

Guidance for editing records in the BSBI Distribution Database

Contents

Introduction.....	2
Validation States.....	2
Only ever add new information.....	2
The database is a store of data, not an interpretation of it.....	3
“Correcting” records.....	3
Record Sources.....	3
Grid references and names.....	4
Redeterminations.....	4
One person’s trivial information is someone else’s essential data.....	4
Document every change.....	5
Don’t get hung up on duplicates!.....	5
Frequently Asked Questions.....	6
Whose records should I work on?.....	6
Where do I start?.....	6
What should I do if the incorrect record came from MapMate?.....	6
When should I mark a record as “doubtful”?.....	7
When should I reject a record?.....	7
What should I do if the grid reference doesn’t match VC.....	7
What should I do if the grid reference doesn’t match site name.....	7
What if there is no site name?.....	7
What if there is no recorder?.....	7
Further Help?.....	7

Please send corrections and comments to Quentin Groom (quentin.groom@bsbi.org). Our use of the DDb has only just begun and it will take time to discover all the ways we can use it.


Introduction

One of the most important roles of a vice county recorder is validation and verification of records. This includes checking that dates are valid and sites, grid references and vice-counties are consistent (validation); that the records are taxonomically accurate and as complete as possible (verification). This important quality check is what makes our data a trusted source of botanical information. And it is critical to the BSBI's reputation as purveyor of quality records to the conservation and scientific communities.

Having our data in a single database has considerable advantages for the checking, display and analysis of data. Inconsistencies become much more obvious when records can be compared side-by-side. A single data store also helps us to maintain a consistent taxonomy and to credit the right people for their work. Nevertheless, the DDb is not intended to replace programs such as MapMate, which plays a crucial role in data entry and local data management – now and for the foreseeable future.

Here are some general guidelines on how to curate records in the Distribution Database (DDb). However, there will always be difficult cases and grey areas. The best way to deal with them is to ask questions on the [DDb message board](#).

Validation States

Records can have one of five validation states. The validation state symbol for each record is listed in the Record column on the left hand side of a results table. The text in the record is coloured as below. If you hover the cursor over the question mark  to the left of the word "Record" in the column header a helpful key will pop up.

 Confirmed

 Unchecked / not verified

 Duplicate

 Doubtful

 Rejected

 Needs checking

Most records are given the validation state of "unchecked" when they enter the DDb and can then be given one of four validation states through editing. Records that have been checked are given the "confirmed" state; the "rejected" state if they are rejected and "doubtful" if they need to be flagged for further checking. Records can also be marked as duplicates, though this is not entirely necessary.

Important functional differences are that "confirmed" and "unchecked" records are mapped by default, whereas, doubtful, rejected and 'needs checking' records are not. Also, doubtful and rejected records are not retrieved in a search by default. There is no functional difference between records flagged as doubtful or rejected, except that doubtful is intended to mark a record for further scrutiny.

Only ever add new information

This is one of the most important aspects of responsible data management. This is why we reject records, rather than deleting them. It allows you to see what the original record looked like and the reason why things changed. Fortunately, the DDb takes care of the version history of a record for you. If you edit a record, the old data is not lost and the original information is kept. Nevertheless,

you can help the traceability of information if you comment on your edits in the places provided. This will remind you and inform others of why a correction was made. For example, if you reject a record because you think the recorder misidentified it, then write something like “rejected because this species is highly unlikely in the county, is easily misidentified and no evidence was provided”.

Never remove information with which you don’t agree without explaining your reasoning!

The database is a store of data, not an interpretation of it

The published Atlas of the British and Irish Flora does not contain maps of records; it is an interpretation of the records at one point in time. On the other hand the maps in the DDb are just maps of records and are constantly changing and lack interpretation. The contents of the DDb are derived from many places, including MapMate, the Vascular Plants Database, spreadsheets and publications. We should not expect the DDb maps to perfectly reflect what we believe is the distribution of a taxon. The difference may be slight for many species, but in some circumstances it is an important distinction. For example, for rare native species, we might not want to map deliberately planted occurrences in an atlas, even though these sorts of records are valuable when trying to understand how plants are dispersed by mankind.

“Correcting” records

Consult the original recorder and/or check the original recording card wherever possible to clarify any apparent mistakes that you wish to correct before making any changes. This helps to ensure that your change is not introducing a new error and that the DDb copy of the record is more likely to stay consistent with an original data set held by the recorder. Never make changes that might transform a record into something different from the original intent of the recorder. If you disagree with a record it is better to reject it altogether rather than to edit it to fit your conception of reality.

Don’t try to “clean-up” records so you get the map that you think is “correct”. If you need a specific map, either use the filters in the DDb to select only those records you want to plot, or alternatively download the data and generate the maps in another mapping program (DMap, QGIS, DIVA-GIS or ArcGIS).

Record Sources

In general, corrections should be made to the top copy (original source) of the records.

If the original source of the record to be corrected is your MapMate – then you MUST correct it in your MapMate and sync through to the hub. Such corrections will update the DDb within a few weeks. If the record is from another MapMate centre then ask them to correct the error and sync the record to you and / or to the hub. Likewise, if you periodically supply updates to the DDb using another recording system, like Recorder, then use it to make the correction.

Since the DDb now holds the top copy of the Vascular Plant DataBase (VPDB) and Threatened Plant DataBase (TPDB) records from those sources should be corrected directly using the DDb, where appropriate. Other data sets that are not being actively maintained outside the DDb can generally be safely edited in the DDb – but if in *any* doubt please post a query on the Message Board.

Grid references and names

It is an important principle in science that you can never know anything; you can only ever measure something with a degree of uncertainty. An enthusiastic recorder might buy a new top-of-the-range GPS and then go into the field to gather the “correct grid references” for rare plants and then edit the historic records in the DDb to their new “correct” grid reference. Of course, that recorder might be unaware of population changes that have occurred, old populations that have since become extinct, of sites with the same name but in different localities etc. Unless someone has made a genuine, significant error in a grid reference, do not correct it to what you think it should be. What is a genuine error? Frequent examples are where someone has used the wrong grid reference prefix such as NZ instead of NY or ST instead of SO; also, cases where grid references are at sea or in some other impossible location.

Sometimes the correct grid reference is obvious from the original data, for example, “Hole Mill” can’t be at [NZ8989](#) as this is in the North Sea. The recorder must have intended to write [NY8989](#), where Hole Mill can be seen on OS maps. This is a good reason to give a record a site name when it is originally made.

Other cases are less obvious; [NZ333779](#) is in the sea but the site name is “Seaton Sluice road-side” which, at its closest point, is about 600m away. However, the site name could refer to anywhere on 5 km of road. Wherever possible, go back to the original recorder and ask them to either correct it or to give you a correction. If you are forced to correct the record you have two choices: either reject the record or enter a new grid reference. You might reject it if the location information is too vague or ambiguous, particularly if the record is also deficient in other aspects such as a vague date, an anonymous recorder etc. If you decide to change the grid reference, select a grid square that is large enough to encompass all of the possible sites. In this case “Seaton Sluice road-side” could mean many places covering several tetrads, so the best we can do is to change the grid reference to [NZ37](#) (hectad), even though this unit is not recommended for general recording.

Whatever you decide, use the comments boxes to explain your decision.

Redeterminations

The identification given by a recorder and/or determiner is linked to their name; you should not change it. If, for example, you know that the record is of a particular subspecies, don’t just change its name, add a new determination and explain your reasoning in the comments box. You can then change your determination to the preferred one. Take care however, as your name will be linked to this new identification.

Adding a new determination on a record where there is no specimen is dicey and should be used cautiously. If you’re unsure of identification, it may be safer to mark the record as “doubtful” or “rejected”, rather than guessing at the actual identification.

One person’s trivial information is someone else’s essential data.

People use plant records for more things than just creating distribution maps. They are used by taxonomists, geneticists, statisticians, historians, social scientists, ecologists and many more besides. While we don’t have to collect data with these people in mind, we should not remove information we consider trivial. Small comments associated with records often give valuable clues to the origin, population size and persistence of plants at a site. Accurate dates are required for phenology and full

recorder name help document the history of biological recording. Take care of these data, even if they are not useful to you. When you add comments don't use abbreviations. These soon become indecipherable and are not as obvious as you may think to the wide variety of people who use these records.

It is tempting, for example, to change site names so that all records for a site come under the same name, but the original site name is important. It helps us understand the history of the site. These names link historic records to old books and maps and while we think of place names as stable, there are many cases where sites have moved, shrunk or grown in time. These are part of the seemingly trivial information that we should preserve.

Document every change

The DDb has many places to add comments and explanations. Use these to document your changes. This will help you, but also those who follow you. If you take the time to validate a particularly unusual record, make sure you document that you have done this. Then someone else won't repeat what you've already done or, worse still, reject the record in ignorance.

Add any literature references, herbarium codes and specimen numbers to a record. These are some of the best forms of documentation.

Don't get hung up on duplicates!

People frequently get preoccupied by duplicates, yet they are more of an annoyance rather than a problem. Duplicates are handled with ease by computers and will generally not be evident when mapped. People, however like to see one record for one person, date and place. The problem is that duplicates often hold complementary data. So you can't just reject one in favour of the other without first combining the information. Merging records is not always easy. Sometimes there are clashes in the data, for example, where the same record has been derived from two different publications. It is important not to lose this information, so the reference on one of the duplicate records needs to be entered on to the other one. This is, however, a lot of work for a limited benefit.

Duplicates often arise when the same data has been computerised by two separate digitisation programmes often with very different aims. For example, when the data for the New Atlas was captured, in many cases only the absolute minimum was digitised to generate the dot maps; just taxon, hectad & date-class. In many cases, more comprehensive digitisations have subsequently captured fuller details. Another example is where a Flora and herbarium are digitised separately. The Flora may have derived some of its data from the herbaria. Nevertheless, these duplicates should be kept, because we might want a complete list of the contents of the herbarium or all the records from a flora.

Even identifying duplicates can be difficult, particularly where people have lumped records into date ranges or if the recorder is anonymous.

In general, be relaxed about duplicates. Try not to create them and do not reject them without due consideration.

Note: you can hide duplicates by scrolling to the bottom of the list of records the DDb has returned for your search and ticking the "hide" button. This is not meant to be a rigorous duplicate hiding

mechanism. It only looks for records with the same taxon, date & grid ref. Records with more details are preferred. Accordingly, a record with a date of 1930-1970 at SD59 is a duplicate (and will be hidden) of a record of the same taxon on 14/5/1962 at SD548978.

Frequently Asked Questions

Whose records should I work on?

First work on your own records and make sure they are correct, complete and unambiguous.

The Vice County Recorders (VCR) should work on their own counties in consultation with their local recorders. When people submit records to you, make sure you get their contact details so you can query records later.

If you discover records which are incorrectly attributed to your VC, correct their VC and get in touch with the correct VC recorder to tell them. Always consider that the older a record becomes, the harder it becomes to accept or refute. There are numerous cases of old records that lack complete documentation. They continue to frustrate analysis and confuse people. Try to keep on top of the verification of recent records as these are the easiest to check.

Taxon referees should work on their own specialties in consultation with the relevant VCRs.

Try to come to agreement, but consider that if a record is contested and it can't be supported by a refereed specimen or photograph then it can't be accepted just because you say so.

Where do I start?

If you have access the summary My County report for your county, then have a look records linked to from the report's *data validation* tab. The report page links to subsets of your data that would be worth scrutinizing more closely. This includes records where the system's automated checks have detected possible problems and also records of rare taxa or singleton occurrences. Break up the validation task into small manageable chunks.

Focus on finding mistakes initially. If you find an error try to understand how it happened – often you can see a pattern that can lead to other records with the same problem, which you easily correct at the same time. For example, if you spot a site with a mistyped grid-reference, you'll often find that a large set of records loaded at the same time or share the same mistake. Weird taxon mistakes may be the result of data entry errors (particularly if data was entered using numerical codes) – if you spot one such mistake then have a look to see if it was repeated more widely.

There are bulk-editing tools to help you edit many records at the same time. If you would like help correcting a batch of records or if you have spotted a widespread mistake and don't know how best to correct it then please ask for assistance on the message board.

Do I need to verify all the records from my county?

In general no. Prioritise records that are interesting, i.e. occurrences of rare taxa; records that have unexpected locations; isolated records (spatially or temporally). Confirming every record of species from a site where it is well-known or confirming uncontroversial records of common taxa is probably not worth the effort.

Don't worry about checking records of critical taxa if they are outside your sphere of expertise. We'll ask taxon referees to look at the distribution of these taxa.

What should I do if the incorrect record came from MapMate?

If at all possible, correct these records in MapMate. As long as you sync with the MapMate hub your corrections will reach the DDb in a few weeks.

When should I mark a record as "doubtful"?

You should do this when you think the record requires more investigation before rejecting it outright. By marking it as doubtful you stop that record being mapped, but you mark it in such a way that you can easily retrieve it for further investigation.

When should I reject a record?

1. When it is obviously erroneous and can't be corrected or re-determined. This may be because the information is irreconcilable, either because it is internally inconsistent or incompatible with other records.
2. When the taxon is rare or unusual and the record is vague and has no links to supporting information. For example, if there is an anonymous record of *Juncus alpinoarticulatus* for hectad NY66 in the date range 1950-1959, you might consider rejecting it, particularly if that species had never been found in that hectad before or since. Nevertheless, if that same record has a literature reference or a herbarium specimen associated with it, then this supports the record and you should be a lot more cautious about rejecting it. Don't reject records based on their summary: look at their details first.

What should I do if the grid reference doesn't match VC

If the site name and grid reference appear correct then it is usually safe to change the VC. If there is no site name then don't just assume that the grid reference is correct. Sometimes, if it is a rare plant, you might be able to use other records to validate that this species grows at this site. Otherwise, don't be timid about rejecting such a record. If there are two conflicting pieces of information and no supporting evidence, we can't assume that one piece of evidence is more likely to be correct than another.

What should I do if the grid reference doesn't match site name

Reject the record, unless there is other supporting information that validates which of these two pieces of information is correct. If you alter the grid reference ensure that the new grid reference covers all possible places that could refer to the site name.

Sometimes mistakes are due to simple typographic errors and can be corrected easily. Also if this is a known site for this species correction can be made with confidence. However, if doubt remains it may be better to reject it.

What if there is no site name?

Only add a site name if you have evidence to justify it such as the original recording card or field notes, or can speak directly to the recorder or someone in the recording party.

What if there is no recorder?

Again it is best to leave blank if you are in *any* doubt. However you could check the original recording card if it is available.

Further Help?

If you need further help, then please post your question on the [DDb message board](#).