority-habitat-inventory-england
[13] https://www.ceh.ac.uk/ukceh-land-cover-maps
[14] https://lle.gov.wales/catalogue/item/TerrestrialPhase1HabitatSurvey/?lang=en

    - Discussions with land managers.
- Agreement level assessments:
    - Specific survey data collected at the start of scheme implementation and resurveyed after 5-10 years. Surveys are conducted using standard protocols to assess the status of features monitored for specific scheme objective. Different methodologies are used within the sampling framework and some surveys are contracted out;
    - Data for non-scheme area comparisons (counterfactual sites) are sourced from partner organisations, such as BTO, RSPB and national monitoring schemes. For example, BTO/RSPB/JNCC Wetland Bird Survey (WeBS) monitoring provides data on breeding waders on lowland wetland grassland for comparisons with land covered by schemes;
- National evaluation project data sources:
    - Internal data including option uptake, locations, management prescriptions, burn plans, popularity of options, members attitude, social aspects;
    - Internal species and habitat data from survey conducted at the start of schemes;
    - Internal protected sites data, such as number of sites within schemes, trends in condition;
    - Internal data from previous agri-environment projects;
    - BTO/JNCC/RSPB Breeding Bird Survey (BBS) data;
    - RSPB data for land sparing vs. land sharing project;
    - Met Office climate data;
    - Verbal contact with landowners to gain qualitative information;
    - LERCs data;
    - Historic England data for historic features projects.

Frequency: The frequency of data required is different for the different purposes linked to agri-environment schemes. When developing schemes all data available will be considered, for example for feeding into models informing scheme design. When implementing schemes, a baseline survey is conducted on sites at the start of an agreement, with a follow up survey at a later date, ideally after 5-10 years. For national evaluations of the effectiveness of schemes all data available will be considered for use.

Scale: National scale data is required for scheme development and national evaluation. Site or land parcel level for implementing schemes.

Resolution: For implanting schemes as fine as possible to cover small scale changes. Depends on the feature assessing, but <100m^2 was often cited as a requirement. Many survey protocols mentioned collecting data to 2m^2 resolution. Coarser data can be used for initial scheme development and national evaluation.

Access:

- Scheme design:
    - Habitat data and existing survey data covering the specific site are accessed internally;
    - Existing data for the site are requested from partner organisations and national monitoring schemes;
    - Qualitative and anecdotal information through discussions with land managers;
- Agreement level assessments:
    - Baseline and repeat survey data, previous project data and habitat maps are accessed through internal systems. For example, an agri-environment monitoring archive has been created for NE, but it was

    - mentioned that this is not currently accessed as widely as possible across the organisation;
    - Data for comparisons with non-scheme areas are accessed through direct requests;
- National scheme evaluation projects:
    - Internal data on scheme uptake, scheme locations, management measures, scheme specific baseline and repeat survey data accessed through in-house systems;
    - National data accessed by direct requests to partners and national monitoring schemes;
    - LERC datasets purchased or available through agreement;
    - Verbal discussions with land managers.

Process:

- Scheme design:
    - All available data are collated, and in-house modelling and assessments inform decisions on sites eligible for schemes and appropriate management options for each site.
- Agreement level assessment:
    - A multi-objective baseline survey is conducted at the start of scheme implementation, and repeat surveys are conducted 5-10 years later to give a “before and after” assessment;
    - Survey data and non-scheme data for counterfactual sites are provided to contractors. Contractors conduct the analysis for individual scheme effectiveness and provide a report and interpretation of results to inform future scheme implementation
- National scheme evaluation projects:
    - All available data are provided to contractors in various formats. Contractors analyse provided datasets according to specific project aim. For example, BTO/JNCC/RSPB Breeding Bird Survey (BBS) data are used to identify the effect of habitat provision schemes on species/habitats by sampling along scheme-uptake gradient;
    - Contractors produce project reports to inform national scheme strategies.

Outputs:

- Scheme design:
    - Site specific aims and details of management options to achieve scheme aims provided to land managers;
- Agreement level assessments:
    - Annual reports covering a range of scheme objectives are published giving the key findings from scheme assessments;
- National scheme evaluation projects:
    - Project reports covering specific aims are anonymised and shared with agency staff to shape future projects;
    - Synthesis across project reports informs internal management decisions and agri-environment scheme strategies.

Future:

- Align data requirements with national monitoring scheme protocols to maximise data collection;
- New technologies, such as eDNA and AI species identification, were recognised as useful tools but are not transferable to agri-environment schemes at present, as these data only provide information on the extent of habitat, but not condition, and only provide species presence information, not population changes;

- Possibility of mobilising survey data to the NBN Atlas, but it was noted that this would mean the data are presented at a coarse resolution, and only show presence/absence information of species. Habitat and abiotic information is currently not within the remit of the NBN Atlas;
- Ambition to make assessments on a wider scale, but this needs further consideration.

3.2.5 Habitat mapping

Creating up-to-date national habitat maps is a current priority. Different approaches of mapping habitat extent are being implemented across the CNCBs.

Reasons behind data use:
Interviewees indicated that all suitable data sources were incorporated into national habitat mapping projects, including national habitat surveys conducted by partner organisations. The habitat classification system and data used by each country needs to meet the requirements of the business needs and mapping approaches used by each CNCB. For example, NIEA and NE use an EO mapping approach, meaning that habitat classes used must be distinguishable in the EO models, and the spatial resolution of training data used needed to match the 10m resolution of the Sentinel imagery used. The NatureScot approach provides enough detail to meet stakeholder requirements, including assessing biodiversity net gain, flagging land use change, informing landscape management decisions, Natural Capital accounting and identifying opportunities for habitat restoration and connectivity. It has recently been announced that Space Intelligence, a satellite data company, has been commissioned to create a national map of Scotland using satellite imagery techniques[15]. NRW update their mapping of semi-natural habitat extent using EO data for their Indicator 43 reporting every few years. This is based on their national habitat map, which was originally created using EO approaches. It was recognised that updating the national maps on at least an annual basis would greatly enhance the application of these resources, particularly to work areas that require current data to inform land management decisions, assess damage or evaluate scheme effectiveness.

Data sources:
- NIEA are working with JNCC to create a national map using Earth observation techniques. Internal NVC surveys conducted as part of ASSI condition assessment surveys were combined with Copernicus Sentinel-1 and Sentinel-2 satellite imagery[14];
- NatureScot combined 750 NVC surveys translated to EUNIS level 4 and national surveys of saltmarsh, coastal vegetated shingle, sand dune and native woodland habitats;
- NRW combined upland vegetation field survey data from 1979-1989 with lowland survey data from 1987-1997 to create a comprehensive habitat map using Phase 1 habitat classifications. NRW then created an updated version of that Phase 1 map using Earth observation approaches;
- NE, supported by JNCC, are working on the “Living England” project to produce maps using Sentinel-1 and Sentinel-2 satellite imagery with training data from a variety of sources including the National Plant Monitoring Scheme[16].
- Marine benthic habitat maps are created using internal and commissioned video/sediment grab/dive survey data, data layers from academic research, survey data from partners (such as Cefas, IFCAs), citizen science data from schemes,

[15] https://www.scottish-enterprise-mediacentre.com/news/satellite-data-company-receives-extra-gbp-100-000-to-help-tackle-the-climate-emergency
[16] https://www.npms.org.uk/

including Seasearch, Shorething and Capturing Our Coast, Wildlife Trusts and LERCs.

Scale: National

Frequency: One-off with ambition to update annually

Resolution: All survey data at capture resolution, often 2m^2, with outputs scaled to 10m to match sentinel imagery for NIEA and NE outputs.

Access:

- Data from protected site monitoring surveys accessed internally;
- National surveys either accessed internally or provided by partners, such as RSPB, Wildlife Trusts and Scottish Forestry, through a sharing agreement.
- Citizen science data accessed for marine habitat maps are provided directly from Seasearch, and are accessed by request from other schemes and partners.

Process:

- NIEA: NVC survey data were used to train machine learning models. Habitat classes were defined by specialist ecologists and relevant columns were extracted and standardised. Only data showing habitats in favourable condition could be used as training data;
- Marine: Predictive modelling of benthic habitat locations, such as seagrass, is undertaken in Python and ArcGIS using available data as ground-truthing data;
- NatureScot: NVC surveys have been re-classified to EUNIS level 4 and will be combined in GIS to create a single layer of habitats.

Outputs: National habitat maps shared internally. Some national maps are currently shared openly, and there is the possibility of making restricted datasets open access in the future.

Future:

- Ambition to have continuously or annually updated habitat maps, but further development of the required technology is needed and a greater volume of robust ground-truthing data would need to be collected;
- Potential to have downloadable layers for each habitat openly accessible;
- NatureScot aim to implement large scale EUNIS level data collection efforts for specific sites and improved coverage of uplands areas. NatureScot aim to create an EO derived national map using robust training data.

3.2.6 Licensing protected species

Planning applications much demonstrate that European protected species[17] are not impacted by proposed developments and may require survey evidence for these species. Developers and contractors require licences to carry out surveys on protected species, and to take other action that impacts them and would be otherwise illegal. CNCBs are responsible for managing these licences and minimising impacts on species. Licences are provided on the condition that surveyors must return species data to assist further licensing decisions.

[17] Species under Annex IV of the EU Habitats Directive, also referred to as European Protected Species.

Reasons behind data use:
It was noted that staff were reliant on internal location information for existing and past licences and EIA data provided by consultants. EIA should incorporate available LERC data, as it is a requirement that consultants purchase data from LERCs as part of the assessment. It was mentioned that NBN Atlas data are rarely used in licensing applications due to the coarse resolution of data. Data returns are provided as a direct requisite of species licences, but are currently not always being applied as widely as possible. There is an ambition to mobilise these licence return data for use in other applications. Bat data are an example of where this is working well, as these data are currently being mobilised. Licence return data are required at a $10m^2$ resolution, as fine resolution is needed to make informed management decisions, assess development impacts and examine roost connectivity.

Data sources:
- Internal data including locations of existing and past licences;
- Licence return information;
- EIA consultant data;
- Accessible species monitoring scheme data, for example dormice survey data;

Scale: Local level, but national coverage

Resolution: Licence return data provided at $10m^2$ resolution, $<25m^2$ are useful on local scale

Frequency: Continuous

Access:
- Internal data management systems;
- EIA provided by consultants

Process:
- Licence requests submitted. Internal staff assess other licences and relevant data (e.g. bat roost locations) from the area and grant licence on specific conditions;
- Staff provide advice for consultants conducting EIA surveys.

Outputs:
- Licence return data, including the location of protected species, provided as a condition of the licence. The level of detail provided depends on the ecologist;
- Mitigation licences require more information, for example for bats, yearly updates of locations, species, numbers, and types of roost;
- Data are published on internal data systems.

3.2.7 Providing species protection guidelines and advice

Species specialists within CNCBs provide detailed guidance for best management practices to conserve protected species on a national scale. Many applications involve working on specific projects, such as species reintroductions, in partnership with agencies, universities and NGOs. For example, the Vincent Wildlife Trust (VWT) have been commissioned by NatureScot, NE and Forestry England to produce a recovery plan for pine martens in Britain, including using modelled outputs to determine locations with the best prospect for pine marten recovery. Projects with CNCB specialist input include Back from the Brink[18], focused

[18] https://naturebftb.co.uk/

on pine marten and long-eared bat recovery, wart-biter reintroductions on the South Downs, dormouse reintroductions[19] and EA-led monitoring of water voles.

Reasons for data use:
Specialist advice is provided on a project basis and guidance is mostly based on data collected as part of the project. National and local species monitoring schemes provide essential datasets. Species-specific surveys mentioned include the pine marten surveys as part of the Back from the Brink project led by VWT[20] and the VWT/EA otter survey data. The sustainability of these schemes was questioned and the need for secured funding to ensure the continuation of scheme surveys was highlighted. Required data resolution depends on the species of interest, as some species are more mobile than others. It is recognised that priority species are subject to change, especially with changes in "Red List" species, which will alter the focus of projects.

Data sources:
- National recording scheme data;
- NGOs and Wildlife Trusts data;
- Project specific survey data.

Scale: Both local and national

Resolution: As fine as possible, depends on the species

Frequency: Ad-hoc

Access:
- National recording schemes provide data through agreements;
- Wildlife Trusts provide data directly for projects.

Process: Data are gathered from various sources and assessments made in-house by specialist species advisors

Outputs:
- Advice to land managers about mitigation measures;
- Project-based advice including most beneficial locations for conservation and reintroduction efforts

3.2.8 Species status assessments

The national status of priority species and species assemblages is assessed over time, identifying trends in threatened taxa.

Reasons behind data use:
The majority of species status reviews use taxon-specific recording scheme data to assess the status on a national scale. NBN Atlas data are also incorporated, as these data have national coverage. Spatial and taxonomic gaps in data coverage were cited as limitations to this work, and specific surveys are occasionally commissioned to fill some of these gaps.

Data sources:
- Recording schemes data, including BTO, BCT, BSBI data;

[19] http://publications.naturalengland.org.uk/publication/5914082255306752?category=32020
[20] https://www.vwt.org.uk/projects-all/back-from-the-brink/

- NBN Atlas data;
- Survey data from partners, such as SEPA monitoring data;
- Third-party data, such as data from Royal Botanic Gardens;
- Specific survey data collected to fill gaps in data coverage.

Scale: National

Resolution: Varies across species, as some are more mobile. For example, 1km^2 can be used for invertebrates.

Frequency: Continuous

Access:
- NBN Atlas openly accessible;
- Requests for third-party data. Datasets provided under licence;
- Internal survey data on data management portals;
- BSBI DDb accessed under agreement.

Process: Specialist staff provide assessments of all data to produce status reviews

Outputs:
- Species status reports, detailing the species status and possible threats;
- Interpretation of results feeds into internal decision making for land management;
- Status assessments are provided to JNCC, who then submit to UKSI.

3.2.9 Marine Protected Area designations

The CNCBs are responsible for designating protected areas within inshore waters, while JNCC is responsible for offshore areas. The process of designating inshore Marine Protected Areas (MPAs) and Marine Conservation Zones (MCZs) involves rigorous data collation to identify priority habitats and species, geological and historic features that require protection from certain marine activities. The boundaries of MPAs, and restrictions within these areas, are dictated by the location of these features.

Reasons behind data use:
All data available were collated to ensure the proposed designations were based on all available evidence. National calls for data helped to improve data access. It was noted that robust evidence for MPA boundary decisions was required to justify decisions during stakeholder engagement discussions.

Data sources:
- JNCC Marine Recorder;
- NBN Atlas data;
- Internal survey data from projects and condition assessments;
- Data provided by partner organisations, such as Cefas;
- Wildlife Trusts data;
- Citizen science data from recording schemes, such as Seasearch, Shorething, Capturing Our Coast;
- National species-specific recording schemes provide rare species records;
- LERCs data.

Scale: National

Resolution: As fine as possible

Frequency: One-off process during designation stage

Access:

- Marine Recorder data snapshot provided by JNCC;
- NBN Atlas openly accessible;
- Partner data provided by agreement or request;
- Internal survey data internally managed;
- Approach national monitoring schemes directly;
- National call for data to gather data from citizen science schemes.

Process:

- Data collated from all sources and pre-processed internally to extract relevant information, standardise species and biotope names. Some interviewees estimated this process took approximately 4 months;
- In-house analyses involved a GIS Python model to provide a high/moderate/low confidence level for presence of priority habitat or species within proposed MPA boundary;
- Quality Assurance processes conducted in-house.

Outputs: Proposed MPA and MCZ boundaries for implementation

3.3 Project case studies

3.3.1 DAERA blue carbon habitat mapping

This work feeds into the Green Growth Strategy and entails mapping the extent of habitats with high blue carbon storage potential, including kelp, seagrass, brittle star beds among others.

Reasons behind data use:
Utilising all possible data sources ensures the mapping outputs are based on robust evidence and include all known areas of blue carbon habitat, which ultimately will improve management capabilities. Internal surveys target areas on a 6-year cycle, while data from the Agri-Food and Biosciences Institute (AFBI) provide coverage in other areas. Citizen science projects help to target areas of interest, such as seagrass and kelp.

Data sources:

- Commissioned surveys targeting areas of potential blue carbon habitats, including drone surveys of seagrass beds and salt marsh;
- Existing internal DAERA survey data;
- AFBI[21] survey data;
- Internal seabed habitat maps;
- EMODnet EUSeaMap[22];
- JNCC Marine Recorder[23];
- EIA data for proposed developments collated through CeDAR;
- Data from academic research, such as kelp layers[24] from The Institute of Zoology, Zoological Society London[25];

[21] https://www.afbini.gov.uk/
[22] https://www.emodnet-seabedhabitats.eu/about/euseamap-broad-scale-maps
[23] https://jncc.gov.uk/our-work/marine-recorder/
[24] https://macroalgalresearchgroup.com/project-reports/
[25] https://www.zsl.org/science/about-the-institute-of-zoology

- Ulster Wildlife Trust seashore surveys;
- Citizen science surveys, such as the Rathlin Island Dive Expedition[26] and Seasearch data.

Scale: National

Frequency: Currently one-off project basis

Resolution: As fine as possible

Access:
- Existing internal survey data stored in Marine Recorder format;
- Commissioned surveys provided directly in Marine Recorder format;
- Survey data from partners (AFBI) provided directly in MEDIN format;
- Seasearch provides data directly in Marine Recorder format;
- Knowledge of EIA survey by staff mentions, request survey data from consultants;
- Marine Recorder snapshot data from JNCC;
- EMODnet EUSeaMap online portal.

Process:
- Data management, Quality Assurance and pre-processing through CEDaR and in-house;
- Gathering all data sources and standardising formats, such as converting data to Marine Recorder and Marine Environment Monitoring and Assessment National Database (MERMAN)[27];
- Mapping analysis in-house using ArcPro.

Outputs: National maps of blue carbon habitats for use in policymaking

3.3.2 NE great crested newt (GCN) modelling tool

This project outlines a more strategic approach to licencing protected species. The impacts of proposed developments are assessed according to the location of GCN and the loss of potential GCN habitat. The work combined GCN distributions data, habitat and abiotic information to create a habitat suitability model. The results from this modelling approach were used to outline risk areas and create a strategic opportunity map to guide licence restrictions, steer development away from high-risk zones and determine the most beneficial location for compensation ponds. This approach to GCN licencing launched in Kent and Cheshire in 2019 and NE are now working with Local Planning Authorities (LPAs) to implement the scheme across England.

Reasons behind data use:
Responding to planning applications requires high-resolution, up-to-date data to ensure guidance and mitigation recommendations are accurate. To ensure suitable data were used for this project, a specific eDNA survey was conducted over three successive seasons across England, which met the data requirements for the modelling approach. The use of additional data sources was constrained by the resolution. Existing internal licence return data were recorded at 100m^2 resolution, whereas this project required the input data to match the 25m^2 resolution of the CEH Land Cover Map. Similarly, the GCN distribution data

[26] https://www.daera-ni.gov.uk/publications/rathlin-island-dive-expedition-2019-citizen-science-project
[27] https://www.bodc.ac.uk/projects/data_management/uk/merman/

obtained from LERCs varied in resolution and quality and most records were too coarse for use in this project. Only the records with $1m^2$ and $10m^2$ resolutions were suitable for use. Ideally the project team would like to incorporate the Living England map instead of using the 2015 Land Cover map, as this will provide updated habitat extent information. It was also noted that GCN population data would improve the project, but presence/absence points were the only data currently available.

Data sources:

- Stratographic eDNA surveys of counties across England conducted for 3 consecutive seasons;
- Internal licence return data;
- GCN distribution data from LERCs;
- Pond locations extracted from Ordnance Survey (OS) MasterMap;
- CEH Land Cover Map 2015[28] 25m raster layer;
- Cranfield University soil data, including pH levels, phosphorus levels, soil types;
- LPA proposed development areas.

Scale: County level with national coverage

Frequency: Ambition to update the model at least every 5 years and to include new data as often as possible, such as pond losses and gains in Living Pond layer.

Resolution:

- Area of interest is 250m radius of GCN habitats;
- All input data scaled to $25m^2$ resolution used for maps;
- Advice outputs scaled to $1km^2$ and $10km^2$ resolution in line with national guidance.

Access:

- Internal surveys and licence return internally stored;
- Approached LERCs directly to purchase data under licence;
- OS Mastermap data purchased and updated every few months;
- LPA proposed development data accessed by direct request.

Process:

- Pre-processing data layers included minor formatting of internal survey spreadsheet data and rasterizing the vector soils data and resampling to 25m resolution;
- Habitats of interest were extracted from the CEH Land Cover Map raster to create new layers including habitats suitable for GCN, for example areas with >50% arable land cover and areas within 250m radius of ponds;
- Input habitat data layers and GCN distribution points were fed into a Species Distribution Model (SDM) in R biomod2 package. Ensemble models were created using the GCN presence/absence points as training data and the model was used to predict GCN distributions over the entire county area. Predictions give a habitat suitability score of between 0-1, with 1 being very likely that GCN are present;
- Exported prediction as raster and created vector layers of habitat suitability using ArcPro;
- Risk zone maps were created using ArcPro modelling to show locations of habitats suitable for GCN. These maps were checked by stakeholders and specialists;
- Strategic opportunity maps showing ponds to protect and the most beneficial locations for compensation ponds were created using an ArcPro model, which removed buffered urban areas, protected areas, roads, rivers, flood zones and

[28] https://www.ceh.ac.uk/services/land-cover-map-2015

applied the habitat suitability scores and LPA allocation data to determine opportunity scores. Scores were based on a whether a location met key habitat requirements (for example GCN prefer to be within 100m of woodland) and the number of favourable habitats within 250m radius;
- Strategic opportunity areas were overlaid with existing pond data to provide assurance that modelled suitable areas could be colonised;
- Development impacts were assessed by overlaying the polygon layer of LPA allocations with habitat suitability model outputs and generating a summary spreadsheet of the area of suitable habitat lost and presence/absence of GCN. Impacts were assessed on 10km^2 basis to produce licence guidance and decide location of conservation ponds.

Outputs:
- Guidance to developers, including costs incurred, conservation and mitigation measures needed and the most beneficial conservation pond locations to inform habitat developers.
- Risk zone map to advise development plans and determine the cost of compensation. Based on a traffic light system, highlighting areas to avoid due to GCN presence or likely presence in red, areas with possible GCN populations as amber and areas without GCN as green. High-risk areas are prohibited from using the compensation scheme and developments in medium-risk areas incur higher remediation costs;
- Strategic opportunity map provides information to habitat delivery contractors about favourable pond locations to bolster and connect GCN populations. The scheme is currently delivering 4:1 pond ratio at present, so for every pond lost due to development, four new ponds are created;
- Living Ponds layer for use in future projects;
- Output maps are available for internal use in other areas of monitoring and are provided to regional steering groups and habitat delivery bodies (for example private contractors and Wildlife Trusts) to decide pond locations.

Future:
- Continuously maintain and update the tool at least every 5 years, including implementing new models and incorporating new data, such as the Living England map;
- Review strategic opportunity map by assessing pond colonisation of current conservation ponds;
- Ambition to make openly accessible tool for EIA, where developers can assess impacts of proposed development before submitting an application. An ArcGIS tool that allows users to upload a vector layer of the proposed development and visualises the impacts and associated remediation costs could be considered;
- Possibility to use this approach for other species, such as dormice, water voles, bats, by incorporating different data layers. Similar work currently in place to support beaver reintroductions.

3.3.3 NatureScot Rivers for Conservation project

This project aimed to establish the “naturalness” of river catchments across Scotland and to categorise rivers in terms of importance for conservation, for example for conservation of Red List species or natural heritage, to influence management policies.

Reasons behind data use:
NBN Atlas data provided national coverage of species data for the pre-defined list of species of conservation concern from multiple sources. However, species records at low spatial resolutions were excluded, as this work required precise location data at <100m^2 resolution.

Data sources:
- NBN Atlas species records;
- SEPA water quality datasets, including macroinvertebrate records.

Scale: National coverage, advice given on catchment scale

Frequency: One-off project

Resolution: Fine as possible, <100m^2

Access:
- NBN Atlas data openly accessible;
- SEPA data provided directly.

Process:
- NBN Atlas data were extracted for a pre-defined list of riparian species of conservation concern;
- Assigning a "naturalness" score by assessing the presence of features important for natural heritage or the presence of species of high conservation concern.

Outputs: Currently being produced. Outputs will inform policymaking.

3.3.4 NatureScot invasive non-native species management

An important use of biological data is facilitating a rapid response to invasive non-native species (INNS). NatureScot have established an alert system to promote early detection and eradication of priority invasive species and examine the extent of widespread INNS to inform national management.

Reasons behind data use:
Data are sourced from citizen science recorders, national schemes and partner organisations to provide national coverage. Alerts from citizen science recording schemes enable a rapid response, but records need to be verified to be useful, as there have been issues with spurious records of Muntjac deer and other species. NBN Atlas data are accessed, but records would need to be verified. The need for an alert system and quicker data flows was highlighted by the fact that it took over 18 months for NatureScot staff to hear about the first recording of floating pennywort, an aggressively invasive aquatic plant, in Scotland. Currency of data is essential when managing outbreaks of INNS, as these species are difficult to control once established. Data are required at a fine resolution of <100m^2, preferably at 10m^2 resolution to match the Fisheries Trust data collection. Coarse resolutions of 10km^2 tend to overlap several catchments, which illustrates the need for fine resolution records.

Data sources:
- Citizen science data, including national monitoring schemes (for example BSBI) and smaller projects, such as Project Splatter[29];

[29] https://projectsplatter.co.uk/

- Fisheries Trust data, mapping the maximum upstream extent of species;
- Scottish Forestry Native Woodland Survey of Scotland[30] used for extent of widespread species, such as rhododendron;
- NBN Atlas.

Scale: National

Frequency: Continuous

Resolution: As fine as possible, 10m^2 for alerts, 1km^2 for trends

Access:
- Automatic alerts or records directly flagged by recording schemes;
- Fisheries Trust and Scottish Forestry data provided directly.

Process:
- List of species identified through horizon-scanning based on international priorities and species movement patterns through the rest of the UK;
- Engage with recording schemes, for example BSBI, for example through blog posts, to highlight target species;
- When INNS record flagged, local teams will implement management practices to eradicate or control the outbreak;
- Changes in extent of widespread species are analysed in GIS.

Outputs:
- Site management decisions;
- Reports on the status of priority INNS provided to ministers and partners every few years;
- EU report on priority INNS every 6 years.

3.3.5 NE HLF/LIFE Hoverton Great Broad restoration project

This project[31] started in 2015 and aims to restore Hoverton Great Broad to its historic state. The broad has been degraded for several decades and has been classed as “unfavourable, no change” in condition assessments. To improve the site condition, the broad will be dredged to restore former depths and biomanipulation of the fish community will facilitate the recovery of zooplankton populations. Restoring zooplankton communities and reinstating the macrophyte population will help to control the levels of algae within the broad.

Reasons behind data use:
Assessments of site suitability for restoration confirmed Hoverton Great Broad as an appropriate site. This project has benefitted from enhanced data availability, streamlined data access and flows, as partner organisations had a comprehensive knowledge of previous surveys and provided the datasets directly. It was highlighted that fish community data were not available, which limited the evidence used when assessing site suitability for the project. Anecdotal suggestions that the broad was an important spawning ground for bream were unsupported due to a lack of data, but the site may have been more important for bream than first thought. This highlights the requirement for accessible data to make

[30] https://forestry.gov.scot/forests-environment/biodiversity/native-woodlands/native-woodland-survey-of-scotland-nwss
[31] https://hovetongreatbroad.org.uk/the-project/

suitability assessments. The longevity of existing survey data enables scheme assessments to be made before and after implementation. Continuous monitoring of the fish and plankton communities and water quality has been conducted as part of the project and will continue after completion, providing the data required to assess scheme effectiveness.

Data sources:

- Internal site condition monitoring data available, as the site is an NNR and SSSI;
- Broads Authority annual macrophyte survey data since 1980s;
- EA water quality monitoring data, including phosphate levels;
- Existing data on fish communities were not available, but assessments of social impacts and changes in fishery yields were made;
- Internal project data collection, including phytoplankton and zooplankton surveys, water quality monitoring collected monthly, invertebrate survey commissioned, sediment cores to assess the seed bank and the presence of pollutants;
- Fish community monitoring data collection to continue for at least 10 years after fish biomanipulation measures have been completed;
- BTO/RSPB/JNCC Wetland Bird Survey (WeBS) data provide trends in wintering waterfowl over time.

Scale: Local, specific site level

Resolution: Fine, capture resolution

Frequency: One-off project, monthly monitoring data collected

Access:

- Previous condition assessments directly provided by the senior site manager;
- Broads Authority data are provided through contact with NNR staff and are also published online;
- EA originally sent interpreted dataset from water quality monitoring, but raw data are now accessed online as a .csv file;
- BTO provide BTO/RSPB/JNCC Wetland Bird Survey (WeBS) data following request.

Process:

- Monitoring programmes established at the beginning of the project to continue throughout and beyond completion, including commissioned surveys and links with universities;
- Pre-processing data to extract the relevant information carried out using Excel formulas;
- Analysis of the collated datasets will be contracted out to assess the scheme effectiveness. All available data will be collated to provide a before/after picture of the broad. Assessment of the impacts of fish community biomanipulation will continue for several years following the removal of fish barriers. There is also a related PhD project analysing fish movements using acoustic tags to provide an understanding of the spawning patterns of fish species and the impacts of the fish barriers;
- WeBS assessments will be conducted in-house to determine the effects of restoration on wintering waterfowl and compare population trends to the Broads as a whole, UK and international trends.

Outputs: Specific project reports

3.3.6 NE hydro-ecology project

This work provided advice to EA regarding water abstraction licences to restore sustainable abstraction levels in the Norfolk Broads, focussing on the impacts within Ant Valley. People working in the fens had suggested the hydro-ecology had changed significantly over the last 50 years, highlighting concerns over excessive water abstraction.

Reasons behind data use:
Plant species data were accessed, as there is a greater volume of plant data compared to invertebrate species. Invertebrate data would have enhanced this study, but data availability is low. The BSBI Distribution Database (DDb)[32] was used because the coverage of plant records is significantly greater than other sources, including the NBN Atlas, and data are available for the last 120 years, which enables full time-series analysis of long-term trends. NBN Atlas data and internal paper files were incorporated, but to a lesser extent. Data analysis was conducted at a 200m^2 resolution to match the EA water abstraction model. The species of interest in this project are rare and were historically recorded at a high-resolution, meeting the requirements for the EA framework of 200m^2. It was noted that availability of high-resolution data for more common species would be significantly lower.

Data sources:
- BSBI plant records for last 120 years;
- NBN Atlas;
- Internal site-level historic data.

Scale: Local, site specific

Resolution: 200m^2

Frequency: One-off project

Access:
- BSBI DDb through existing agreement;
- NBN Atlas data openly accessible;
- Internal historic data files stored on data management system.

Process:
- Compiled data from sources and conducted an extensive literature review of existing data;
- Analysed data in-house to map extinctions and declines of species requiring clean, nutrient-free water over the last 120 years;
- Downloaded species list for each of the EA 200m^2 model cell squares from BSBI DDb and applied Ellenberg scores (detailing the nutrient requirements) to each species. Analysis included plotting the date each species was last recorded within each square against ecological requirements. This provided a timeline of the disappearance of species within each 200m cell and across the site. A clear peak in disappearances was found at 1971;
- NE staff provided strong guidance to EA about water abstraction licences;
- EA conducted in-house modelling to decide which abstraction licences to renew and remove.

Outputs: Project reports, management advice to EA.

[32] https://database.bsbi.org/

3.4 Limitations

Many limitations in current data workflows were identified by interviewees. The main limitations are discussed in order of how often each were mentioned.

3.4.1 Data access

Constrained data access is a major limitation to data use across the CNCBs. Interviewees felt CNCBs were making statutory decisions based on approximately 50% of all available data. Knowledge of all available data is required to assess absence points and data coverage gaps. Staff recognised the need for discussions across the agencies about how best to engage and collaborate with the recording community to enhance data access.

Staff felt there are now huge disparities in the coverage of accessible data, as a result of limited data sharing. Data are not reaching the NBN Atlas and are instead held in restricted repositories, including LERCs and national scheme databases, which are only accessible through an official data sharing agreement or by direct purchase. Licence conditions stipulated by data providers can be restrictive with regard to the sharing of derived data products, which can limit the scope of projects. Concerns around the concept of fully open data were raised, as this was seen as a barrier to data sharing, particularly at capture resolution. Questions around how data collection, curation and verification would be funded using the open data model were also raised. Interviewees outlined difficulties with the open data model, including the reluctance of LERCs and recording schemes to share datasets openly, in some cases due to the business models of these providers. Some interviewees felt that if there was no monetary value to be gained from data, some providers may lack the motivation to update the datasets.

Interviewees felt important data sources are not being harnessed as much as possible. Survey data collected to support planning applications, including EIA data, are not shared with CNCBs due to commercial sensitivity, which restricts the use of this information in future assessments. Advisors are currently reliant on the knowledge of other members of staff to confirm that a previous EIA survey was conducted in the same area as a new proposal. Political issues were raised as a barrier to data providers sharing their data. Some recorders are uncomfortable sharing data due to concerns that the records will be exploited commercially in some way. Staff mentioned this lack of data access is causing damage to biodiversity, as development proposals are being approved in areas important for protected species and habitats due to the lack of evidence available. For example, development licences have been granted based on an incomplete dataset and then recorders have presented important bat data meaning the conditions of the licence need to be changed. Similarly, tree planting licences have been granted on areas that appeared to be suitable, then recorders shared data that classifies the area as an important dry heathland habitat.

Difficulties with relationships between LERCs and CNCBs were raised, especially by NE staff. Staff stressed that available data were accessed through collaboration between individuals rather than an official agreement. The loss of accessible habitat data, particularly on a local scale, was identified as a consequence of the change in LERC relationships, as these datasets are currently not on the NBN Atlas. Regardless of these political issues, it was widely recognised that the effectiveness of LERCs varied regionally and concerns were raised about the sustainability of the LERC framework, mainly due to a lack of resourcing. Opinions about the LERC role within data flows varied massively. Many members of staff felt that some form of agreement needed to be reached between CNCBs and LERCs to restore and enhance data access, while some interviewees wanted to rely solely on the NBN Atlas or other national repository. However, it was widely recognised that discussions across the recording community, including the NBN and Association of Local Environmental Records

Centres (ALERC)[33], were required to identify barriers to data sharing and improve collaboration.

Restricted data access leads to further difficulties with data licensing, which constrains the use of datasets in projects and involves huge data management tasks. Interviewees felt there was a lack of knowledge within teams about data licensing, and the time this process involves, and highlighted the need for wider staff training opportunities. It was often the case that correct permissions were not in place at the start of a project, meaning staff needed to retrospectively request changes to licences. If this request was denied, the project output needed to be adapted to remove the restricted dataset or the derived products were not shared openly.

3.4.2 Data flow

Most interviewees felt that current data workflows include too many recording platforms and data pipelines and that data providers were offered no assurance of the ultimate location of their records. It was noted several times that often records do not reach the NBN Atlas, including records submitted through iRecord, which highlights the deficiencies in current data flows. There is a high level of confusion about "who is collecting what from where" and concerns that important data are being overlooked or "falling through the cracks". 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. There was a general feeling that current data flows consist of several siloed databases with no links between them and a lack of clarity around the function and purpose of each system. The fragmented workflows described lead to a myriad of data sources for staff to access data, which significantly increases the data collation time. Revised data flows will require more integration between data repositories, archives and management systems.

Current data flows are subject to significant time lags, with estimates ranging from at least 6 months up to 6 years. Lengthy time lags were identified as a major limitation to the NBN Atlas data, as quicker updates are required for responding to planning applications, casework and INNS management. Verification of records was identified as a pinch-point in data flows, as species have different verification requirements and it was mentioned that the number of verifiers is dwindling.

Data sharing between partners was highlighted as a difficulty, particularly with marine acoustic, video footage and imagery survey data and "big" data, such as satellite images and datasets derived from automatic sensors. The need for a Government-approved file transfer system for large datasets was mentioned.

3.4.3 Data coverage

Spatial and taxonomic gaps were identified as a limitation to current data applications. Spatial gaps included remote areas, uplands, intertidal and shallow subtidal areas, and are prominent in Scotland. Identified taxonomic gaps exist in fungi, invertebrates, lichens and bryophyte groups. Marine teams also mentioned the low habitat data coverage in the marine environment. Current CNCB data uses focus on protected sites, as condition assessment programmes require data collection in designated areas. Conducting surveys outside protected sites was widely recognised as an ideal scenario, but it was mentioned that this is

[33] http://www.alerc.org.uk/

not possible with current funding. Staff are therefore reliant on citizen science and national monitoring scheme data outside protected areas.

3.4.4 Data quality

Data quality was cited as a limitation to data use, especially with some citizen science data. Interviewees noted that spurious and duplicated records of species are present on the NBN Atlas and in wide-participation citizen science datasets, including records outside the normal range of a species. This raises concerns around data quality and undermines the application of these datasets. However, interviewees were keen to emphasise the huge value of citizen science data and wanted to harness more of these records. The need for rigorous verification of records was highlighted, which would assure users that the datasets are reliable and enable them to apply these data with confidence. It was mentioned that the number of verifiers is declining and there is a loss of taxonomic expertise, particularly in younger generations. The volume of records of common species puts pressure on verifiers, who need to spend more time on the rare species that are difficult to identify.

3.4.5 Internal data management

Most interviewees mentioned the urgent need to update internal data management systems. Many CNCBs have started the process of updating management systems, and staff were hopeful that these changes would ease internal workflows.

The main issue raised was that there is often no clear repository to store internal data, particularly data collected as part of a discrete project or casework, including licence return data. There is an appetite to share these data for use in other areas of work or university research projects to maximise the applications of these datasets, but there are no clear pathways for doing so. Staff felt that data from projects and casework are difficult to find, not properly archived, with limited metadata and version control. Local teams also highlighted the difficulties of accessing local scale data internally, as central management systems often only hold national datasets. Data sharing across teams is also restrictive and staff often resort to emailing copies of datasets, which leads to version control issues and complications in data management.

Internal data systems were described as clunky, confusing and overly complicated, especially if staff are not regular users, and many mentioned having difficulties accessing raw data in a usable format, rather than a PDF. Each CNCB has many data portals, some internal and some external, which leads to version control issues. Data managers felt internal data flows are currently fragmented and that management tasks became more time-consuming as a result of system inadequacies.

Staff often mentioned the fact that a large amount of internal data was locked in unusable formats, including notebooks, paper files and CD disks. Mobilising these historic data would enhance current data applications, but there was no time set aside for this process. Improved data management processes will help to mobilise these legacy records for future use.

3.4.6 Low resolution

All interviewees would like to access capture resolution data, as this was seen to be the “gold standard” of data and it was recognised that capture resolution data would present more opportunities and enable more effective protection of wildlife. If capture resolution data are not available, there was a general feeling that data resolution should be “as fine as possible”. All interviewees felt that resolutions of 1-10km^2 were often not fit for purpose, with the exception of some national applications, including the Natural Capital Atlases developed

by NE, which display national data at a 25km^2 resolution. Many mentioned that 10km^2 resolutions are not based on ecosystem principles, as these data can straddle many catchments with completely different characteristics. For casework, responding to planning applications and monitoring rare species a resolution of <100m^2, preferably <25m^2, is required. In these use cases, the exact location of a species is required to ensure its protection. Many species have very particular habitat requirements, for example some species found in soils have a range of 30-40cm dependent on groundwater levels. In these cases, it is imperative to access high-resolution data to determine if a species is present. For EO applications, ground-truth data are required at <10m^2 resolution to match the resolution of satellite images.

Interviewees felt that coarse resolution undermined the NBN Atlas data and many recommended reviewing which species and habitats were presented at blurred resolutions. In the marine environment, resolution was not seen as a prevalent issue, as staff were able to access capture resolution benthic data in most cases.

3.4.7 Data formats and consistency

Inconsistencies in data collection formats increase the need for pre-processing and limit the application of datasets. Common pre-processing tasks included standardising species names and classifications and removing duplicated records. Staff felt these tasks were a necessary step in the data workflow, but some suggested this stage could easily be automated if more coding expertise was available in-house. Interviewees felt that surveys should be designed to collect data in a consistent format, using standardised methods, to ensure data can be used in a wider range of applications. More efficient internal data collection, including site condition monitoring and casework surveys, could help to fill gaps in data coverage for use in other applications. Staff emphasised the need for better metadata collection and suggested using an existing tool, such as iRecord, to standardise internal data collection.

3.4.8 Funding and resourcing

Many interviewees considered the lack of secured funding across the recording community as a limitation to future data use. The sustainability of national recording schemes and LERCs is uncertain and heavily dependent on long-term funding commitments. Changes in internal resourcing across the CNCBs has put pressure on data services staff. It was also observed that the recording community is ageing, and concerns were raised about a loss of taxonomic expertise among amateur recorders.

3.4.9 NBN Atlas functionality

The NBN Atlas was widely perceived as a powerful tool in biological data applications and interviewees felt the capabilities of the Atlas were very valuable. However, many users outlined some limitations in the functionality of the Atlas, including data types available, data visualisation and analysis capacity.

It was recognised that the NBN Atlas currently only supports species records as point data, rather than habitat polygons, tagging and tracking data. These limitations are particularly restrictive when including marine data, which increasingly uses biotope polygons and tagging data for marine animals. Visualising marine data on the NBN Atlas also suffers from the conversion from latitude/longitude to British National Grid coordinate reference systems. Marine teams described the fact that when marine point data are submitted to the Atlas, the location of the point is shifted to the centre of a 100m^2 square, which reduces the accuracy and reliability of the data and fosters a reluctance to upload marine survey data to the Atlas.

Many interviewees highlighted the fact that the ability to search data using specific criteria is no longer available in the Atlas framework. Staff identified useful previous search criteria including date range, data class, data source, species conservation status and protected site or area. Users also mentioned that visualisation options offered by the Atlas are limited and no longer include the ability to colour code records according to the last date recorded, date classes, conservation status. It was mentioned that the analysis capabilities of the NBN Atlas could be improved to offer functions similar to the Pantheon platform. Users also identified difficulties navigating the NBN Atlas and most interviewees felt that the Atlas interface could be more user-friendly. Other functions that are no longer provided by the Atlas include the ability to control the resolution of sensitive species data and editing the metadata of a record as a data provider.

3.4.10 Low awareness of available data

A low awareness of what data exists, where data are stored, any caveats around datasets and how to access data was often cited as a limitation to data use. Teams are reliant on the knowledge of other staff members when accessing key data sources, and are often only aware a survey has been conducted through a passing mention from a colleague. Staff are also not familiar with all data access systems and often struggle to extract relevant data and process to an analysis-ready stage.

3.5 Improvements

Possible improvements to current data workflows are presented in the order of how often each were mentioned by interviewees. The first three improvements were mentioned by most interviewees, including the need for connected workflows, improved internal data management and improved data access.

3.5.1 Connected, simplified workflows

Interviewees were unanimous in identifying the need for transparent and completed data flows. Crucially, data flows need to be simplified, with fewer routes and platforms to submit and access data. Clear signposting of what data are collected and where data are ultimately hosted would assure recorders that submitted records reach a central repository. One analogy given was modelling data flows on the postal service, where records are guaranteed to reach the correct destination. It was mentioned that trackable DOIs could be utilised to determine where data were applied and provide summary statistics to data providers to monitor data flows.

Some interviewees suggested data could flow into a single central repository, either the NBN Atlas or another platform, to give users more confidence that the records accessed are the complete dataset. However, it was emphasised by many interviewees that reformed data flows do not necessarily need to comprise a single repository, but it is essential that all repositories need to link together to create one flow. Data management systems would need to communicate with each other to create a streamlined and coherent workflow. Clarity around the role and purpose of each data system is also crucial, for example one platform could be used for analysis and a different platform could be used for visualisation and mapping. It was also suggested that the data itself does not necessarily need to be hosted in one place, but a central hub would signpost existing data meeting criteria, identify the data provider and the data location. A data archive platform was also suggested as a solution to ensure the preservation of historic data for future use, similar to DASSH[34].

[34] https://www.dassh.ac.uk/

Understanding resistance to data sharing and identifying barriers to data flows is essential and changes in the biological recording culture are potentially required to facilitate data flow reforms.

3.5.2 Improved internal data management

Staff outlined important aspects of an updated internal data management system, with the overarching aim of mobilising all internal datasets for use across teams. Improving internal data access would increase data use across multiple applications. For example, site condition assessment data could be incorporated into long-term species status assessments or discrete projects. Mobilising more data for use by site managers and ensuring local datasets are available centrally were highlighted as important features of an improved system. Some local staff felt that accessing internal datasets, such as previous site assessments, in the field would be useful in casework.

A standardised, searchable metadata catalogue or inventory of all data held within the organisation would help to signpost suitable datasets to staff. Creating a simple catalogue detailing existing datasets, where data are stored, and a point of contact would improve data use across teams. The need for a data archive was also recognised to ensure legacy data are safely stored for future use. Another option identified was to employ a central system to capture, store and interrogate data, including analytical tools and an automated function to mobilise records to the NBN Atlas. The importance of web server hosting was emphasised, which would improve data sharing and storage capacity, and help to avoid staff downloading individual copies of datasets, which currently causes version control issues. All marine-based staff interviewed stressed the importance of an update to Marine Recorder and highlighted the need for the new marine data management system to communicate with terrestrial data systems. Bridging the gap between marine and terrestrial datasets was seen as a solution to mismatched datasets, for example the extent of the intertidal zone differs between OS and UK Hydrographic Office maps. Interviewees suggested an internal system update could be funded centrally, or 5% of individual project funding could contribute.

Creating an organisation-wide, standardised approach to data acquisition would ensure the correct data sharing agreements and permissions were obtained from the beginning of a project. More support for staff is required during the data acquisition stages of projects.

3.5.3 Improved data access

Improved access to high-resolution data was identified as a key priority. Interviewees felt that improved data access, and having confidence staff are accessing all available data, would promote more effective wildlife conservation. Increasing data sharing across the recording community would enhance data applications across the agencies. Staff mentioned the need for more data sharing agreements and more transparent data licencing to improve access. It was also reinforced that building an understanding of the barriers to data sharing, particularly with the NBN Atlas, is essential in tackling data access issues in the wider recording community. It was suggested that Defra bodies could all have access to the same datasets and share data using a Defra-wide system, and interviewees felt internal CNCB data should be openly accessible to the public. It was widely acknowledged that EIA datasets are incredibly valuable for multiple applications, including site condition assessments, species status reviews, mitigating impacts from further development, and there was a definite appetite to mobilise these data. This would involve a change in legislation to establish data sharing as a pre-requisite for planning applications.

All interviewees agreed that enhanced access to capture resolution data would present more opportunities for data use. A review of which species and habitats are presented at blurred resolution on the NBN Atlas was suggested, as not all species and areas are sensitive.

Many interviewees felt a reform of LERC relationships was required to improve data access. Staff in NE mentioned re-establishing some form of funding agreement with ALERC would solve data access issues and avoid spending on individual cases. This agreement could be more flexible, such as a hybrid funding agreement based on ad-hoc data provision, or an agreement in areas where data coverage from other sources is low. The potential of establishing a data exchange system, whereby CNCBs provide historic internal data in exchange for required datasets, was also suggested. It was widely recognised that LERCs need more support across the UK to improve data access and spatial coverage. Increased funding or possible reforms in the LERC infrastructure were suggested to solve issues of low resourcing. For example, Shropshire Ecological Data Network[35] is no longer a physical record centre, but a group of organisations and individuals who collect and provide data under a common licence.

3.5.4 Communication across recording community

Maintaining volunteer engagement would improve the sustainability of national recording schemes and raise awareness of the importance of biological data use across the UK. Providing feedback to volunteer recorders, for example thanking them for records and offering advice about verification requirements, would help to maintain engagement. The potential for automated feedback provision is discussed by (Wal et al., 2016). Investing in the recording infrastructure was also suggested, including offering training in species identification, covering expenses, and offering opportunities for recorders to increase their expertise or become verifiers. A national recording newsletter would help to link groups together and promote collaborative working. Creating fun, user-friendly formats for sharing data uses and findings, such as videos and short articles could increase and diversify volunteer engagement. Innovative approaches are needed to encourage recorders to visit remote areas, which would improve data coverage. Improved communication with the recording community would help to understand current barriers to data sharing and would cultivate stronger, more trusting relationships with recording groups and individuals. Communicating CNCB data requirements and priorities with the recording community, such as creating a watchlist, would improve data coverage in these areas and promote efficient data collection.

Creating a single online hub to engage volunteers through one location was suggested as an improvement to the current framework where each project has individual portals. This would mean volunteers were able to get involved in multiple recording schemes and would establish efficient feedback channels. Data from the NBN Atlas could be extracted using Web Map Services (WMS) to visualise the importance of volunteer efforts in an area, or as part of a national scheme, which would promote further engagement.

3.5.5 Harness additional data

It was often noted that significant amounts of collected data are not being utilised across the CNCB remit. Accessing and applying citizen science data was seen as an option to fill data coverage gaps and an opportunity to engage the next generation of biological recorders. Interviewees felt further discussions across the agencies were needed to facilitate more effective use of citizen science data. The BTO, CEH and JNCC Terrestrial Surveillance Development and Analysis (TSDA) partnership has helped in exploring applications of unstructured data. There were concerns about the quality of some citizen science records and verification was highlighted as a vital step in the data workflow. Improving in-house QA

[35] https://www.shropshire.gov.uk/natural-shropshire/ecological-data-network-sedn/

and easing the current pressure on verifiers are possible options for streamlining citizen science data flows.

Mobilising EIA and historic datasets was seen as a major priority across the CNCBs. Possible solutions included a change in legislation to ensure EIA survey data are shared with agency staff, potentially after a time period, for example 5 years, so the data are no longer commercially sensitive. Mobilising historic records, both from internal files LERCs, would require substantial resourcing in-house or engagement with volunteer groups. The BSBI Herbarium at Home project is an example of a volunteer programme to mobilise historic data in paper files. Another possibility would be recruiting university students to mobilise historic data for use in research projects.

3.5.6 Standardised data format and metadata

Ensuring standardised formats are used in survey data collection would streamline pre-processing across agencies. The Darwin Core format was cited as the best option for a standardised format and it was suggested that automated processes could be used to convert current data into Darwin Core. Marine teams proposed a change in the Marine Recorder infrastructure could convert data into a Darwin Core format to streamline data sharing across teams. The importance of metadata was also emphasised, with reference to the MEDIN standards in marine data collection. It was suggested that mandatory metadata fields could be included in data submission platforms to standardise metadata collection.

3.5.7 Improve NBN Atlas functionality

Increasing the analytical capabilities and improving the user-interface of the NBN Atlas were suggested by many interviewees. The ability to create powerful search criteria was cited as a possible improvement, including filtering by date range, species status, and protected site or area. Improved visualisation options, such as displaying species records coloured according to the last recorded date and being able to download maps for use in reports, would increase usage. Many felt the NBN Atlas was best placed to offer analytical capacity centrally, rather than each organisation investing in in-house systems. The inclusion of a separate analysis tab with more ecological and statistical analysis tools, such as alpha hulls, would expand the remit of the Atlas.

An expansion in the data types hosted by the NBN Atlas was mentioned as a major improvement to its capabilities. Including habitat layers, tagging and tracking data, abiotic data and contextual data layers, such as air quality, climatic variables and socioeconomic factors, would massively enhance interpretation of species records. The ability to overlay species data with habitat and contextual layers would present opportunities in all aspects of data use.

Many interviewees felt the NBN Atlas interface was complicated and difficult to use. More intuitive tab labelling, improved instructions for how to use the Atlas and the ability to log in and edit metadata fields of submitted records were all cited as possible improvements.

3.5.8 Mobilise CNCB data

Many CNCB members of staff stressed the importance of mobilising internal data to complete data flows. Interviewees felt that public body data funded by the state need to be published openly in order to meet statutory duties. Local teams mentioned that internal site monitoring survey data are not mobilised to the public. Some members of staff are using iRecord to put data into the public domain, but it was noted that this was not a requirement. A need for greater transparency around CNCB decisions and associated evidence is needed to allow discussion and justification.

3.5.9 Open data model review

Many identified open data requirements as being a possible barrier to data sharing and data access across the CNCBs. Interviewees identified that in light of this limitation, it was more important to have access to as much data as possible than to adopt open data principals. Many mentioned the need to reinstate the two-tier access system on the NBN Atlas, as this would encourage data providers to submit their data at capture resolution. The FAIR data principles, encouraging data that is findable, accessible, interoperable and reusable, were cited as an alternative to open data. Many interviewees highlighted the fact that the FAIR data principles are more nuanced and allow data providers to control access, while users can easily find all data and metadata. This is an active area of discussion, see (Pearce-Higgins et al., 2018).

3.5.10 Efficient data collection

A review of data collection was suggested as a solution to data coverage issues. Using data for multiple applications was mentioned as an ideal scenario to maximise resources. For example, national recording schemes could be adapted to collect habitat data for use in ground-truthing EO data applications while collecting species records. It is important to align analysis aims with data collection at the beginning of a scheme or project to enable data collected to be applicable to many work areas. Establishing a network between recording community members, including national and local monitoring schemes, CNCBs, JNCC, NGOs and academia would facilitate collaborative working to answer critical research priorities. For example, university research projects could be aligned with overarching research questions to collect information to fill data gaps for use in CNCB work. Collaboration between users of biological data would enhance data availability and promote efficient use of resources.

3.6 Future use cases

Interviewees identified a range of future research priorities, which are discussed below and ranked according to how often each were mentioned.

3.6.1 Remote sensing

The main research priority identified by interviewees was harnessing remote sensing technologies to enhance data applications. CNCBs have already started to apply remote sensing techniques to nature conservation, including using high-resolution sensors and satellite data. Specific research aims included:

- Establishing spectral signals for each habitat class using drones to create high-resolution habitat maps;
- Creating continuously updated habitat maps to measure change in extent and condition;
- Detect significant changes in land parcels to target field survey effort (similar to Copernicus User Uptake projects with JNCC[36]);
- Develop time series datasets of wildfires and muirburn events;
- Apply commercial high-resolution satellite data, such as Planet[37] data, to assess peatland condition, monitor restoration progress, and woodland regeneration;

[36] https://jncc.gov.uk/our-work/copernicus-project/#:~:text=As%20leading%20UK%20experts%20in,Agreement%20on%20Copernicus%20User%20Uptake.
[37] https://www.planet.com/

- Use remote sensing of chlorophyll concentrations to assess the status of water bodies.

Interviewees highlighted the need for robust ground-truth data, and it was clarified that remote sensing techniques would not replace field surveys but work in tandem with more traditional data collection.

3.6.2 eDNA

eDNA techniques are already widely used in freshwater environments and interviewees identified the potential for applying these methods in terrestrial and marine monitoring. eDNA methods were viewed as a cost-effective option to increase data collection capacity within field teams. The possibility of automatically assessing soil communities using eDNA methods at field sensor stations was discussed.

3.6.3 Data used to answer current Government themes

Policies across the four UK administrations seek to promote and facilitate nature conservation, including Welsh Government's Sustainable Management of Natural Resources framework[38] and the UK Government's 25-Year Environment Plan[39]. Applying data to achieve policy commitments and exploring growing areas of research were identified as priorities. The main policy areas mentioned were Nature Recovery Networks and Biodiversity Net Gain, followed by implementing Natural Capital and Ecosystem Services approaches. Links to the Green Growth Strategy, including mapping blue carbon habitats and focusing on species and habitats sensitive to climate change, were also made. Other priority areas identified were:

- Peatland restoration;
- Ecosystem connectivity, including developing a habitat connectivity indicator for long-term monitoring;
- Providing advice to achieve tree planting commitments, including "Right Tree, Right Place" projects and novel approaches such as using data layers of vascular plant species associated with good quality habitats to classify areas according to suitability for tree planting;
- Ecosystem resilience;
- Controlling invasive species;
- Monitoring Red List species;
- Ecosystem Health;
- Changes in biodiversity in the urban environment;
- Monitoring changes at the edge of species ranges, as these individuals are likely to hold climate change resilient traits;
- Conserving genetic diversity, for example a recent NatureScot report details a scorecard approach to assessing genetic risks facing populations of 26 species[40].

3.6.4 Increased data collection

Collecting more data outside protected sites, updating habitat inventories and filling key data gaps, both spatial and taxonomic, were identified as future priorities. Monitoring areas outside protected sites will build an understanding of UK-wide trends and explore the impacts of broad-scale pressures, such as air pollution, on ecosystems. More habitat data

[38] https://gov.wales/sites/default/files/publications/2019-06/sustainable-management-of-natural-resources-guide.pdf

[39] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/693158/25-year-environment-plan.pdf

[40] Hollingsworth et al., 2020

are required to increase mapping capabilities, particularly in the marine environment, while invertebrates, fungi and lichens were identified as significant taxonomic gaps to focus on, as these groups are useful indicators of change.

3.6.5 Artificial Intelligence (AI)

Interviewees mentioned the growing capacity of AI techniques as an opportunity for future data uses, but it was recognised that ongoing development is required in this area. The application of AI to automatically identify species would ease pressure on manual verification. AI could also be employed to count large populations, including seal and seabird colonies, using high-resolution drone or satellite imagery.

3.6.6 Different data types

Exploring the application of different data types, including population information, tagging and tracking data, and acoustic monitoring data, would expand current data use beyond traditional presence/absence records. Using sensors, both in-situ and mobile satellite platforms, to collect data automatically would create large amounts of "big data" to apply to multiple objectives. For example, NatureScot are currently exploring automated data collection methods as part of a CivTech Challenge[41] to help farmers achieve nature conservation outcomes.

3.6.7 Upscale current work

Many interviewees identified upscaling current proof-of-concept work or continuing to update a current project as future priorities. For example, NIEA are expanding a pilot project mapping habitat extent in one county to the entire country level.

3.6.8 Improved analysis techniques

Novel data analysis techniques, including Bayesian statistics and machine learning methods are priorities in future work.

3.6.9 Integrated data collection

Adopting a more integrated approach to data collection would allow data to be applied to multiple areas of work and maximise resources. For example, surveys focusing on species monitoring could also collect habitat type ground-truth data for EO applications. The need to integrate marine and terrestrial monitoring was also highlighted as a future aim.

3.6.10 Working on landscape or catchment scale

Many interviewees stressed the importance of shifting the focus of monitoring and reporting to a landscape or catchment scale in order to capture the "bigger picture" when assessing broad-scale impacts, such as climate change. This is highlighted by NRW's recent shift to a "place-based" approach[42] for environmental management.

[41] https://www.civtechalliance.org/civtech-5-challenge-3-scottish-natural-heritage
[42] https://naturalresources.wales/about-us/news-and-events/blog/developing-area-statements/?lang=en

3.6.11 Impacts of abiotic factors on ecosystems

Interpreting species changes in terms of contextual information, such as air pollution, nutrient loading and climatic factors, is an important area of future research. Performing abiotic scenario modelling to find loading thresholds of ecosystem damage would help to create effective mitigation policies.

3.6.12 Applying new technologies in workflows

New technologies present options for workflow improvements. Suggestions for future development included creating a data mapping platform delivering real-time updates of changes in habitat extent and condition and species distributions using WMS. Utilising the Internet of Things network to enhance data collation and application was also suggested.

3.6.13 Changes required to facilitate future aims

Interviewees identified several changes in workflows and attitudes that would facilitate the future research priorities outlined above. The following changes are not presented in any order:

- Adapt current systems to handle “big data” and new data types, such as tagging and tracking and population information;
- Building in-house capabilities in EO and analytical techniques;
- Increased, long-term funding;
- Political will and recognition that the information needs to be collected;
- Societal change to increase awareness and desire to protect nature;
- Need to demonstrate the benefit of biological data to the general public, which will increase the uptake of citizen science schemes;
- Assess gaps in data coverage to target future data collection;
- Tackle the resistance to new technologies by talking to staff about requirements;
- Learn from other organisations to apply new technologies.

4 Conclusions

The common use cases outlined in this report highlight the wide range of biological data applications across local and national teams within CNCBs. The value of biological data in nature conservation was highlighted throughout interviews, and there was a strong feeling that CNCBs cannot perform their function as evidence-led organisations without improved access to available data. Interviewees identified the requirement for high-resolution data as a priority, to make informed, justified decisions to effectively protect wildlife. Improving data sharing and streamlining data flows across the recording community would present more opportunities for data use in future. Maximising data use in multiple applications and harnessing the increasing volume of citizen science data will expand the capabilities of CNCBs to conserve the UK's natural resources, while emerging technologies such as remote sensing and eDNA, present opportunities to monitor landscape scale changes in almost real-time. There is a strong appetite across CNCBs for capitalising on these opportunities and reforming biological data workflows to meet user requirements.

5 References

Alison, J., Maskell, L.C., Smart, S.M., Feeney, C., Henrys, P.A., Botham, M., Robinson, D.A. & Emmett, B.A. 2020. *Environment and Rural Affairs Monitoring & Modelling Programme (ERAMMP).* ERAMMP Report-30: Analysis of National Monitoring Data in Wales for the State of Natural Resources Report 2020. Report to Welsh Government (Contract C210/2016/2017)(UK Centre for Ecology & Hydrology Project 06297). https://erammp.wales/sites/default/files/ERAMMP%20Rpt-30%20GMEP%20re-analysis%20for%20SoNaRR2020%20v1.0.pdf

August, T., Harvey, M., Lightfoot, P., Kilbey, D., Papadopoulos, T., Jepson, P., 2015. Emerging technologies for biological recording. *Biol. J. Linn. Soc*. 115, 731–749. https://doi.org/10.1111/bij.12534

Department of the Environment. 1995. *Biological Recording in the United Kingdom*. Published by Department of the Environment. http://nora.nerc.ac.uk/id/eprint/7868/1/BR_in_UK_v1.pdf

Foster, G.N., 2015. Taking the oldest insect recording scheme into the 21st Century. *Biol. J. Linn. Soc.* **115**, 494–504. https://doi.org/10.1111/bij.12531

Hollingsworth, P.M., O'Brien, D., Ennos, R.A., Ahrends, A., Ballingall, K.T., Brooker, R.W., Burke, T., Cavers, S., Dawson, I.K., Elston, D.A., Kerr, J., Marshall, D.F., Neaves, L., Pakeman, R.J., Trivedi, C., Wall, E., Wright, F., Yahr, R., Bean, C., Blake, D., Campbell, R., Comont, R., Finger, A., Fraser, K., Genney, D., Hall, J., Hannah, A., Jehle, R., Jones, S., Kohn, D., Llewellyn, M., Lurz, P., Macdonald, I., McIntosh, J., Mitchell, R., O'Dell, J., Page, S., Pemberton, J., Pérez-Espona, S., Piertney, S., Sime, I., Thompson, D., Ogden, R. 2020. *Scotland's Biodiversity Progress to 2020 Aichi Targets: Aichi Target 13 – Genetic Diversity Maintained – Supplementary Report 2020.* Inverness, Scottish Natural Heritage. www.researchgate.net/publication/339149105_Scotland's_Biodiversity_Progress_to_2020_Aichi_Targets_Aichi_Target_13_-_Genetic_Diversity_Maintained_-_Supplementary_Report_2020

JNCC 2019. *The UK Approach to assessing Conservation Status for the 2019 Article 17 reporting under the EU Habitats Directive*. Joint Nature Conservation Committee, Peterborough. Available to download from https://jncc.gov.uk/article17

Pearce-Higgins, J.W., Baillie, S.R., Boughey, K., Bourn, N.A.D., Foppen, R.P.B., Gillings, S., Gregory, R.D., Hunt, T., Jiguet, F., Lehikoinen, A., Musgrove, A.J., Robinson, R.A., Roy, D.B., Siriwardena, G.M., Walker, K.J., Wilson, J.D., 2018. Overcoming the challenges of public data archiving for citizen science biodiversity recording and monitoring schemes. *J. Appl. Ecol*. **55**, 2544–2551. https://doi.org/10.1111/1365-2664.13180

Pescott, O.L., Walker, K.J., Pocock, M.J.O., Jitlal, M., Outhwaite, C.L., Cheffings, C.M., Harris, F., Roy, D.B., 2015. Ecological monitoring with citizen science: the design and implementation of schemes for recording plants in Britain and Ireland. *Biol. J. Linn. Soc*. **115**, 505–521. https://doi.org/10.1111/bij.12581

Pocock, M.J.O., Roy, H.E., Preston, C.D., Roy, D.B., 2015. The Biological Records Centre: a pioneer of citizen science. *Biol. J. Linn. Soc*. **115**, 475–493. https://doi.org/10.1111/bij.12548

Roy, H.E., Pocock, M.J.O., Preston, C.D., Roy, D.B., Savage, J., Tweddle, J.C. & Robinson, L.D. 2012. *Understanding Citizen Science & Environmental Monitoring. Final Report on*

behalf of UK-EOF. NERC Centre for Ecology & Hydrology and Natural History Museum. http://www.ceh.ac.uk/sites/default/files/citizensciencereview.pdf.
Roy, H.E., Rorke, S.L., Beckmann, B., Booy, O., Botham, M.S., Brown, P.M.J., Harrower, C., Noble, D., Sewell, J., Walker, K., 2015. The contribution of volunteer recorders to our understanding of biological invasions. *Biol. J. Linn. Soc*. **115**, 678–689. https://doi.org/10.1111/bij.12518

Wal, R. van der, Sharma, N., Mellish, C., Robinson, A., Siddharthan, A., 2016. The role of automated feedback in training and retaining biological recorders for citizen science. *Conserv. Biol*. **30**, 550–561. https://doi.org/10.1111/cobi.12705

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*. https://nbn.org.uk/wp-content/uploads/2018/11/SBIF-Review-Final-Report-and-Recommendations.pdf