Enhancing data quality of bryophyte records for the National Biodiversity Network

Author
Mark Hill

This report was produced the British Bryological Society under the Defra/NBN Trust contract 2011-2014 to improve data provision, management and coordination in the National Biodiversity Network

April 2012

NBN Contract BBS.a

Verification rule sets for NBN Record Cleaner and recommendations on species whose records should be treated as sensitive

Report by Mark Hill
on behalf of British Bryological Society (BBS)

24 April 2012

Contents

Background

This note accompanies the deliverables submitted by Mark Hill. Work started in February 2012 and was completed in April 2012 following a meeting of the BBS Conservation and Recording Committee on 30 March 2012.

Deliverables based on tender document

1. Specification of essential attribute fields for new records

Database fields currently used by BBS

The information on database fields was sent as part of the tender submission and is listed here as Appendix 1. It is based on fields present in the BRC database, but in a modified form that is (a) flat-file with few special codes and (b) easily understood by non-technical bryologist.

Aggregate definitions

The file *1_Aggregate_definitions.xlsx* is included with this submission so that the definition of aggregates is fully explicit. This table provides a record of precisely the data used to produce the hectad distributions of complex entities such as *Schistidium apocarpum* s.l. The table of aggregates derives from four sources:

0. An existing table of aggregates held in an ORACLE table at BRC;
1. Additional aggregates or members of aggregates that came to light during scrutiny of the ORACLE table;

2. A set of partial synonymies where a taxon is available for recording, although not currently on the official list. An example is *Dicranella howei*, which is arguably an entity worthy of recognition although in the BBS official list it is treated as a form of *Dicranella varia*; and
3. Infraspecific taxa of species.

To derive the new aggregates table, these sources were combined. All entities with qualifiers such as s.str., sensu A.J.E. Smith (1978) or s.l. were examined critically to make sure that they were either in an aggregate (s.str.) or were defined by their membership (s.l.). Several missing aggregates were detected and their membership has now been supplied.

The aggregate records were combined in a table, and those taxa that were members of an aggregate contained within a higher-level aggregate were enumerated. These 'grandchildren' were added to the membership of the higher-level aggregate if they were not already in it. The resulting table was joined to itself to find out if there were further 'grandchildren'. There were none.

2. Documented procedure for dealing with records highlighted by Record Cleaner

As pointed out in the tender, the BBS is about to produce a substantial *Bryophyte Recording Handbook* with about 38 A4 pages. The BBS currently applies the following checks.

1. Set data in the fields outlined in Appendix 1.
2. Code up species and standardize recorder names and herbaria.
3. Check georeferencing, mainly to ensure that given vice-county matches the grid reference, to at least 1-km resolution.
4. Flag up species that have not been recorded from the vice-county since 1960.
5. Flag up species that have a probability of less than 0.3 of being found in the hectad.
6. Where an altitude is specified, check that it lies within limits for 1-km square (monad).

Much detail on checking is available as a download from the BBS website by following Recording > Resources for Recorders > Procedures for checking bryophyte data. More information will be given in the forthcoming *Recording Handbook.* In the mean time, Table 1 provides an overview of the actions required when checking data.

Problem	**Action**
Georeferencing incorrect? [Commonest error in data submitted to BBS is incorrectly specified 100-km square.]	Check map against (a) name of site and (b) vice-county. Search for alternative grid references in a different 100-km square.
Altitude out of range for location?	Treat in the same way as an ordinary georeferencing error.
Identification beyond expertise of recorder?	Ask recorder whether (s)he is confident of record. If no then request that specimen is checked by an expert (e.g. BBS referee). If check not possible, mark the record as doubtful. Allow for possibility that recorder is over-confident; such recorders are rare but do exist.
Recorder probably does not have access to a microscope?	Encourage the recorder to acquire a microscope. In the mean time mark all records except for those in identification-difficulty categories 1 and 2 as doubtful.
Taxon outside known range of occurrence? [Range is not explicitly defined, but a threshold of 0.3 probability of occurrence in the 10km square has been set. This criterion is well tested, having been in use for checking BBS data for about two years.]	Ask recorder whether (s)he is confident of record. Records that are flagged as unlikely by Record Cleaner are potentially the most interesting. Non-experts are recommended to consult an expert member of the BBS about these, preferably the BBS Regional Recorder if there is one. If the record turns out to be a vice-county record, the specimen should be checked by BBS Recorder for Mosses or BBS Recorder for Liverworts.

Table 1. Procedure for dealing with records highlighted by Record Cleaner.

3. Identification difficulty rule set

Results are contained in the file *3_Bryo_identification_difficulty.zip*. This consists of a text file in the format specified for the contract, together with an Excel spreadsheet showing how the categories were derived, with notes on the species. Most of the work for this rule set was done jointly by Mark Hill, Sam Bosanquet and Chris Preston during autumn 2009. Mark Hill revised and updated it in March 2012, adding a category 1, of species that can be identified from photographs.

4. Spatial distribution rule set

A copy of the BBS database was downloaded from BRC in early February 2012. Hectad distributions of taxa based on records since 1950 were calculated. The distribution of aggregates was derived using the definitions in *1_Aggregate_definitions.xlsx*. In addition, 14 'quasi-aggregates' were defined in a separate worksheet. These consist of entities where one component of an aggregate is very much commoner than the other. An example is *Acaulon muticum* s.l. A record of *Acaulon muticum* s.l is deemed to be a record of *Acaulon muticum* s.str. except in hectads where the very rare *Acaulon mediterraneum* has been found. In an extreme case, records of the quasi-aggregate *Cratoneuron filicinum* s.l. must consist almost entirely of *Cratoneuron filicinum*, because the rare segregate *Cratoneuron curvicaule* has not been found in Britain since 1913.

The program Frescalo (Hill, 2012) was applied to the distribution data, including aggregates. Frescalo uses a smoothing procedure to calculate probabilities of species occurrence from observed frequencies in the neighbourhood of each hectad. Taxa with probability of occurrence less than 0.3 are rated as notable, and are flagged when present. The rule set, of about 1.3 M records with probability greater than 0.3, is contained in the zipped file *4_Spatial_distribution_rule_set.zip*. In total 1246 taxa (including genera and aggregates of lower rank) are included in the distribution rule set, which lists probabilities in 2834 hectads. The single text file can easily be divided into pieces, one for each species, using the records beginning "$" as separators.

5. Temporal rule set – seasonal range

The number of bryophytes that can be identified only at a particular season is small. By far the majority of apparently seasonal species are those of high mountains, especially snowbeds. An extreme example is *Andreaea nivalis*, of which 83% of records were in July or August and the remaining 17% in June or September; but *A. nivalis* could surely be found in October if someone were to search for it then. Records, arranged in descending order of 'summerness', are shown in the second worksheet *Month_counts* of *5_Bryo_months.xlsx*. The rather numerous high-mountain species appear at the top. High-mountain species do not require checking for season, because improbable occurrences will almost always be picked up by the spatial check.

The most notable example of a species that is often misidentified at the wrong season is *Bryum caespiticium*. It was in the past substantially over-recorded, so that its true season of apparency is not clear from the database. Flagging out-of-season records of this and a few other *Bryum* species would be useful. The species in question are:-
NHMSYS0000309475 *Bryum archangelicum*
NBNSYS0000036615 *Bryum caespiticium*
NHMSYS0000309486 *Bryum creberrimum*
NHMSYS0000309514 *Bryum salinum.*
Only records from the period May-October should be accepted. Outside this period, the essential details of peristome structure cannot be observed.

6. Temporal rule set – year range

Records of extinct species will automatically be picked up by Record Cleaner, because such species were always exceedingly rare and will be flagged on the grounds of improbability. 28 taxa have not been seen since 1972 (Table 2), and a further 6 taxa have not been seen since 1986. Dates, finders and TVKs are listed in *6_Extinctions.xlsx*.

Taxon	Year	Comment
Not seen since 1972		
Atrichum undulatum var. gracilisetum	1927	
Bryum turbinatum	1947	
Bryum uliginosum	1953	Recent records from Ireland
Cratoneuron curvicaule	1913	
Cynodontium fallax	1868	Single record only, not refound
Dicranum elongatum	1964	Later record by D.G. Long thought to be incorrect
Encalypta brevicolla	1871	Single record only, not refound
Fossombronia mittenii	1972	
Grimmia sessitana	1966	Single record only; disputed taxon
Gyroweisia reflexa	1938	Casual, persisted 1933-1938
Helodium blandowii	1882	According to Hodgetts Red List, last found in 1901
Hypnum revolutum var. dolomiticum	1932	Single record only, not refound
Lescuraea saxicola	1911	
Lophozia longiflora	1956	Taxonomically difficult species
Mielichhoferia mielichhoferiana	1971	Corrie Kander 1996 is doubtful
Neckera pennata	1835	Single record only, probably casual
Orthotrichum shawii	1873	Transient at single site, 1860-1873
Paludella squarrosa	1916	One extant site in Ireland
Philonotis cernua	1961	Last Irish record 1987
Plagiothecium piliferum	1939	
Pohlia crudoides	1968	Single record only, not refound
Pterygoneurum lamellatum	1970	
Sphagnum obtusum	1911	
Tetrodontium repandum	1958	Only 2 records, 1954 and 1958
Tortella limosella	1906	Single record only, not refound
Trematodon ambiguus	1883	Single record only, not refound
Weissia controversa var. wimmeriana	1956	Single record only, not refound
Weissia mittenii	1920	
Last record 1973-1986		
Campylopus atrovirens var. gracilis	1975	
Tayloria tenuis	1977	Mentioned in database as comment to 1974 record
Andreaea rupestris var. papillosa	1978	
Aplodon wormskioldii	1981	
Scapania parvifolia	1981	
Seligeria diversifolia	1984	

Table 2. Bryophyte taxa not found in Great Britain for 25 or more years

Twenty-five species (Table 3) are thought to be introduced since 1800, though there are differences of opinion about the status of *Lophozia herzogiana*. Dates and TVKs are listed in *6_Bryo_non-native_first_dates.xlsx*. The rate of new arrivals rose sharply after 1940 to about 3 species per decade (Figure 1). Most introduced species are so rare that they would be flagged because they are improbable. There is probably little point in putting in an explicit check on the year in which records were made.

Name	**Year**	**Place**	**Finder**
Achrophyllum dentatum	1991	Cornish garden	Rumsey, F.J.
Atrichum crispum	1848	nr Rochdale	Nowell, J.
Bryum apiculatum	2007	Cornwall	Holyoak, D.T.
Bryum valparaisense	2005	Scilly	Finch, R.A.
Calomnion complanatum	1998	Kerry	Labeij, W.
Calyptrochaeta apiculata	1967	Tresco, Isles of Scilly	Paton, J.A.
Campylopus introflexus	1942	Washington, Sussex	Marshall, J.K.
Dicranoloma menziesii	2008	Kerry	Holyoak, D.T.
Didymodon umbrosus	1958	Stanmore, Hampshire	Paton, J.A.
Ephemerella readeri	2006	Lindley Wood Reservoir	Hooper, E.J.
Hennediella macrophylla	1965	R. Mole, Box Hill	Crundwell, A.C.
Hennediella stanfordensis	1958	Lizard	Coombe & Whitehouse
Heteroscyphus fissistipus	1999	Kerry	Long, D.G.
Leptophascum leptophyllum	1964	Brook, Isle of Wight	Crundwell & Warburg
Leptotheca gaudichaudii	1998	Kerry	Labeij, W.
Lophocolea bispinosa	1962	Scilly	Paton, J.A.
Lophocolea brookwoodiana	2004	Brookwood, Surrey	Sheahan, M.C.
Lophocolea semiteres	1955	Scilly	Wanstall, P.J.
Lophozia herzogiana	1986	Hampshire, Woolmer Forest	Crundwell, A.C.
Orthodontium lineare	1910	Overton Hills, Cheshire	Richards, E.A.
Riccia crystallina	1922	Scilly	Duncan, J.B.
Riccia rhenana	1952	Surrey, R Holloway College	Hooper, S.S.
Telaranea murphyae	1961	Scilly	Murphy, R.J.
Telaranea tetradactyla	1970	Wisley, Surrey	Milnes-Smith, M.D.
Tortula amplexa	1973	Moira, Leicestershire	Side & Whitehouse

Table 3. Bryophyte species introduced to Britain and Ireland since 1800

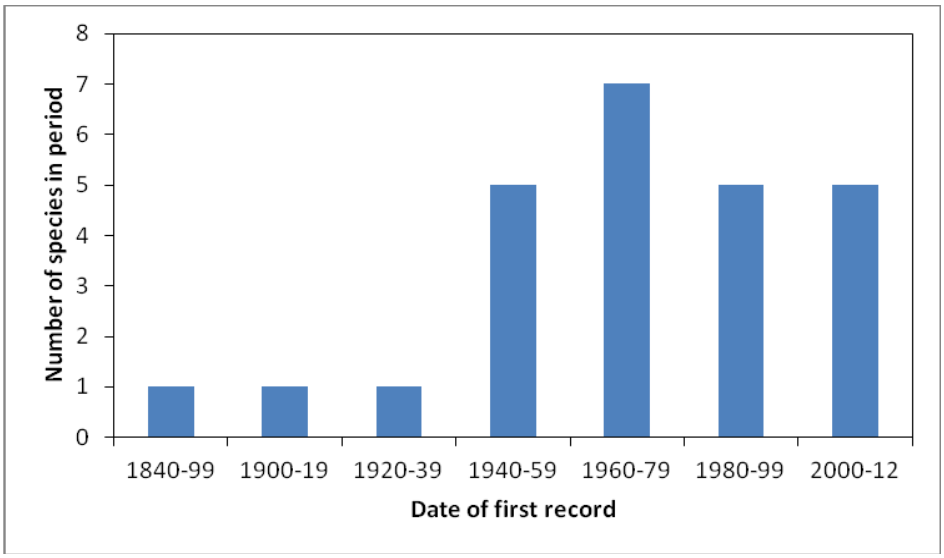

Figure 1. Numbers of first records of non-native bryophyte species since 1840

7. Identify criteria for sensitive records

Almost all BBS data are available at full resolution on the NBN Gateway. Exceptions are:-

- Proprietary datasets supplied by third parties, e.g. LRCs; these datasets are for BBS use only and are not exported to the Gateway unless the proprietor gives approval;
- A very small number of sites where the land owner forbids electronic publication;
- A very small number of records that a conservation agency such as CCW wishes to keep under wraps; and
- 11 species that are thought to be vulnerable to collecting (Table 4).

This list will be reviewed by the BBS Conservation and Recording Committee when it meets in April 2012. The following text will be used in the agenda paper.

Review of sensitive species for NBN Gateway

In the current list of sensitive species, the BBS recognizes three main types of potential harm: damage by trophy-hunters, antagonization of land owners, and trampling damage to vulnerable peatland. Access controls for 'sensitive' data are currently under review by the NBN. In future there will only be two levels of access to data: public and full. The 'sensitive' flag could be used to blur records to a non-sensitive level (e.g. 10 km square) or to remove them from public view, according to the wishes of the data provider, without duplicate records appearing on the NBN Gateway. It is not yet clear whether the NBN will be using more than one type of 'sensitive' flag. The committee was asked to consider the following.

1. Is the current list right or should some species be removed or added?
2. Does the Committee recommend that some species should be accorded variable levels of sensitivity depending on geographical location?
3. Would the Committee be content to see all records of sensitive species mapped at 10-km resolution?
4. Are the arguments against disclosure correctly stated? Are there other criteria that should be applied?

The Committee made the following recommendations.

1. Scottish Natural Heritage expressed a preference for opening up the full data for all bryophyte species. Taking account of this view, the Committee decided to remove 7 species

from the list of 11 (Table 4). Of the four remaining, two are retained on the advice of the Irish National Parks and Wildlife Service, and two Cornish species are retained, reflecting the views of the Committee itself.

2. The Committee was not in favour of variable levels of sensitivity, on the grounds that this would be an unnecessary complication.
3. The Committee would like to see all records of the remaining 4 sensitive species mapped at 10-km resolution.
4. The arguments against disclosure are correctly set out in Table 4.

The information for the BBS dataset on the NBN Gateway 'Bryophyte data for Great Britain from the British Bryological Society held by BRC' has been changed to take account of these decisions. The shorter list of sensitive species will be used for the next upload of the dataset.

Reference

Hill, M.O. (2012) Local frequency as a key to interpreting species occurrence data when recording effort is not known. *Methods in Ecology and Evolution,* **3,** 195-205.

RDB	NAME	BBS C & R Committee decision 30 March 2012	Argument against disclosure	Argument for disclosure
EN	Adelanthus lindenbergianus	REMOVE from sensitive list	Unique site in GB, which might attract trophy hunters. In Ireland it has been depressed by high stocking levels.	Fairly widespread on Islay and more likely to by affected by over-grazing than to collectors
VU	Athalamia hyalina	REMOVE from sensitive list	Land owner may be antagonized, which would make SNH job difficult	Anyone who wanted to find this famous site could do so any way
CR	Bruchia vogesiaca	Retain	Unique site of species; potentially a target for trophy hunters.	Was not refound during a search in 2011
CR	Bryum schleicheri	REMOVE from sensitive list	Unique site in GB, currently the subject of translocation activities. Land owner may be antagonized, which would make SNH job difficult	Unlikely to be a target of trophy hunters; it grows much better in the Alps
EN	Cyclodictyon laetevirens	REMOVE from sensitive list	Cornish locality is a cave, where the species was badly damaged by collection in the past. The single good patch of Cyclodictyon is not difficult to get to and it's also vulnerable if people pull bits off	Localities outside Cornwall need to be known by local conservation staff
EN	Ditrichum cornubicum	REMOVE from sensitive list	Small population of an endemic vulnerable to collecting.	Main threat is ecological succession and cessation of mining.
VU	Jamesoniella undulifolia	REMOVE from sensitive list	Globally uncommon and scattered species. Two of the three Cornish locations for Jamesoniella undulifolia have tiny populations that could be damaged by attempts to collect material	Localities need to be known by local conservation staff
EN	Leiocolea rutheana	Retain	Irish localities sensitive to damage by collectors and trampling and it has only 3 small (known) populations	British localities are mostly well known
	Paludella squarrosa	Retain	Extant locality is in the Republic of Ireland. Site is extremely vulnerable to trampling damage from visiting botanists since it is in very wet fen, partly floating at times	
EN	Thamnobryum angustifolium	REMOVE from sensitive list	A valued endemic, with two locations, potentially vulnerable to trophy hunters. Only a single patch in Derbyshire. It was damaged by collection at a BBS meeting in the 1970s.	Information already mainly in public domain; Tullie House give a tetrad grid reference for Cumbrian locality on the NBN Gateway
CR	Weissia multicapsularis	Retain	Globally threatened species which might be eliminated from its Cornish location by collection. Its small patches are very persistent in exactly the same places from year to year and hence it seems highly vulnerable to collectors.	Locality needs to be watched by local people, not just David Holyoak

Table 4. Bryophyte species recognized as sensitive on the NBN Gateway with observations on those removed from the sensitive list in 2012

Appendix 1. Database fields currently used by BBS

Database field	Explanation
(a1) Source	
Rec_no_ex_source	Serial number in dataset submitted to BRC by BBS
TO_SOURCE_ID	Record key (e.g. key in Recorder 6) of record in source database; null for new records
REF literature	Reference to literature source where this is origin of record
(a2) Record status	
Confidential	Used where land owner or conservation body wants data withheld from NBN Gateway
NS_ID	Native status code (e.g. introduced species that persist on substrate without spreading); rarely used
(a3) Record confidence	
RC_ID	Record confidence code; 1 = checked in expert review, 2 = doubtful, 3 = incorrect, 7 = superseded, 10 = vice-county record needing confirmation
Micr chk	Specifies whether a microscopic check has been made
(b1) What	
Concept	BRC code for the taxon; all taxa added before 2010 should have an NBN TVK, but these are used only for submitting data to the NBN Gateway
Species	Taxon name in current checklist
(b2) Biological	
Fruit	Sporophytes present
Male	Plant with male organs (it may by monoicous)
Female	Plant with female organs (it may be monoicous)
Gemmae	Gemmae present
Bulbils	Bulbils present
Tubers	Rhizoid gemmae, also known as tubers, present
(c1) Habitat	
Habitat_site	Habitat of the site where records are a list from an area
Habitat	Habitat where taxon found, e.g. bark of elder, base-rich mire
(c2) Location	
Alt m	Altitude in metres
Location	Name of location
Grid ref	Grid reference using GB, Irish or UTM (Channel Islands) grid
Tetrad	Single letter; used only when accuracy of record is 2000 m
Quadrant	NE, NW, SE, SW - used only when accuracy of record is 5000 m
Accuracy	Not used for data input - this is inferred from the entries for Grid ref plus Tetrad or Quadrant; this is necessary for data output from database
VC	Vice-county – a number, e.g. 56 for Nottinghamshire
(d) Time	
Day	Day if known; day of beginning of date range if not
Day2	Day of end of date range; null if exact day is known
Month	Month if known; Month of beginning of date range if not
Month2	Month of end of date range; null if exact Month is known
Year	Year if known; Year of beginning of date range if not
Year2	Year of end of date range; null if exact Year is known

(e1) Persons	
RNAME	Name of recorder or group of recorders
Conf/det	Confirmer or determiner; set to null if same as RNAME
Compiler	Compiler of records when multiple records sent to BBS; often the same as RNAME
(e2) Specimens	
Coll No	Recorder's collection number, e.g. Blockeel 40/458
HERB	Herbarium where specimen, if any, exists; PVT for private herbaria
Comment	General comment, e.g. identification note, colony size, capsule maturity, whether VC record etc.