



Improving the quality of botanical records available via the NBN Gateway

Contract report BSBI10-a to BioD Services Limited

R.W. Ellis and K.J. Walker

2011



This report was produced the Botanical Society of the British Isles (BSBI) under the Defra/NBN Trust contract 2008-2011 to improve data provision, management and coordination in the National Biodiversity Network.

The Botanical Society of the British Isles (known as BSBI) is a registered charity in England and Wales (212560) and Scotland (SC038675)
Registered address: Department of Botany Department,
Natural History Museum, Cromwell Road, London, SW7 5BD

www.bsbi.org.uk

Contents

Summary	3
1. Introduction	4
1.1 Background	4
1.2 Objectives	4
2. Methods	6
3. Results	6
3.1 Objective 1: Guidance on what information records should contain	6
3.2 Objective 2: Rules for quality assuring records in the NBN data validation software	9
3.3 Objective 3: Processes for verifying records	12
3.4 Objective 4: Managing and sharing records across the NBN	15
3.5 Objective 5: Identification of sensitive records	20
4. Conclusions	21
5. Acknowledgements	22
6. References	22

Summary

- In the UK botanical recording is well-organised via a number of recording schemes that operate from local (county) to UK-levels. As a result, millions of plant records from throughout the UK are available on the NBN Gateway.
- Generally, data quality is good although improvements, particularly regarding systematic verification, are required. In particular, the provision of online facilities to allow experts to verify records in a more efficient manner.
- A range of mechanisms to standardise recording and improve the efficiency of data verification and data-flow are recommended in this report.
- Baseline information has been provided to form a series of taxon-specific rules to assess the spatial, temporal and identification accuracy of plant records in datasets via NBN validation software. This should enable data compilers to rapidly generate subsets of records that merit scrutiny from the County Recorder/Local Co-ordinator for verification purposes. It is hoped that the accuracy of plant records at local and national (including NBN Gateway) levels will be improved further as a result.
- The flow of botanical records between data providers and the NBN is currently complex, and in some cases inefficient. The development of the Society's own central collation of records, coupled with the recommended data-flow guidelines included in this report, will hopefully help to resolve this problem in the future.
- The guidance on preferred data-flows presented in this report, in particular how and by whom records should be submitted to the BSBI and ultimately the NBN, will hopefully reduce duplication and improve access to more up-to date records in the future.

1. Introduction

1.1 Background

The National Biodiversity Network (NBN) Gateway is the main mechanism for accessing biological recording data in the UK and collates records from a large and diverse group of organisations. The Gateway currently holds over 60 million species records contributed by over 140 different organizations. A key aim of the Gateway is to improve both the coverage of these data (temporal and geographical) and the quality of the records submitted to it. To achieve this, a set of priority taxa has been identified for which work should be undertaken to ensure relatively complete coverage on the Gateway. These are vascular plants, Lepidoptera (specifically butterflies and macro-moths), lichens and birds.

In taking forward work on these priority groups, the NBN is developing data validation software to improve quality control. This is intended to be used by data collators to highlight records requiring scrutiny by experts or where a minimum threshold of information has not been achieved. In order to do this effectively a set of rules is needed, relevant to each taxon group, against which records can be checked. Once such rules have been developed and integrated into the software tool it will be distributed to data collators, together with general guidance on recording and data flow within each species group.

The Botanical Society of the British Isles (BSBI) is keen to improve the quality and flow of vascular plant records submitted to, and available on the Gateway. As vascular plants are one of the priority groups, BSBI were contracted to develop rules for the NBN's validation tool and to provide general guidance on recording and data flow.

The BSBI is a UK charity dedicated to furthering the recording and study of the British and Irish flora. With nearly 3000 members and a large network of volunteer vice-county recorders and local recording groups, the Society is the leading source of high quality botanical information in the UK. These key volunteers are supported by a small team of staff with a wide range of specialist skills and knowledge in species identification, life-history and ecology, biological recording issues and systems, and database management and analysis.

A key strength of the BSBI is its voluntary principle. Since the Society's inception in 1949 its network of 157 vice-county recorders has contributed generously to a number of local and national recording-schemes. BSBI coordinates recording for national atlases, most recently in 2002 (Preston *et al.*, 2002) and vice-county recorders have produced numerous county floras that underpin these atlas initiatives. Structured surveillance surveys include *Local Change* (1987-2004; Braithwaite *et al.*, 2006), and more recently the *Threatened Plants Project*, a stratified random sample survey of 50 of the UK's most threatened species (2008-2012). Local (county) groups also collate high resolution records for nationally and locally restricted species (*County Rare Plant Registers*). BSBI and the Biological Records Centre (BRC) have collaborated to share data and expertise and since 1964 the Society has contributed over 12 million data points of plant (vascular and charophyte) records to the *Vascular Plants Database* (VPDb). All these records are available via the NBN Gateway.

1.2 Objectives

The main objectives were to produce the following:

Objective 1: Guidance on what information records should contain

To define which attributes are essential and which are desirable for higher plant records and produce guidance on the preferred format and content of these records.

Objective 2: Rules for quality assuring records in the NBN data validation software

To develop three separate rules for plant taxa: identification difficulty, geographic range and recorded year range.

- Prepare a classification system for identification difficulty of plant species using categories based on the difficulty of identification combined with the recorder's expertise and need for additional evidence to substantiate a record.
- Prepare baseline distributions for each species to act as geographic rules. Develop these at 10 × 10 km grid square resolution for the UK and provide them in a suitable format for inclusion in the NBN data validation tool. The recent recorded distributions of each species in the VPDb were to be used to generate these baselines.
- Prepare a set of temporal rules defining the acceptable year range of records of each taxon (i.e. the year in which the species was first recorded in the UK and where appropriate the year in which a species became extinct in the UK). The recorded year ranges of each species in the VPDb would be used to generate these baselines.

Objective 3: Processes for verifying records

Guidelines would be produced to ensure the accuracy of vascular plant records submitted to and available on the NBN Gateway. This would include a protocol for submitting plant records to approved taxonomic experts to assure the identification accuracy of new records (based on the classification of identification difficulty) and guidelines for experts determining the accuracy of new species records including how to record decisions and their justification. Finally, a strategy for assessing and classifying the accuracy of existing records already shared through the NBN Gateway would be provided including comments on the NBN's suggested framework for classifying the status of records (i.e. into the categories Correct, Considered correct, Requires confirmation, Considered incorrect, Incorrect and Unchecked).

Objective 4: Managing and sharing records across the NBN

Publish guidance on the preferred processes of verification, validation and data-flows through which new records should be checked, managed and shared between data providers and NBN. This should define who should hold and manage plant records and how these should be shared across the NBN.

Objective 5: Identify sensitive species

Agree and publish information on those species for which full public access to observation records might lead to environmental harm. For this a review of each plant species would be undertaken to identify those taxa that should be considered sensitive. The criteria used to justify the selection would be stated.

2. Methods

Contract work was carried out between September 2010 and August 2011 by Bob Ellis (BSBI Projects Officer) and Kevin Walker (BSBI Head of Research and Development) with the assistance of Trevor James (NBN Trust) for Objectives 2 and 5.

All native and alien higher plant taxa including microspecies of *Hieracium*, *Rubus* and *Taraxacum* as well as recognized subspecies and hybrids were included in the assessment of rules for Objectives 2 and 5, including species that are regionally extinct in the UK. Many alien species and hybrids that have only been recorded a few times in the UK were excluded, as well as all species for which records in the UK are considered doubtful. The total number of taxa was 4720 (2961 native, 1756 alien and 3 aggregates with native and alien taxa).

Species names follow Stace (2010) or Sell & Murrell (1996 *et seq.*) where not included in the former. Note that due to recent nomenclatural changes many names differ from those in the NBN Species Dictionary.

3. Results

3.1 Objective 1: Guidance on what information records should contain

For the purposes of the national recording schemes run by the BSBI, the following data fields are considered essential and desirable for each record. We recommend that essential fields are the minimum common standard that all other collators of botanical records, including Local Records Centres, should adhere to, in order to ensure data quality and efficient flow of records across the NBN partnership.

Essential data fields:

- *Species name* - all records must have a scientific name that follows a common standard. In the absence of an up to date list of British plant names, BSBI recommends using the third edition of the *New Flora of the British Isles* (Stace 2010) as the preferred source for names of British and Irish taxa although this should not preclude the use of names for additional taxa not included in Stace (2010), especially subspecies and critical taxa included in the *Flora of Great Britain and Ireland* (Sell & Murrell 1996 *et seq.*). At the moment not all recorders or recording packages conform to Stace (2010) and therefore it must be accepted that data compilers will have to resolve cases where species names do not match. This may seem straightforward but in addition to species and subspecies, aggregates, segregates and microspecies need to be considered. In general lower taxonomic units such as 'forma' and 'variety' should be avoided. One rule of thumb might be to record to the most detailed level, given (a) the abilities of the recorder and (b) the critical identification features of the plant available at the time of observation. When using an aggregate (i.e. an aggregate of difficult to identify species) it is important that its scope is clearly defined somewhere (i.e. associated with the taxon dictionary). There are very few aggregates (other than genus) in the NBN species dictionary and therefore recorders should be discouraged from recording genus only except where the taxon is critical.
- *Grid reference* – an accurate grid reference is one of the most important attributes of a biological record as it allows a species or habitat to be located spatially. A grid reference

defines the presence within a square at a scale of 1m, 10m, 100m, 1km or 10km by the coordinates of its southwest corner on the national (OS) grid or, in the case of Ireland the Irish national grid. The grid reference of the square in which the plant occurs should be recorded NOT the grid reference whose southwest corner is closest to the occurrence. If there is an extensive colony of a plant that spans more than one square at the preferred resolution, records should ideally be made for all the relevant grid references at that resolution. Site 'centroid' grid references (i.e. those in the middle of a site) should be avoided unless the site falls entirely into the square that the grid reference defines. Alternatively one could use a six figure grid reference (i.e. a 100 × 100 m square) rather than multiple eight figure grid references (i.e. 10 × 10 m squares), or even a 1 × 1 km square (monad) instead of several six figure grid references. A frequently encountered problem occurs when recording a large site which spans more than one 'monad'. The ideal solution is to record the two or more areas of the site separately but sometimes this may be difficult in the field. The BSBI's minimum recommended scale for a record is the tetrad (2 × 2 km resolution) preferably recorded using the 'DINTY' naming convention as shown in the diagram below. In this system each 10 × 10 km square (hectad) is divided into 25 2 × 2 km squares, each with a unique letter code. The standard convention using this system is the hectad followed by the equivalent letter – in this example TL20A, TL20J and TL20W for the occurrence of a species occurring in the tetrads highlighted in orange. Note that a grid reference such as TL2468 implicitly indicates a resolution of 1 × 1 km but NOT a tetrad. The BSBI's preferred resolution of recording varies depending on the conservation interest of the species as set as shown in the table below (Walker *et al.*, 2010):

E	J	P	U	Z	Resolution of recording	<100m	1Km	2Km
D	I	N	T	Y	Nationally rare & scarce species	<input checked="" type="checkbox"/>		
C	H	M	S	X	UK Priority Species (e.g. Red List, BAP)	<input checked="" type="checkbox"/>		
B	G	L	R	W	Other threatened (e.g. Welsh Red List)	<input checked="" type="checkbox"/>		
A	F	K	Q	V	County rare & scarce species	<input checked="" type="checkbox"/>		
					Axiophytes (indicator species)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
					New county or hectad records	<input checked="" type="checkbox"/>		
					Refinds of 'extinct' species	<input checked="" type="checkbox"/>		
					All other species		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

- *Site/location name* – a site name, ideally one that is recognizable from an Ordnance Survey map, is important not only as a check on the grid reference but also in locating the record more accurately if a coarser resolution grid reference is used (i.e. monad, tetrad). For example if a common species is generally recorded at a tetrad scale, adding a site name would allow it to be used subsequently in a species list for that site. Therefore the locality name should be as precise as possible without being unduly long. For example, 'Monks Wood, junction of Badger and Hotel Rides' is preferred to 'Monks Wood, 20 paces along the east side of Hotel Ride, just west of the junction with Badger Ride'. Also it is often better to start with the locality rather than supplementary information, for example 'Monks Wood, field near' rather than 'Field near to Monks Wood' as this allows more efficient sorting of site records. In addition it is always better to use names (and spellings) which appear on OS maps to help make records accessible and consistent and to apply both names and name formats consistently.

- *Recorder name* – the name of the recorder (i.e. normally the observer) of the sighting is essential, in part for verification purposes and in part to define ownership of the record. Recorder names should be unequivocal and include initials (not full names) but exclude prefixes such as 'Dr', 'Rev.' etc. As a general rule group names should be avoided (e.g. BSBI field meeting) and a lead recorder(s) should be attributed to the record. However multiple recorder names are acceptable as long as individuals recorders names are separated by a semi-colon as these are easier for computers to recognize i.e. Walker, K.J.; Ellis, R.G. The preferred format is to have the surname followed by initials (i.e. Walker, K.J.) as this allows more efficient sorting of records.
- *Date* – is an essential part of a record and should be the date that the species was seen and should follow the format dd/mm/yyyy. It is important to use four digits for the year to avoid confusion between centuries. Recording the exact date is preferable but records at month and year and year only are acceptable for commoner species. If the date is a month only then use the format 00/mm/yyyy; if a date is year only use 00/00/yyyy. Date ranges should be avoided as these are difficult to manage electronically and can cause problems when records are assigned to date-classes which start within the date range.
- *Determiner name* – if the species recorded has been identified by a different person from the recorder, usually by an expert, then this person's name should also be provided as the 'Determiner', in a separate data field. The naming conventions should be the same as for recorders (above). Note that in some recording programmes, such as Recorder, determiner is a required field even if it is the same as the recorder.

Desirable data fields:

- *Watsonian vice-county* – provides an unequivocal spatial reference around which botanical recording has been organized for over 150 years. The recommended BSBI numbers and names for these are given on the BSBI website (www.bsbi.org.uk). Difficulties only arise on the borders between vice-counties. If in doubt, you can find out which vice-county you are in by using the mapping utility on the Biological Records Centre website (www.brc.ac.uk/vcList.asp). If applied correctly at the outset vice-county provides a useful check of the record's grid reference (and location).
- *Native status* - this should be the status (native or alien) of the record and not the national status of a species as many species that are native to the British Isles are often introduced within and outside their native range. Perhaps only exceptions need to be recorded. For introduced species status should ideally indicate whether it is planted, established (i.e. self-regenerating) or casual (i.e. not regenerating).
- *Quantity* – the quantity or abundance of a species can be measured in a variety of ways but those that provide an objective assessment are preferred. General notes (e.g. 'several large colonies', 'plentiful' etc.) are better placed in a comments field. For plants it is important that any count or measure should specify what unit is being used and scales of abundance (e.g. DAFOR, DOMIN) should follow standard conventions. Kent & Coker (1992) provide a useful summary of abundance measures commonly used for botanical purposes. For practicality it is often better to estimate the size of a population using broad categories such as <10, 11-50, 51-100, etc. with an indication of the unit

counted (e.g. clumps, rosettes, flowering spikes, etc.). If no abundance is recorded we recommend the use of “Present”. It should be noted however that the scale at which the abundance measure is applied should be appropriate. For example, it is not acceptable to use the DAFOR scale at a tetrad resolution unless the entire tetrad has been assessed (which is highly unlikely). It is better to use the comment field if the abundance refers to an area smaller than the square implied by resolution of the grid reference.

- *Habitat* - is probably only necessary when recording rare, scarce or threatened species. In other circumstances habitat might be better as an attribute of the site (particularly if a particular habitat is being surveyed). The most widely used frameworks for UK habitats include (a) UK Broad Habitats (see Hill *et al.* 2004), (b) EUNIS Level 1 & 2 (Davies *et al.* 2004) and (c) the National Vegetation Classifications described in British Plant Communities (Rodwell 1992 *et seq.*).
- *Altitude* – these are perhaps only relevant if they are of particular note such as the minimum and maximum for a particular area or region. They are probably more relevant in upland areas.
- *Herbarium* - this should be noted if the record has been derived from a voucher specimen or a voucher specimen accompanies the record to allow determination. We recommend that public herbarium names are abbreviated to the standard conventions given by Kent and Allen (1984). If known, specimen reference numbers can also be recorded.
- *Literature reference* - records taken from literature sources (e.g. papers, survey reports, etc.) can be submitted in the same way as a field record although we recommend that the source reference is submitted along with the record in a separate column. Ideally these should be full citations with author name, year of publication, title of paper/book/report, volume and page numbers if a paper, publisher (for books and reports) and place of publication.
- *Comment* – these are clearly optional but the presence of a comment field is often a very useful way of qualifying other record attributes or capturing important information that does not fit in any of the other recommended fields. Examples might include notes on flowering behaviour, management of the site, changes since a previous visit, whether it is a new discovery or refind of a much older record. The comment field can also be used to flag up an unverified record, although in some databases, e.g. Recorder, this is done with a flag in the Determination field.

3.2 Objective 2: Rules for quality assuring records in the NBN data validation software

Three separate sets of rules were developed for use with the NBN validation software. For each taxa information was provided to define rules of identification difficulty, geographic range and recorded year range.

3.2.1 Classification of higher plants according to identification difficulty

UK plants were classified for identification difficulty into four categories according to the difficulty of species identification combined with the necessary level of a recorder's

expertise and the need for additional evidence to substantiate a record. The development of this classification system was undertaken by Trevor James and included the following phases. First, a complete taxonomic checklist of UK plants was extracted from the BSBI Dictionary from which very rare aliens (fewer than three current records on the NBN Gateway) and hybrids were removed (both of which would automatically require a relevant 'expert' in the group concerned to adjudicate on the record). Second, this checklist, comprising 4720 native, archaeophyte, neophyte and casual species, together with all other hybrid taxa, was annotated to indicate to the level of expertise needed to satisfactorily identify it. A description of the four levels of identification difficulty is given in the following table with examples of the species included:

Level	Description	Examples
1	Species that can be relatively easily identified even by inexperienced recorders using up-to-date field guides. This includes 137 taxa for which records would be accepted from anyone without the need for additional evidence.	<i>Bellis perennis</i> , <i>Urtica dioica</i>
2	Species where care is needed for identification (1436 taxa) for which records would be acceptable from a reliable botanist at the local or national level, or from a source that is able to assure that this is the case with reasonable certainty.	<i>Alchemilla alpina</i> , <i>Poa compressa</i> , <i>Stellaria pallida</i> , <i>Vicia sylvatica</i>
3	Species that are difficult to identify (1323 taxa) for which a record would only normally be acceptable if it has been checked and validated by the BSBI Vice-county Recorder for the area concerned (or by another regional expert or other authority recognised by the BSBI).	<i>Ajuga pyramidalis</i> , <i>Carex montana</i> , <i>Lycopodium annotinum</i> , <i>Rosa tomentosa</i>
4	Species that can only be identified following critical assessment (1825 taxa). A record would only normally be acceptable from an expert in the taxonomic group concerned, or from a person recognised by such an authority. Such experts should be known to the BSBI or its related taxonomic community as a recognised authority on the group concerned. This would include, but not be exclusive to, the BSBI's own panel of referees and experts.	<i>Hieracium</i> , <i>Rubus</i> , <i>Taraxacum</i> , most hybrids

This classification applies to Britain and Ireland including the Isle of Man and the Channel Islands and was based purely on identification difficulty, not on the likelihood of occurrence in a particular place. To overcome this 461 taxa classified at Level 2 were 'flagged' to indicate where regional variation in the level at which a record might be acceptable exists. These taxa may require determination by a vice-county recorder or equivalent because they are uncommon in the region (and thus many competent recorders will be unfamiliar with them).

In addition, 3123 taxa were 'flagged' as requiring the collection of specimens for critical examination and/or retention as voucher specimens (guidance on how to collect and submit these is given in Section 3.3.1).

It is planned that this annotated checklist will be validated by a small group responsible for data quality and it is recommended that the classification is reviewed periodically and amendments and corrections made in response to user feedback and changes in the UK flora (e.g. due to taxonomic revisions, species colonisations, etc.).

Finally, we would stress that this checklist can only be regarded as guidance, and that the BSBI, along with its vice-county recorders and taxonomic experts/referees would always reserve the right to 'call-in' records that, to them, seem to be in error, regardless of whether or not a record appeared to have met these recommended levels of checking.

3.2.2 Geographic ranges of higher plants

Baseline geographic distributions for each taxon were defined and provided to Graham French of the NBN Trust in text file format for incorporation into the NBN's data validation software.

Baseline distributions were defined from all hectad records, regardless of year, in the Vascular Plants Database (VPDb) provided by the Centre for Ecology and Hydrology in June 2009. This incorporates the entire *New Atlas* dataset (Preston *et al.*, 2002) plus a few additional checked datasets loaded onto the VPDb loaded since 2002. The taxa used were based on NBN preferred names and matched to the species codes held in the VPDb.

This dataset represents a reasonably accurate assessment of the known UK distribution of most taxa, although some geographic areas and taxa are under-recorded (e.g. recently described, difficult to identify hybrids and critical microspecies, uncommon aliens, etc.). In addition, a number of very recent (post-2000) colonisations are unlikely to be represented. Therefore it will be important to revise the baseline distribution rules over time as more records are gathered.

A total of 7,871 geographic test files were produced from the VPDb, of which 5,655 were populated with hectad information. We included the blank test files, mainly for very rare aliens and hybrids, in order to alert recorders of any new records for them, as we felt such taxa should be reviewed and new test files produced at a later date.

Two problems came to light after testing the geographic files. First, some very widespread species threw-up distribution errors in under-recorded areas where they were probably either correct or insignificant. Second, a few taxa, mainly segregates of aggregates, were frequently mis-matched for a number of reasons and therefore threw-up spurious failures. The following is recommended to avoid these problems in the future:

- Blur the known distribution of widespread species by including adjacent hectads unless these are in a different vice-county
- Develop a method of updating the test files when failures are determined as good records
- Remove the 'empty files' for the moment and investigate errors on a case-by-case basis creating appropriate files as new information comes to light

3.2.3 Recorded year ranges for higher plants

Temporal rules were defined to cover the year ranges of records for each taxon in Britain and Ireland including the Channel Islands and the Isle of Man. A precise start year was available for 1543 native and 1457 alien taxa based on unpublished research carried out by David Pearman. Where a start year was unavailable a default of 1800 was used. Native taxa were only given an end year where they are known to have become extinct in Britain and Ireland, equating to the year when a taxon was last recorded. For all other taxa the default end year was not specified (to signify the present year). Thus potentially important records

of species recorded either before the first published occurrence or after the last known sighting will be highlighted by the NBN validation software.

As with each sets of rules, the temporal rule will inevitably become outdated over time as new species are recorded for the first time, species become extinct or earlier records come to light. Consequently, it will be important to periodically update the baseline over time as more records are gathered.

3.3 Objective 3: Processes for verifying records

3.3.1 Protocol for submitting specimens for expert determination

BSBI panel of referees and specialists

Members of the BSBI may submit specimens for identification directly to a panel of national referees and specialists, but it is perhaps preferable to do so in consultation with the relevant vice-county recorder. Non-members must approach the vice-county recorder who may forward specimens or other relevant evidence to the panel of referees, where considered necessary. Details of referees and vice-county recorders are available in the *BSBI Year Book* and vice-county recorder details are increasingly available on the BSBI web site (www.bsbi.org.uk). The general guidance for submitting specimens is provided below, but note that there may be specific requirements for particular genera or species as outlined in detail in the *BSBI Year Book*, details available on request from vice-county recorders to non-members.

One of the benefits of being a member of the BSBI is having access to this panel of specialist referees. These volunteers are national experts in their particular plant groups and their work, which is done in an entirely voluntary capacity, helps maintain the quality of BSBI projects. Please note that the service is only available for **non-commercial** identifications.

Communicating with a referee

Referees may be contacted by post or email, as shown in the list of referees in the *BSBI Year Book*. Specimens that are few in number may be sent unannounced unless otherwise indicated, but contact the referee to seek agreement before sending a large number of specimens all at once. Note that the address to which specimens should be sent is that shown in the list of referees, and is not necessarily the same as in the list of members.

Plant specimens

Any specimen sent to a BSBI referee should normally be of British Isles provenance and should be accompanied by the following information: locality, including grid reference, name of the collector(s), date of collection, description of features likely to be lost on 'pressing' (e.g. flower colour, plant height, etc.). It is often also useful to include any habitat details or photographs of the plant in situ and/or small envelopes containing detached material such as seeds, flowers, basal leaves, etc. Digital images are preferred by many referees as identification features can be enlarged on screen.

Specimens must have been collected legally: see the *BSBI Code of Conduct* for the collection of plant material for details (www.bsbi.org.uk/Code-of-conduct.pdf).

Referees will usually require dry, pressed material with representative rhizomes, leaves, flowers or fruits any other diagnostic features required for determination (see the appropriate taxon entry for guidance on what to send). Note that specimens of some taxa may need to be preserved differently, for example in fluids, or must contain relevant parts of the plant. Please consult the *BSBI Year Book* for the specific requirements for particular genera or species. If hybrid origin is suspected, specimens should be accompanied by material of or notes on any putative parental taxa at the locality.

In some cases, photographs or fresh material may be required or accepted (again, see the appropriate taxon entry). If sending fresh material, please phone/email the referee before posting material to make sure it can be dealt with immediately. Fresh material should be sealed in a plastic bag and posted as soon as possible following collection.

Additional information

BSBI members contacting a vice-county recorder or referee should quote their BSBI membership number and say whether the sender would like specimens to be returned. Bear in mind that the referee may like to keep the material or send it to a herbarium, as reliably determined material is of greater value when kept in regional or national herbaria, where future workers can study it easily. Collection of duplicates would allow the supplier and the referee to keep one specimen each. The referee may want to publish details of the specimen or add them to the BSBI Distribution Database (DDb). If this is not acceptable, this should be stated when the specimen is supplied. Lastly, senders need to supply their postal address and email addresses, and a stamped, addressed, suitable envelope in case a written reply or return of specimen is necessary.

3.3.2 Guidelines for experts determining the accuracy of new species records

The BSBI recommends that experts who determine records keep details both of the record itself, including data that have been submitted with the specimen (or photographs), and details of the determination. Ideally these should include the following fields supplied by the collector; collector name, date of collection and location (ideally including at least a 6-figure grid reference). In addition the following fields supplied by the expert should also be recorded; taxon, determiner name, date of determination and reason for the decision reached.

The amount of detail in the last item will depend on how critical the identification is. In straightforward cases, nothing may be necessary. If the taxon is highly critical and perhaps likely to be subject to taxonomic revision it may be useful to record in much greater detail how the decision was reached. In these cases it is also appropriate to lodge the specimen either in the expert's own collection or with a herbarium in an appropriate establishment.

The BSBI recommend that all records determined by experts should be sent to the following:

- *Recorder* - normally the sender will have provided a stamped self-addressed envelope for a reply. If this is not the case, then clearly it is up to the expert whether or not they respond.
- *Relevant BSBI vice-county recorder (if not the sender)* - if the specimen was submitted by someone other than the vice-county recorder, then it is envisaged that all the expert

need do is submit the data to the *BSBI Distribution Database* and the data will automatically be made available to the vice-county recorder.

- *BSBI Distribution Database (DDb)* - it will be possible for the data to be entered into the DDb through an on-line facility, either as individual records or as batch updates from the expert's own database. Appropriate instructions for both methods will be made available in a user guide. The submission of the records in this way will ensure that they become available to researchers, appropriate conservation bodies and others via the NBN Gateway.

Clearly there is a danger of duplication here and ideally the vice-county recorder and the expert will agree who is responsible for the curation of the top copy of the data.

3.3.3 Strategy for ensuring the quality of records available on the NBN Gateway

Data quality rules

It is in everyone's interest that the botanical data available through the NBN Gateway and used elsewhere is of the highest quality possible. The BSBI recommends the following general 'rules' are applied to ensure the quality of botanical records:

- Existing datasets should be validated and verified using NBN validation software and any other tools available, and this validation process documented for each dataset
- Doubtful records should be 'flagged' on the Gateway
- Datasets should carry an indication of their level of quality to be displayed in the dataset metadata on the Gateway
- A threshold of acceptable quality should be established (e.g. they should meet the minimum criteria for acceptability indicated by the process of data validation using the NBN validation software); datasets that do not achieve this threshold should be removed from the Gateway until the threshold has been reached through validation and verification.

Validating and verifying plant records using NBN validation software

The following procedures are available and could be effective in checking large amounts of data:

- Basic validation of grid references and date formats
- Validation of records in relation to vice-county boundaries (if vice-county is present in the data)
- Validation by matching grid reference to coastal boundaries (with the exception of *Zostera* spp. all vascular plants are land- or freshwater-based)

Failure of these tests could be used to automatically indicate erroneous records (although there might be some concerns about small offshore islands). The following tests are also available but failures do not necessarily indicate an error; for instance they may well be genuine records for new parts of the country. Records could, however, be automatically marked as unconfirmed or doubtful and then investigated further:

- Verification of dates against first known dates and known extinction dates
- Verification against known distributions. The tests are available but they probably require further testing and refinement before being used for bulk checks. Once

established such checks would also need updating on a regular basis in order to reduce apparent failures due to incomplete test data

- Verification against the known identification difficulty of the taxon with records ‘flagged’ depending on the level at which their identification needs to be checked before it can be confirmed

Mechanism for marking the verification status of a record

It is important to establish a mechanism by which records that are subsequently confirmed as correct observations from outside the known distribution are fed back into the test files of the validation software to update the known-distribution tests. The BSBI recommends that individual records are marked as follows (the equivalent NBN terminology is given in parentheses):

- Verified (Correct)
- Unconfirmed (Considered correct, Requires confirmation)
- Doubtful (Considered incorrect)
- Erroneous (Incorrect)
- Not checked (Unchecked)

Dataset quality

The BSBI recommends that an indicator of quality should be calculated for each botanical dataset held on the NBN Gateway using a scoring system based on the percentage of records marked in the categories given above. The simplest approach would be the proportion of verified (correct) records, but it might be of greater value to give some weight to unconfirmed (and even doubtful) records as well. A dataset with a number of doubtful records, so long as they are marked as such, is probably of more value than one that is unchecked. The threshold of acceptable quality would need to be agreed by all parties, but the BSBI recommends a high level threshold. The BSBI would also welcome a clear indication of data quality for all botanical datasets on the Gateway.

Implementing this strategy

Clearly the testing of existing data is a matter for negotiation between the data owners, data providers and the NBN Trust. The BSBI will maintain and improve the known-distribution test files for the validation software but would not expect to carry out validation and verification on behalf of other data owners, unless contracted to do so.

3.4 Objective 4: Managing and sharing records across the NBN

Botanical records available through the NBN Gateway come from a variety of different sources, with very little or no central coordination of data-flow. Consequently the flow of botanical records between organizations and then onto the NBN is often inefficient. For example, datasets may pass through several routes increasing the likelihood of duplication, or near-duplication, due to the editing of records by different data collators. This is often compounded further by transmission of data in a variety of different formats. A ‘preferred’ dataflow model for botanical records is therefore highly desirable, and essential if data mobilization via the NBN Gateway is to be improved in the future.

3.4.1 Recommended data flow and validation procedures for new plant records

BSBI data-holdings and management

The BSBI has a long-established network of vice-county recorders who act as the local custodians of vascular plant data in their counties. Until relatively recently, with a few exceptions, data were collated in card indexes and on hectad (10Km square) master cards produced to enable mapping of all British and Irish taxa at the hectad level as part of national atlas initiatives. Data were digitised on behalf of the Society by the Biological Records Centre (BRC) primarily for the *Atlas of the British Flora* (1962) and the *New Atlas of the British and Irish Flora* (2002). The BRC also collates data from sources other than the vice-county recorders which are available to the society (and others) and researchers. This database, which is available via the NBN Gateway, is known as the 'Vascular Plants Database' (VPDb).

More recently, particularly during and since the *Local Change* project (2003-2004), BSBI data have been collated centrally and stored electronically by the vice-county recorders themselves. Since the onset of this project, the recording package MapMate has been supplied (and supported) by the BSBI and the data collated centrally in a central MapMate Hub (now split into 6 regional hubs to increase storage capacity). In addition the BSBI holds copies of vice-county recorder data from a variety of other recording packages including Erica, Recorder, Biorecs and DMap, primarily as backup, as well as national collations on specific groups of species such as the *Threatened Plant Database* (TPDB), some critical groups such as dandelions, brambles and hawkweeds and published floras that are yet to be added to the BRC database. Currently there are about twelve million records on the VPDb and eight million on the MapMate Hub but with a little duplication between the two.

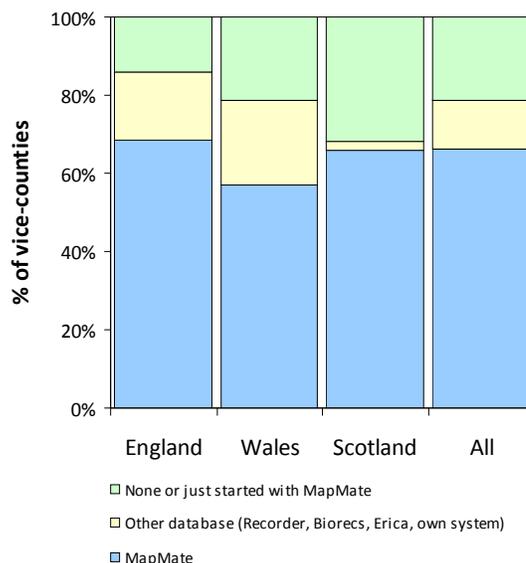


Figure 1. Database systems used by BSBI vice-county recorders in England, Wales and Scotland (as of March 2011).

Over the last few years the BSBI has significantly increased its capacity to handle electronic data through the development of its own central database - the *Distribution Database* (DDb). This is an online facility that draws together all the data across BSBI holdings to better facilitate data interrogation, analyses, dissemination and validation and verification. However, it is important to note that the DDb is a 'working-set' of records, as distinct from

the VPDb which is intended to host finalized, complete data-sets. A central feature is the ability to interrogate all the records for a vice-county or species which can be queried, mapped and edited online by authorized users. The DDb is still under development but is currently (August 2011) available to authorized users (mainly VCRs and partner organizations). Current development work includes the loading of datasets, improving functionality, particularly with respect to querying, mapping and online editing and the development of automated validation and verification procedures. Facilities for online records submission are planned in the near future.

3.4.2 Recommended data flow model

A simplified data flow diagram showing the recommended routes through which BSBI and non-BSBI plant records should 'flow' is given in Figure 2 below. Currently the majority of records originate from local botanists, mainly VCRs (or local groups coordinated by VCRs), who submit datasets to the BSBI directly either through the MapMate Hub or as one-off datasets when surveys or floras are published. At present these are manually collated into the DDb although in the future these routines are likely to be automated. A smaller subset of records, originating from non-BSBI sources, also reach the VCR from Local Record Centres, although the volume is highly variable depending on vice-county. In addition, small volumes of data enter central collations (either DDb or VPDb) from national BSBI referees or national surveys. Currently only 'completed' datasets (i.e. published) are forwarded from the BSBI to the VPDb annually (including published MapMate Hub datasets but not those that are considered to be 'work-in-progress'). The entire VPDb is then periodically loaded onto the NBN Gateway.

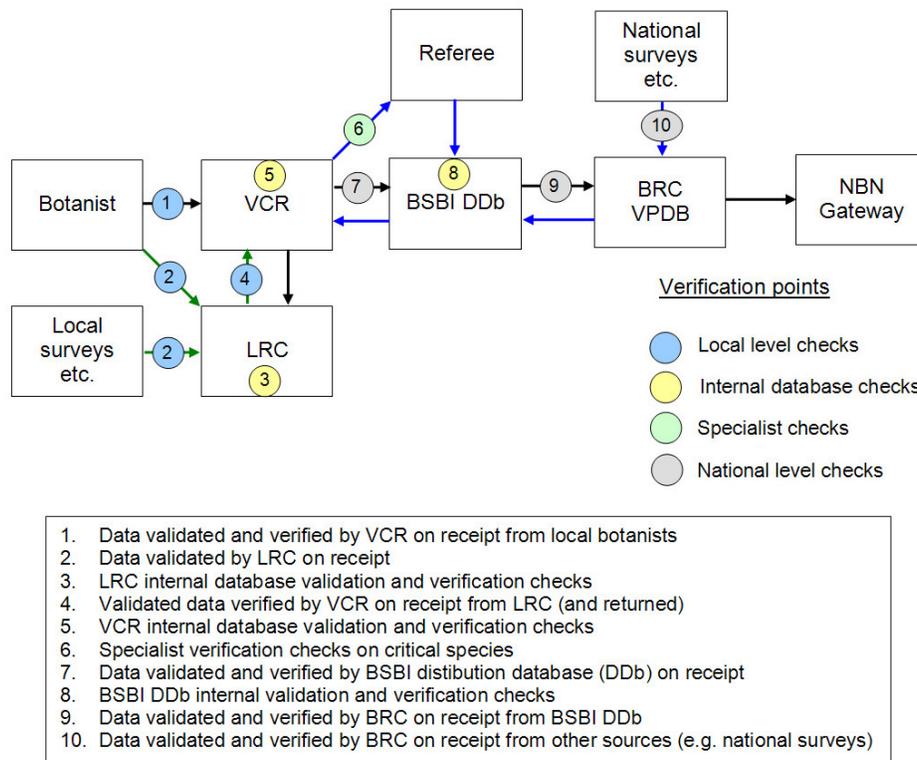


Figure 2. BSBI simplified data flow diagram showing possible points at which validation and verification might be applied.

BRC and NBN Gateway

At present, only BSBI data that has been forwarded to the VPDb is available through the NBN Gateway. In the future it is envisaged that the entire DDb will be forwarded to the VPDb, thereby substantially increasing the volume of recent records available through the NBN Gateway. Direct access to the DDb will be made available to key partners, with the caveat that a proportion of the data will remain unchecked.

Local Record Centres (LRCs)

BSBI encourages vice-county recorders to enter into data sharing agreements with their local LRCs, with some emphasis being placed on data quality and data exchange. The BSBI recommends that LRCs only pass on validated records to VCRs for verification. In some cases this may involve an unacceptably high workload for the VCR who is a volunteer, often with limited spare time. Some agreement therefore may be needed to restrict the species for which verification is likely to be critical, for example species included in a County Rare Plant Register (CRPR), species that are difficult to identify, axiophytes, species with fewer than say 10 vice-county records etc. It is also possible that the NBN validation software may help, especially if local verification test files are developed and applied by both the LRC and VCR. Ideally, the VCR and LRC share all validated and verified data but the path to the NBN Gateway should be made clear at the outset to avoid excessive duplication. In an attempt to avoid duplication the BSBI recommends that all verified records are submitted to the NBN Gateway via the BSBI (LRC>VCR>DDb>NBN) whereas all others should go via the LRC.

3.4.3 Validation and verification procedures

Botanical records submitted to the NBN Gateway via the BSBI are subject to validation and verification at various points along the dataflow model presented (Figure 2). Probably the most important are the local checks carried out by the VCR or the LRC where the record has come from a different source. These provide an initial filter on third party records collected during commissioned surveys (as part of development schemes, etc.). In addition VCRs can provide an important validation 'service' for data held by the LRC, and in some cases VCRs and LRCs employ automated checks within their own data-holdings (now made much easier with the development of the NBN validation software). National referees provide a further vital check for taxa requiring expert determination and internal checks within national databases (DDb and VPDb).

Currently there are differences in the level of checks implemented in the VPDb and the MapMate Hub. Records on the VPDb have been carefully checked during the production of national atlases and can therefore be viewed as 'completed' (i.e. published). The VPDb also attempts to code each record as to its verification status (i.e. whether it is thought to be dubious or not). Although carefully compiled, most data in the MapMate Hub have not been through such a systematic process of validation and verification. Consequently datasets in the MapMate Hub are best viewed as 'work-in-progress'. This does not necessarily mean they are of poor data quality, just that routine checks have not been applied systematically.

The BSBI is keen that its own data is of the highest quality and so a system of on-going validation and verification is being developed as part of the DDb. This is currently being achieved in two ways:

Automated checks

The DDb is duplicating many of the procedures included within the NBN validation software namely taxonomic, temporal and spatial (geo-referenced) tests against known values. A number of basic checks that are currently being implemented include:

- Is the name valid?
- Does the grid reference occur within the VC?
- Is the grid reference in the sea?
- Is the record within the known hectad distribution of the taxon?
- Is the record within the known VC distribution of the taxon?
- Is the record within the known date range of the taxon?

Online editing of records

One of the most powerful aspects of the DDb is the ability to edit records online by authorized users. This is applied at a number of levels within 'workspaces'. Those users with the highest administrative rights for a vice-county (i.e. VCRs) or taxon group (i.e. referees) can mark records as 'confirmed', 'dubious' etc. or alternatively other 'trusted' users can do this with a justification (comment) which can then be 'signed off' by the administrator for that area or group of species. Once this has been completed there is an option to hide these dubious records from view to others.

3.4.4 NBN validation software

NBN validation software provides a powerful and easy to use utility for verifying records and is likely to be incorporated into verification procedures across the dataflow model shown in Figure 2. Significant parts of the rule set used by the software have been developed as part of this contract and clearly its success will depend on the BSBI providing periodic updates to the test files, especially geographic range data, as new information is made available. The BSBI recommends that the Society receives an annual contract from the NBN Trust in order to carry out this work.

Clearly the use of validation software by others will 'flag up' lots of records requiring further investigation. However, it seems unlikely that VCRs will be able to provide a complete 'service' for checking all these records, with the exception of more interesting taxa for which a VCR would probably want to have the final say. In the case of commercial queries the BSBI recommends that VCRs should charge for any services rendered although most VCRs are unlikely to take on such work. In these cases requests could be dealt with centrally via the DDb.

3.4.5 Online recording

Online submission of records is likely to revolutionise the management of biological data over the coming decade, potentially removing the need for recorder's to maintain their own datasets locally. The technology to store, manipulate, edit and submit data online (held on a central 'warehouse' server) is rapidly advancing, and the DDb has followed this model to ensure it can meet demands of recorders over the coming years. Ideally the Society would like to see their VCRs managing the 'top copy' of their datasets via the DDb as this rationalizes the flow of data as well as validation and verification needed. In addition, it provides a more efficient way to check potential errors or report on dubious records.

Currently there is no online data entry facility on the DDb and it is likely that recorders will continue to use MapMate or other recording packages for the foreseeable future. However, we would hope to implement such a facility as part of the ongoing development of the DDb over the next 2 years. This would utilize the current 'work space' utility of the DDb whereby data can be quarantined into a private bay before being signed off by an authorized person for that area or group of species, and then made publicly visible to all other users. In addition, this would ideally include an alert system that told a VCR or expert when data for their area or group of species had been submitted and was ready for validation.

3.5 Objective 5: Identification of sensitive records

The release of high resolution records of certain species could potentially lead to environmental harm, either because a species is rare, collectable, sensitive to disturbance or in most cases a combination of all three. In addition, some taxa are highly visible, especially when occurring on accessible sites. The BSBI therefore recommends that records of these species should not be released to the general public if they could lead to the identification of colony locations.

A review of all British and Irish native and archaeophytes was undertaken by Trevor James to identify those taxa that should be considered sensitive using the following four criteria:

- *Desirability* - how collectable is the species?
- *Accessibility* - how accessible are populations of the species generally?
- *Visibility* - how easy is it to find populations even without a 6-figure grid reference?
- *Fragility* - what is the risk of damage to the plant and/or its habitat whilst looking?

Initially these criteria were used to quickly select species that were thought to be most at risk from environmental harm caused by release of records. This produced a shortlist of 65 species, mainly rare orchids and ferns that have suffered from excessive collecting in the past. These species were then scored against the above criteria from 1 (low) to 5 (high) and an 'overall risk score' (out of 20) calculated by summing the individual scores. The twenty species included in the table below were assessed as having very high scores (>15) and the BSBI recommends that high resolution records for these species are not made publicly available. However, it must be recognized that the sensitivity, as set out above, may change with time, and must be kept under review.

Although the initial selection was largely 'subjective', the species were identified in relation to potential negative impacts of different kinds. The final scoring was more 'objective' although the positioning of any one species against a particular criterion was inevitably to some extent subjective, based on personal experience and understanding of its habitat requirements.

A key weakness of this assessment is that, in certain circumstances, at a local level, other species might be deemed sensitive because of local factors, local rarity, local attractiveness etc. In these cases, the BSBI's vice-county recorder would want to exercise judgement on such species, regardless of how this national (UK) assessment had classified them. However, it would be difficult to administer this from a national perspective and so it must be left to vice-county recorders to administer such restrictions at a more localized scale.

Scientific name	Common name	Risk score
<i>Cypripedium calceolus</i>	Lady's-slipper	18
<i>Dryopteris cristata</i>	Crested Buckler-fern	18
<i>Orchis militaris</i>	Military Orchid	18
<i>Orchis purpurea</i>	Lady Orchid	18
<i>Cephalanthera rubra</i>	Red Helleborine	17
<i>Draba aizoides</i>	Yellow Whitlowgrass	17
<i>Himantoglossum hircinum</i>	Lizard Orchid	17
<i>Orchis simia</i>	Monkey Orchid	17
<i>Cystopteris montana</i>	Mountain Bladder-fern	16
<i>Diapensia lapponica</i>	Diapensia	16
<i>Gentianella ciliata</i>	Fringed Gentian	16
<i>Ophrys fuciflora</i>	Late Spider Orchid	16
<i>Primula scotica</i>	Scottish Primrose	16
<i>Arabis alpina</i>	Alpine Rock-cress	15
<i>Astragalus alpinus</i>	Alpine Milk-vetch	15
<i>Clinopodium menthifolium</i>	Wood Calamint	15
<i>Epipogium aphyllum</i>	Ghost Orchid	15
<i>Saxifraga hirculus</i>	Yellow Saxifrage	15
<i>Trichomanes speciosum*</i>	Killarney Fern	15
<i>Woodsia ilvensis</i>	Oblong Woodsia	15

* sporophyte.

4. Conclusions

In the UK botanical recording is well-organised via a number of BSBI recording schemes that operate from local (county) to UK-levels (Walker *et al.*, 2010). As a result, millions of plant records from throughout the UK are available on the NBN Gateway, both from BSBI recorders, and, in some cases, also direct from other local data compilers such as LRCs. Generally, data quality is good although improvements, particularly regarding systematic verification, are required. In particular, the provision of online facilities will allow expert volunteer County Recorders/Local Co-ordinators to verify records in a more efficient manner. The flow of records between the BSBI, other data collators and the NBN is currently complex, and in some cases inefficient. The development of the Society's own central collation of records, coupled with the recommended data-flow guidelines included in this report, will hopefully help to resolve this problem in the future.

A range of mechanisms to standardise recording and improve the efficiency of data verification and data-flow are recommended in this report. In particular, baseline information has been provided to form a series of taxon-specific rules to assess the spatial, temporal and identification accuracy of plant records in datasets via NBN validation software. This should enable data compilers to rapidly generate subsets of records that merit scrutiny from the County Recorder/Local Co-ordinator for verification purposes. It is hoped that the accuracy of plant records at local and national (including NBN Gateway) levels will be improved further as a result. Guidance on preferred data-flows, in particular how and by whom records should be submitted to the BSBI and ultimately the NBN will hopefully reduce duplication and improve access to more up-to date records in the future.

5. Acknowledgements

Many of the ideas included in this report have been developed through discussions with numerous botanical colleagues over many years. In particular we would like to thank Alex Lockton, David Pearman, Jim McIntosh, Trevor James and Tom Humphrey for their comments and advice on various aspects of this report. The authors are particularly grateful to David Pearman for access to unpublished research on first dates and Trevor James who carried out a large aspect of the work in a private capacity (identification difficulty, sensitivity to environmental harm). We are also grateful to Graham French for technical support and to Oliver Grafton and latterly Paula Lightfoot of the NBN Trust for their guidance and support during this contract.

6. References

- Braithwaite, M.E., Ellis R.W., Preston C.D. 2006. *Change in the British Flora 1987-2004*. Botanical Society of the British Isles, London.
- Davies, C., Moss, D., Hill, M.O. 2004. *EUNIS habitat classification revised 2004*. Centre for Ecology and Hydrology, Monks Wood.
- Kent, D.H., Allen, D.E. (comps.) 1984. *British and Irish herbaria*. Second edition. Botanical Society of the British Isles, London.
- Kent, M., Coker, P. 1992. *Vegetation Description and Analysis*. John Wiley & Sons, Chichester.
- Hill, M.O., Preston, C.D., Roy, D.B. 2004. *PLANTATT. Attributes of British and Irish Plants: Status, Size, Life history, Geography and Habitats*. NERC Centre for Ecology and Hydrology, Monks Wood.
- Preston, C.D., Pearman, D.A., Dines, T.D. 2002. *New Atlas of the British and Irish Flora*. Oxford University Press, Oxford.
- Rodwell, J. 1992-2000. *British Plant Communities. 5 Volumes*. Cambridge University Press, Cambridge.
- Sell, P. & Murrell, G. 1996- *Flora of Great Britain and Ireland. Volumes 3-5*. Cambridge University Press, Cambridge.
- Stace, C. 2010. *New flora of the British Isles*. Third edition. Cambridge University Press, Cambridge.
- Walker, K.J., Pearman, D.A., Ellis, R.W., McIntosh, J., Lockton, A. 2010. *Recording the British and Irish flora 2010-2020*. Botanical Society of the British Isles, London (<http://www.bsbi.org.uk>).