

NBN Standards for integrated online recording and verification

Introduction

Online recording has significant potential to improve the efficiency of biodiversity data flow, to facilitate structured surveillance and monitoring, to increase data mobilisation from the commercial, academic and public sectors and to provide much-needed support to the network of experts who provide verification services, often on a voluntary basis. The benefits of online recording are already widely acknowledged by the biodiversity data management sector, and various online recording systems have been developed by key organisations including several local environmental records centres and national recording schemes. Online recording will only achieve its full potential if these systems can interoperate and are based on a shared set of standards. The NBN already has a suite of tools, standards and principles which provide a secure and equitable basis for the dissemination and use of biodiversity data. These provide a firm foundation, but are likely to need some extension and development to meet emerging requirements.

This document presents a set of existing NBN Standards and Tools for integrated online recording, together with proposed next steps for development.

1) Uses centrally updated NBN Species Inventory and Habitats dictionary

Why is this necessary?

Use of older dictionaries or developing new lists can cause problems further down the line where the various sources need to be reconciled into a single picture. By setting up automated use of the centrally updated NBN Species Inventory and habitats dictionary the hard work is done for you. These resources are updated centrally reducing the overhead to those who need current lists of species names or habitats.

What is required?

There needs to be a single centrally managed source of names and another for habitats which is responsively updated as new entries are required and an agreed mechanism to periodically publish updates to users of the list including the various on-line systems.

Next steps

Review the content of the Species Inventory - identify any obvious gaps that need to be tackled by the NHM dictionary team. This includes removal of duplicates and ensuring the taxonomic hierarchy is fit for use within the on-line recording domain. Also need to establish a format in which Inventory updates will be supplied. Work on this has already commenced but needs to involve other implementers of on-line systems to ensure all needs are covered through a single supply route.

Review terms and conditions around use – currently the use of species inventory is governed by a set of terms and conditions that limit its use within commercial applications. Clearly at least some of the on-line data capture systems could be defined as commercial and this would need to be reviewed to ensure that the species inventory can be used in such systems.

Review the content and publication of the habitats dictionary – the habitats dictionary has not been updated for some time. Need to review its current state and gaps against the requirements for data capture. In addition there is a need to review the keys used within the system to eliminate overlap with the Species Inventory.

2) Exchange data in a standard format

Why is this necessary?

Making use of a standard format eases the exchange of records between systems to facilitate data sharing via the NBN Gateway and to enable verification by experts in specialist systems, for example exporting bird records from iRecord to BirdTrack and non-bird records from BirdTrack to iRecord

What is required?

The NBN Exchange format (<http://nbn.org.uk/Share-Data/Providing-Data/NBN-Data-Exchange-format.aspx>) is currently the primary format for exchanging data across the NBN. This defines the elements that are required and a set of additional that can be provided. Currently the elements required are:

- Record identifier – a unique identifier within the dataset for the record
- Species identifier (from NBN Species Inventory) – requires matching to, or use of the NBN Species Inventory, to provide the standard “Taxon version Key” for each species name used.
- Date – date observation was made on formatted in standard way
- Position – either the grid reference, co-ordinates or the key of a known site against which the record was made.

For more complex data the NBN Exchange Format can be extended (through the creation of additional attributes). However, where the data that are being exchanged relate to more complex methodologies there may be a need to begin to consider a more encompassing standard. Darwin Core is the primary contender here and it would be worth considering whether this should be adopted as the de-facto standard from the outset.

Next steps

Initially we need to agree what format(s?) will be the primary mechanism for exchanging records. As noted above the primary contenders here are the NBN Exchange format and Darwin Core.

Agree how record identifiers will be assigned in order to ensure uniqueness across the network and allow records to be freely exchanged without the risk of duplication of keys for different records.

Individuals also need to be uniquely identified into order to allow them to be recognised across systems.

3) Recorders give consent for their records to be shared and used at the point of data entry

Why is this necessary?

NBN Data Exchange Principle 4 states that “a clear transfer of authority should be made when a biodiversity resource is put together, to allow biodiversity data managers to act on behalf of biodiversity data owners”.

Recorders hold the intellectual property right in their records, and uncertainty over recorder permissions has been a barrier to data sharing in the past (for further information see NBN guidance [‘Clarifying permissions to use data’](#)).

The NBN Standard for online recording sites is therefore that the recorder gives implicit consent for their records to be shared and used for all purposes, including commercial use, simply by entering the records on the system.

What is required?

The NBN Model Recording Form Permissions Statement was prepared for use on paper and electronic recording forms to give permission for the dissemination and use of the records submitted:

By submitting information on this form I agree that it may be collated and disseminated electronically for environmental decision-making, education, research and other public benefit uses. Your contact details will be used for administration and verification purposes only, whilst your name will form part of the record that is collated and disseminated.

This model statement secures the recorder’s permission for their name to be associated with their record. The online recording system should enable the recorder’s name to be stored and disseminated as part of the record. This can be useful information during the verification process and ensures that the recorder can be credited if their record is used in a publication.

Next steps

None

4) Uses NBN validation and verification checks

Why is this necessary?

Defra, the NBN Trust and the community are investing in a set of standard validation and verification checks. These are intended to help ensure that the data is in the correct format and helps flag unusual records for closer examination. Using these tools is intended to help reduce the amount of time required from experts to verify the data and is designed to generally help improve the quality of data across the network.

What is required?

The rules are currently published in standard text file format. Additionally the Record Cleaner tool is a desktop system for checking a record set against the currently published rules. For

other systems investing in data capture it is recommended that the rules are run as the records are entered and the user entering the records immediately notified or any potential problems which could potentially be eliminated immediately. The NBN Record Cleaner rules have already been incorporated into iRecord and are being used to provide feedback to recorders and to provide a filtering mechanism for verifiers, for more information click [here](#).

Next steps

It is anticipated that the rules will be available as web services during 2013

The page on Record Cleaner (<http://www.nbn.org.uk/Tools-Resources/Recording-Resources/NBN-Record-Cleaner.aspx>) also includes a set of outputs from various schemes on data relating to those schemes. We need to integrate this better within the NBN website so that the list of rules is clearly available for download.

We will also consider publishing the rule file format and the rule files themselves as an independent web page.

Rules have been created for just under 18,000 taxa including 77% of the taxa on the UK BAP list. We need a strategy for funding and managing the creation of verification rules for the full UK taxonomy or at least the priority groups which are most likely to be recorded and for which sufficient baseline data exists to create the rules. We also need a sustainable mechanism for keeping the rules updated automatically as new verified records are added.

5. Records are collated into datasets and shared via the NBN Gateway under the administration of appropriate organisations

Why is this necessary?

Records need to be collated into logical datasets so that they can be accompanied by appropriate metadata. Collation will usually be based on taxonomic divisions, but could be based on survey methodology or geographic area in some cases. Datasets must be administered by organisations able to maintain the top copy of the data in response to comments received by users of the NBN Gateway, and able to grant approved users access to sensitive records where this is justified. The organisation who administers the dataset will be acknowledged if the data are used in a publication.

Although it is technically possible to download data from many online recording systems, the NBN Gateway is the preferred route of data dissemination because:

- Data use is governed by Terms and Conditions
- Detailed feedback is provided to the dataset administrators on who has accessed and used their records and for what purpose, enabling this information to be fed back to the recording community
- Mechanisms exist for protecting sensitive records from public access while allowing access for approved users
- Use of data from the NBN Gateway requires that the data providers are acknowledged
- A suite of NBN web services and WMS are available to meet the requirements of a range of users
- Some data users (e.g. Natural England, Environment Agency) have established business systems that rely on data provision via the NBN Gateway
- The NBN Gateway allows automated data dissemination via the Global Biodiversity Information Facility (GBIF)

What is required?

As outlined above, the NBN Gateway already provides a suite of data administration tools for data providers, data delivery mechanisms for a range of users, and terms and conditions governing the use of data from the NBN Gateway. It also provides feedback and reporting mechanisms for dataset administrators, which are being expanded and improved as part of the rollout of Gateway v.5 in 2013.

Next steps

The Indicia/iRecord community warehouse and NBN Gateway database are currently two separate databases hosted on CEH servers. It is possible that these may become a single database in future, although the iRecord and NBN Gateway portals will continue to perform separate functions:

iRecord functions:

- ingestion of individual records
- ingestion of data from Indicia-based online recording systems for verification and onward dissemination to the NBN Gateway
- ingestion of unverified datasets from non-Indicia sources for verification and onward dissemination to the NBN Gateway (e.g. non avian records from BirdTrack)
- facilities to enable data verification by experts supported by NBN Record Cleaner rules
- facilities to enhance the recording experience for individual recorders (data summaries, mapping, photo galleries, recorder league table etc)

NBN Gateway / NBN web service functions

- ingestion of collated datasets accompanied by metadata
- data dissemination to all users (general public, recording community, local records centres, NGOs, academia, national and local government etc)
- data dissemination via GBIF
- ongoing improvement of data quality through peer review via record commenting
- provision of feedback to data providers on the use of their data
- provision of data management tools to data providers

6) Passes data directly to the NBN Gateway via web services/without manual handling

Why is this necessary?

NBN Data Exchange Principle 1 states that “biodiversity data should be easily accessible to enable their use for not-for-profit decision-making, education, research and other public-benefit purposes.”

Passing data to the Gateway is intended to simplify the current model of data flows and reduce the lag between observation and use. By passing data directly to the Gateway this will ensure that users have access to the latest information. This should help improve biodiversity outcomes and avoid time-consuming data handling required to perform bespoke requests required when the latest data is not on the NBN Gateway.

What is required?

Sharing data with the NBN Gateway can be conducted via NBN exchange format or a reduced version of Darwin Core. The Gateway also needs to be able to handle automated

ingestion of data provided in this standard format (due for completion later this FY).

Next steps

Agree what the primary format for exchange of data between systems will be and trial direct publication from on-line systems to the Gateway.

7) Identify sensitive data using a clear application of the NBN criteria

Why is this necessary?

NBN Exchange Principle 2 states that “making biodiversity data available should reduce the risk of damage to the environment. If it is likely to have the opposite effect, availability may need to be controlled”. It is important that sensitive records are appropriately handled to mitigate the risk of damage, and that all recorders are made aware of the potential sensitivity of such records. In addition, in the interests of data being as openly available as possible it is important that data are not flagged as sensitive where there is no justifiable reason for this.

What is required?

The NBN has developed a set of criteria in order to assess sensitivity. These criteria are being used by national recording schemes to develop a list of species (and other conditions) that are regarded as sensitive.

Next steps

Continue to develop the standard list of species and also need to ensure this is published through the website.

8) Automated synchronisation of verification comments between systems

Why is this necessary?

Most taxonomic experts operate on a voluntary basis and have great demands on their time. It is therefore critical that they can operate as efficiently as possible. One way this can be delivered is by ensuring that they can provide feedback on records using a single system /interface with which they are familiar. It is then important that any comments they make on records are associated with the original top copy of the record as quickly as possible. This mechanism is anticipated to be as fully automated as possible.

What is required?

There needs to be a standard format and protocol for exchanging annotations on records, These annotations include those generated manually (i.e. by an individual commenting on a single record) as well as those generated by more automated checks. Establishing a format / protocol allows the comments to be rapidly moved back to the original copy of the records even if the annotations were not made in that system (e.g. where the expert is not using / familiar with the same system the recorder is using to enter and manage the records).

Next steps

This area would be best taken forward through a meeting of those involved in the development of on-line systems to begin to sketch out and agree a protocol for exchanging record annotation information. The outputs of this would then be worked up into a proposed standard for wider circulation and, ultimately, publication. Finally the standard would need to be integrated into the various systems.

9) Controlled vocabulary for verification status across all systems

Why is this necessary?

Most online recording systems incorporate a system for attributing records with their verification status. The verification status may be derived from a combination of the recorder's confidence, the outputs of automated verification rules and the decision of one or more experts. In order for records to be interchanged between systems with no loss of information, a standard vocabulary for verification status needs to be adopted.

Agreement needs to be reached on the combination of factors (recorder confidence, automated checks, expert decision) that would lead to a record being attributed with each verification status.

What is required?

The NBN recommended framework for defining the verification status of records is:

- Correct
- Considered correct
- Requires confirmation
- Considered incorrect
- Incorrect
- Unchecked

This framework was used by national schemes and societies who created verification rules for Record Cleaner as part of their documented procedure for dealing with records which 'fail' the Record Cleaner rules.

Next steps

NBN Trust and BRC will produce general guidance on the combinations of factors which equate to each verification status, e.g. 'considered correct' could apply to a record that has high recorder confidence and passes all automated checks but has not yet been checked by an expert, while 'considered incorrect' could apply to a record that has low recorder confidence and/or fails some automated checks and has not yet been checked by an expert. This will take account of the taxon-specific guidance produced by the national schemes who produced the Record Cleaner rules.

Developers of online systems will be invited to review this guidance and evaluate whether their system's verification status labels can be easily mapped to the standard, e.g. does 'probably correct' in Cofnod's ORS mean the same thing as 'considered correct' in the NBN standard?

10) Unverified data needs to be made available promptly to key users

Why is this necessary?

As stated above, verification is largely carried out on a voluntary basis and taxonomic expertise is a scarce resource. This can result in delays in the data being made available. In some cases, such as a new outbreak of a non-native species, it is imperative that key users (e.g. the Environment Agency and local environmental records centres) have access to records even before they are verified.

What is required?

Online recording systems need to export unverified data to the NBN Gateway with access restricted to key users. The NBN Gateway Data Access Controls already include a control to flag certain records as sensitive, which makes them invisible to the public but allows access to be granted to approved users. This technology could be adapted for unverified records.

Next steps

Explore options for incorporating an 'unverified' flag into the NBN Data Access Controls and the mechanism for exporting data from online systems to the Gateway with this flag.

11) Data should have accompanying information (meta-data)

Why is this necessary?

NBN Data Exchange Principle 3 states that “when biodiversity data are supplied, accompanying information (metadata) on its ownership, methods and scale of collection and limitations of interpretation, should be provided. In general there should be sufficient metadata to allow biodiversity data users to assess the scope and potential uses of their information holdings.”

Records captured through online recording range from effort-based structured surveillance and monitoring data to *ad hoc* wildlife sightings. Records may be collated according to survey methodology, taxonomic group, geographic area or a particular project, e.g. a Bioblitz.

What is required

All data shared should have associated metadata developed in line with the NBN standard.

Next steps

Not much to do here. It could be worth exploring with the on-line data capture systems what a sensible unit of dataset is and ensure the systems capture (or can otherwise generate) adequate metadata to describe these units.

Currently metadata is provided manually by the dataset administrator – can this be automated (bearing in mind the NBN Standard metadata form has many free text fields)?

12) There should be a single top-copy version of the data

Why is this necessary?

There is a risk in exchanging copies of electronic records that the copies gradually diverge as individual copies are edited. There should be mechanisms in place to ensure that the top copy of any record is clearly identified and that duplicate copies of that record can be viewed and additional attributes associated with it but the copy of the record itself cannot be edited. This then allows the original record to be periodically republished to other systems without worrying about resolving changes that have potentially occurred in the multiple copies.

What is required?

Need to ensure that where a system includes the ability to import records from another system that there is a mechanism to identify whether the transfer represents the top copy of the records (e.g. where the recorder is moving all of their records to a different system) and ensure that if duplicate copies are being imported that sufficient mechanisms are in place to ensure that the records cannot be edited.

Next steps

Review the exchange formats and ensure this use case is covered. As systems develop the capability to import records ensure that the rule limiting record editing is implemented.

13. Clarity is needed over who can edit records and add determinations

Why is this necessary?

In some online recording systems (e.g. BirdTrack, iRecord), only the recorder can edit their own records and change determinations. In other online and offline systems (e.g. iSpot, Recorder 6, Marine Recorder) experts or data managers can add a determination without removing the recorder's original determination. Data managers, such as Local Records Centres, may wish to edit other details of the record, for example to correct mistakes in site names.

If records are to be exchanged between systems, clarity is needed over who can edit records and add determinations in each system.

What is required?

The NBN Data Model sets the standard for changing the determination of a record:

“In The NBN Data Model, the central premise relating to identifications is that once a determination or identification is applied to an observation or specimen, it cannot (normally) be changed or removed. Any change of name, either through disagreement over the correctness of the identification or for simple nomenclatural update must be added to the identification history of the record or specimen and may be set as the preferred identification.” (Copp, 2004).

Next steps

The Standards cannot be prescriptive over who can edit records and add determinations, as different approaches have arisen in response to different user needs. The Standards can only specify that changes must take place only in the top copy of the record, and that determination changes made by a party other than the original recorder must be added to, rather than replacing, the original recorder's determination.

14) Personal details should be handled appropriately

Why is this necessary?

NBN Exchange principle 6 states that "Personal data must be managed in accordance with the principles of the Data Protection Act 1998 and/or any subsequent legal provisions." There is obviously a legal obligation to hold and manage personal information appropriately and there is therefore a need to ensure that system developers are aware of the requirements and have implemented these appropriately.

What is required?

Need to clarify exactly what the requirements are in this area (once the technical frameworks have been established) and ensure that package developers are aware of these and are implementing them appropriately.

Next steps

Clarify the technical frameworks, particularly around user registration and login, and the measures that need to be taken around this to manage personal data are clear.

Written in October 2012 by Paula Lightfoot (NBN Trust) and Steve Wilkinson (JNCC)