

Guide to downloading and using the NBN Nameserver Version 1.1 (last updated 03/03/2008)

Compiled by John Tweddle and Charles Hussey

Introduction

This document provides an overview of the NBN Species Dictionary Nameserver, including functionality and how to obtain a copy. Guidance on how to match taxonomic names held in external datasets to those used by the NBN are also included, along with contact details for the project team.

Overview of the NBN Species Dictionary Nameserver

The Nameserver is a facility that is being compiled as part of the NBN Species Dictionary Project. It provides a controlled naming source for use across the NBN and is suitable for wider use within the broader UK biodiversity community. It is being compiled by the Natural History Museum, London and draws upon the taxonomic expertise of a wide range of experts, including Natural History Museum staff and Species Dictionary data providers.

The same organism can be known by a variety of different names: current, obsolete and incomplete scientific names, one or more common names and a variety of misspellings. This can lead to confusion, particularly when combining datasets or searching across disparate resources. In such situations it can often be difficult to ensure that all relevant data for an organism are linked and retrieved if, for example, they have been stored against different names. The Nameserver is designed to solve this problem by matching all of the names that have been used for an organism to its current NBN recommended scientific name. This serves two main purposes. Firstly it acts as a query expansion tool to ensure that all relevant data are returned, and secondly it provides a mechanism whereby the current scientific name for an organism can be checked and retrieved. This allows changes in taxonomic understanding to be monitored and a taxonomic dataset to be kept up to date. The facility contains all the names that are currently included within the Species Dictionary (currently more than 245,000) and will be expanded over time to include other names that are in use for UK wildlife.

As well as providing information on synonymy, the Nameserver also serves to characterise each name. It contains three flag fields which combine to identify the type (scientific or vernacular), status (synonym or current NBN preferred name), and accuracy (well-formed or incorrectly-formed) of each name. Informal taxonomic group information is also provided, with each name being assigned to one of 109 informal groups (e.g. moss, flowering plant, amphibian). This provides a user-friendly way of identifying the type of organism that a name applies to, as well as allowing datasets and search results to be filtered and organised in a simple manner. Hence, it is possible to easily filter the Nameserver so that it just provides data for the types of name and taxonomic groups that are required.

The Nameserver forms a core component of the NBN Species Dictionary, as well as providing a nomenclature service for the NBN Gateway and reporting functionality within Recorder 6. It thus ensures that we are able to provide consistent naming across NBN products. We hope that the facility will also prove useful within the broader UK biodiversity community. The Nameserver is likely to be of interest to database owners, website developers, NBN data providers and any other professional and amateur organisations that require up to date taxonomic data.

Envisaged uses include:

1. Validating and maintaining taxonomic nomenclature: adopting NBN recommended naming and naming standards; ensuring that accurate and consistent naming is employed in databases and on websites; keeping nomenclature abreast of taxonomic advances.
2. Providing the NBN codes (TAXON_VERSION_KEYS) for names: delivering observational data for use on the NBN Gateway; aiding accurate exchange of biodiversity information; simplifying creation of links to NBN websites; utilising NBN Web Services.
3. As a query expansion and filtering tool: on websites and in databases.
4. To assign informal taxon group information to a list of names: to improve clarity of information, aid reporting and allow simple filtering of data.

Structure of the NBN Species Dictionary Nameserver

The Nameserver is provided for external use in flat-file format (i.e. as a single table), the structure of which is detailed in Appendix 1. It does not include any hierarchical information; this role is fulfilled by individual checklists within the Species Dictionary. Instead its role is to link each name to its current NBN recommended scientific name (via the NBN_TAXON_VERSION_KEY) and to provide the flag fields and informal group information that is discussed above.

Care notes:

1. Scientific names are stored within the NBN Species Dictionary in three parts. For the purposes of the Nameserver, the relevant fields are TAXON_NAME, TAXON_AUTHORITY and TAXON_QUALIFIER. Examples of how these fields are populated are provided in Appendix 1 (note that the latter two fields may be empty). These three fields combine with RANK and INFORMAL_GROUP to complete the name concept. **Each unique combination of these five fields is assigned its own 16-character alphanumeric NBN_TAXON_VERSION_KEY** and it is this code which is generally used for data exchange within the NBN.

The net effect is that variations of the same name will have different NBN_TAXON_VERSION_KEYS. For example, the name *Poa annua* L. would have a different key to the misspelled form of the name *Poa anua* L., or a version of the name that lacked the naming authority (*Poa annua*).

2. Vernacular name information is stored solely within the TAXON_NAME field (it is not split into three components like a scientific name). Each unique combination of TAXON_NAME, RANK and INFORMAL_GROUP will be assigned its own NBN_TAXON_VERSION_KEY. Hence the name 'Redshank' as applied to the moss *Ceratodon purpureus* will have a different key to the name as applied to the bird *Tringa totanus*. It is important to separate such differing interpretations of a name.
3. In a relatively small number of cases, the same TAXON_NAME may match to more than one RECOMMENDED_SCIENTIFIC_NAME, hence there is not always a one-to-one relationship between the NBN_TAXON_VERSION_KEY and the NBN_TAXON_VERSION_KEY_FOR_RECOMMENDED_NAME. This situation most

commonly results when a taxon name is demoted to a synonym of two or more species, or where a vernacular name can apply to more than one subspecies of an organism.

4. The Nameserver is a work in progress and not all entries have yet been verified by an expert. Such records have an entry of U (unverified) within the NAME_STATUS and/or NAME_FORM field(s). Such entries will be validated as new taxonomic checklists and expert time become available. Please inform us of any errors that you spot.

See Appendix 2 for other hints and tips for name matching and assigning NBN_TAXON_VERSION_KEYS.

How to obtain a copy of the Nameserver

A date-stamped copy of the Nameserver can be sent as an e-mail attachment upon request to speciesdictionary@nhm.ac.uk. The Nameserver is not static and the downloadable files will be updated on a six-monthly basis to incorporate taxonomic changes, additions of new names and other general improvements. When requesting a copy, please let us know of your intended use for the Nameserver. Providing that your request satisfies our general Terms and Conditions (see below), you will then be emailed the file free of charge. All personal information will be kept secure and treated confidentially.

The Nameserver can currently be downloaded in two file formats:

1. A zipped tab-delimited text file (currently 5.1 MB zipped and 44 MB uncompressed). This includes information for all of the taxonomic groups that are currently contained within the NBN Species Dictionary and is suitable for importing directly into databases. Because of the large number of records contained (> 245,000), it is not suitable for direct import into some software packages (e.g. Microsoft Excel).
2. A series of zipped tab-delimited text files (each less than 65K rows to allow import into most spreadsheet packages). The spreadsheet contains separate worksheets for the following nine groups: higher plants, lower plants, fungi and fungoids (excluding lichens), lichens, vertebrates, insects - beetles, insects - other, other invertebrates, remaining taxon groups.

We are currently considering the possibility of creating Web Services and will do so if there is sufficient user demand. Please contact us if you would find this service useful, or have a request for alternative file formats.

Terms and conditions for use of the Nameserver

The Nameserver forms a component of the NBN Species Dictionary and as such is covered by the general terms and conditions for use of the database. These may be viewed at <http://nbn.nhm.ac.uk/nhm/sdtandc.shtml>. We may change these terms and conditions at any time, so please check them each time you intend to use the Nameserver. If you are in any doubt regarding whether your intended usage of the Nameserver is permitted under these terms and conditions, then please contact us and we will be happy to advise.

Contact details

Mr Charles Hussey
Species Dictionary Manager, Natural History Museum, Cromwell Road, London SW7 5BD. email: speciesdictionary@nhm.ac.uk; phone: 0207 9425213

Appendix 1: Nameserver data structure

Field name	Field type	Description
INFORMAL_GROUP	Text (50)	Informal name for the taxonomic group. e.g. bird; moss; insect - hymenopteran.
NBN_TAXON_VERSION_KEY	Text (16)	Unique 16-character key for the taxonomic name. This key should be used for data exchange with the NBN. Referred to as the 'TVK' below.
TAXON_NAME	Text (85)	Taxonomic name string e.g. <i>Poa annua</i> ; House Sparrow. Includes standardised infraspecific rank term as appropriate. e.g. <i>Stenus crassus</i> var. <i>crassus</i> .
TAXON_AUTHORITY	Text (80)	Naming authority e.g. (L.) Gaertn.; Linnaeus, 1758. This field can contain extended characters, such as é. This is not a problem in Access, but in text files these may be invalid characters and they will not always import into databases or spreadsheets correctly. Therefore, ensure that all new authorities with extended characters import correctly.
TAXON_QUALIFIER	Text (95)	Contains any qualifying information for the name concept. e.g. <i>sensu stricto</i> ; <i>sensu lato</i> ; <i>nec</i> Smith
RANK	Text (25)	Taxonomic rank for name.
NAME_FORM	Text (1)	Flags whether the name corresponding to the above TVK is well-formed (W) or incorrectly formed (I). Well-formed names are spelt and formatted correctly (to NBN standards). A third category of unverified (U) applies to names that have not yet been checked by an expert.
NAME_STATUS	Text (1)	Flags whether the name corresponding to the above TVK is a recommended (R) or synonymous (S) name. A third category of unverified (U) applies to names that have not yet been checked.
NAME_TYPE	Text (1)	Flags whether the name corresponding to the above TVK is scientific (S) or vernacular (V).
RECOMMENDED_SCIENTIFIC_NAME	Text (85)	The current NBN recommended scientific name for the TAXON_NAME. Taken from an NBN preferred checklist, or other authoritative source.
RECOMMENDED_NAME_AUTHORITY	Text (80)	Naming authority for the RECOMMENDED_SCIENTIFIC_NAME
RECOMMENDED_NAME_QUALIFIER	Text (95)	Any qualifying information for the RECOMMENDED_SCIENTIFIC_NAME
RECOMMENDED_NAME_RANK	Text (25)	Taxonomic rank for name for the RECOMMENDED_SCIENTIFIC_NAME
NBN_TAXON_VERSION_KEY_FOR_RECOMMENDED_NAME	Text (16)	Unique 16-character TVK for the RECOMMENDED_SCIENTIFIC_NAME
DATE_RECORD_ADDED	Date/Time	The date when the record was added to the Nameserver (useful for tracking new entries).
DATE_RECORD_LAST_CHANGED	Date/Time	The date when the record was last updated (useful for tracking changed entries).

Appendix 2: Hints and tips for name matching and assigning NBN TVKs

The following suggestions will hopefully help you to match names used in your datasets to those used by the NBN, and ensure that you assign the correct NBN_TAXON_VERSION_KEYS. Assigning NBN_TAXON_VERSION_KEYS to your data is a prerequisite for providing observational data to the NBN Gateway. Comparison and matching of your data to the Nameserver can either be carried out by eye, or through the use of programmatic queries (e.g. within Microsoft Access).

- Ensure that you match to the name string that is exactly the same as the one used in your dataset. If you spot any errors, you may wish to correct these first (or notify us if the error is within the Nameserver).
- Ensure that you choose the NBN_TAXON_VERSION_KEY for the name and author combination that matches your dataset. i.e. Do not just match on the TAXON_NAME, but match on TAXON_AUTHORITY and TAXON_QUALIFIER as well. The naming authority and qualifier are important and define what the name means and how it is handled within the Nameserver. Hence, picking a different combination to the one that you require could result in a changed interpretation for your name and any attached data. If your dataset does not contain author information, then select the NBN_TAXON_VERSION_KEY for the name that has a blank TAXON_AUTHORITY field. Alternatively, you may wish to select the well-formed version of a name, if you are sure that your data refer to the same taxon. To do this, simply pick the version of the name that has a value of 'W' within the TAXON_FORM field.
- Naming authority (i.e. author) abbreviations as used for botanical and fungal names can be highly variable. Try to match to the name that has the same abbreviation as your dataset, or alternatively you may wish to match to the well-formed version of the name and amend your dataset appropriately.
- When you have matched your name strings to those used by the NBN, make sure that the INFORMAL_GROUP and RANK information are what you expect. Some names can be used at more than one rank and homonyms (names that apply to more than one organism) can occur within both scientific names and vernacular names. For example, *Ctenophora* can refer to a genus and subgenus of flies, a genus of diatoms or the phylum that contains comb jellies. Similarly, Redshank can be used for a moss, a flowering plant or a bird. It is important to check that you have selected the correct usages of your names.
- If you cannot find the name that you are looking for then first check for obvious misspellings, for example transposed letters, or differences in the gender of scientific names. Similarly, we include an infraspecific rank term within the name string, whereas you may not have done so. Differences like these are obvious if matching is done by eye, but will often be missed if automated querying is used to match names. If you still cannot find the name that you require, **then please let us know**; we will check the data and, if appropriate, add the name to the Nameserver and issue you with a TAXON_VERSION_KEY.
- When providing data to the NBN Gateway, you should ensure that the NBN_TAXON_VERSION_KEYS that you assign are for the names that are associated with each observation, not those assigned to the RECOMMENDED_SCIENTIFIC_NAME for these taxa (where this differs). It is important that associated observational data remain attached to their original determinations - the Nameserver within the Gateway will handle any synonymy.
- Finally, please notify us if you spot any errors or omissions, as this will help us to improve the product.