

DDB WORKSHOP:

1. INTRODUCTION TO VALIDATION

2. ASSESSING YOUR ATLAS 2020 COVERAGE

Jim McIntosh, BSBI Scottish Officer



VALIDATION USING THE BSBI DDB

Jim McIntosh, BSBI Scottish Officer



WHAT IS VALIDATION?

A record has the four essential parts (sometimes called the four **Ws**):

What taxon (species, subsp., hybrid, var. etc)

Where was it recorded (site name and grid reference)

When was it recorded (date)

Who was it recorded by (and determined by if different to recorder)



WHAT IS VALIDATION?

A record has the four essential parts (sometimes called the four **Ws**):

What taxon (species, subsp., hybrid, var. etc)

Where was it recorded (site name and grid reference)

When was it recorded (date)

Who was it recorded by (and determined by if different to recorder)

And just as importantly, two **Vs**: **Validation** and **Verification** – confirmation that the record is considered correct and who confirmed it.

Validation - checking that dates are valid and sites, grid references and vice-counties are consistent

Verification - that the taxon ID is considered to be accurate



VALIDATION & THE BSBI

- It is the validation and verification that our VC Recorders and Referees do that sets BSBI data apart from all others.
- It is one of the most important roles of BSBI VC Recorders
- We are keen to be seen to be purveyors of fine records to research, science, conservation and land managers!



VALIDATION & ATLAS 2020

We would like all VC Recorders to validate their VC records

1. There is no one better placed than local VCRs to do it : With their local knowledge of plants, places, people, access to local floras, checklists, raw data recording cards, notebooks, card indices, etc.
2. There is too much data coming in for it to be done centrally



NOW

However large the vc dataset, and however large the number of unchecked or incorrect records, it is a **finite** quantity.

Do not let the backlog of unchecked records get any bigger.

Check, (correct where necessary), and **validate** records as soon as they arrive at the **DDb**



NOW

However large the vc dataset, and however large the number of unchecked or incorrect records, it is a **finite** quantity.

Do not let the backlog of unchecked records get any bigger.

Prioritise record checking

Check, (correct where necessary), and **validate** records as soon as they arrive at the DDb










Check & validate all records

Checking and validating records takes c. 2 - 3 minutes per record card










VALIDATION STATES IN THE DDB

	'Validation' state	Mapped?	Returned by default queries?
	Unchecked (i.e. newly added to the DDb)	✓	✓
	Accepted / Confirmed	✓	✓
	Needs verification / May need verification	X	✓
	Doubtful	X	X
	Rejected	X	X
	Deleted externally (in the origin database)	X	X
	Duplicate	X	X

- **Needs verification** should be used when you plan to, or want someone else, to follow the record up, and there is sufficient doubt about the record that you don't want it mapped (yet).
- There is no functional difference between **Doubtful** and **Rejected**. Rejected means that you are virtually certain the record is incorrect in some way; Doubtful means that it is highly likely to be incorrect.



VALIDATION STATES IN THE DDB

	'Validation' state	Mapped?	Returned by default queries?
	Unchecked (i.e. newly added to the DDb)	✓	✓
	Accepted / Confirmed	✓	✓
	Needs verification / May need verification	X	✓
	Doubtful	X	X
	Rejected	✗	✗
	Deleted externally (in the origin database)	✗	✗
	Duplicate	X	X

- Detecting **Duplicate** records is not (an important) part of validating records.
- There is always something better to do with your limited time than searching for duplicates!
- But avoid creating new duplicate records as far as is practical.



VALIDATION STATES - DEMO

Unchecked records are shown in blue when they arrive at the DDb

Recorders can **reject bad records**, **mark records as doubtful** or **confirm good ones**

results table download map help

4 records matched your query.

?	record	taxon	recorder	vc	locality	grid ref	date	status
<input type="checkbox"/>		<i>Carex acuta</i>		VCH22	VCH22 Meath	N86	1950-1969	
<input type="checkbox"/>		<i>Carex acuta</i>		VCH22	VCH22 Meath	N86	-1996	
<input type="checkbox"/>		<i>Carex acuta</i>	Webb, D.A.	VCH22	Navan, River Boyne	N86	3/6/1950	
<input type="checkbox"/>		<i>Sorbus aucuparia</i>	{ Lockton, A.J. }	VCH22	VCH22 Meath	N67	1950-1969	

select all select none selected records action change record validation state ▼

mark validation state as ▼

apply change to selected records

☐ hide duplicates

Hover cursor over “?” to get list of validation states

Blue: unchecked (default)



Keeping on top of ongoing validation - regularly sync small batches of records to the DDb, and then search for just those records; then check and validate.

search for Records (specimens or observations) ▼

county/region East Inverness-shire

→ creation or update stamp 4/3/2017 equals ▼ created ▼

mapmate centre 2dd ☐ exclude

validation state Unchecked ▼

[more options ▼](#)

Current search-form filter
records where validation state unchecked and vice-county: VC96 and [data set 'MapMateCentre:2dd'](#) and record created on 4/3/2017

results table [download](#) [printable map](#) [zoomable map](#) [help](#)

64 records matched your query.

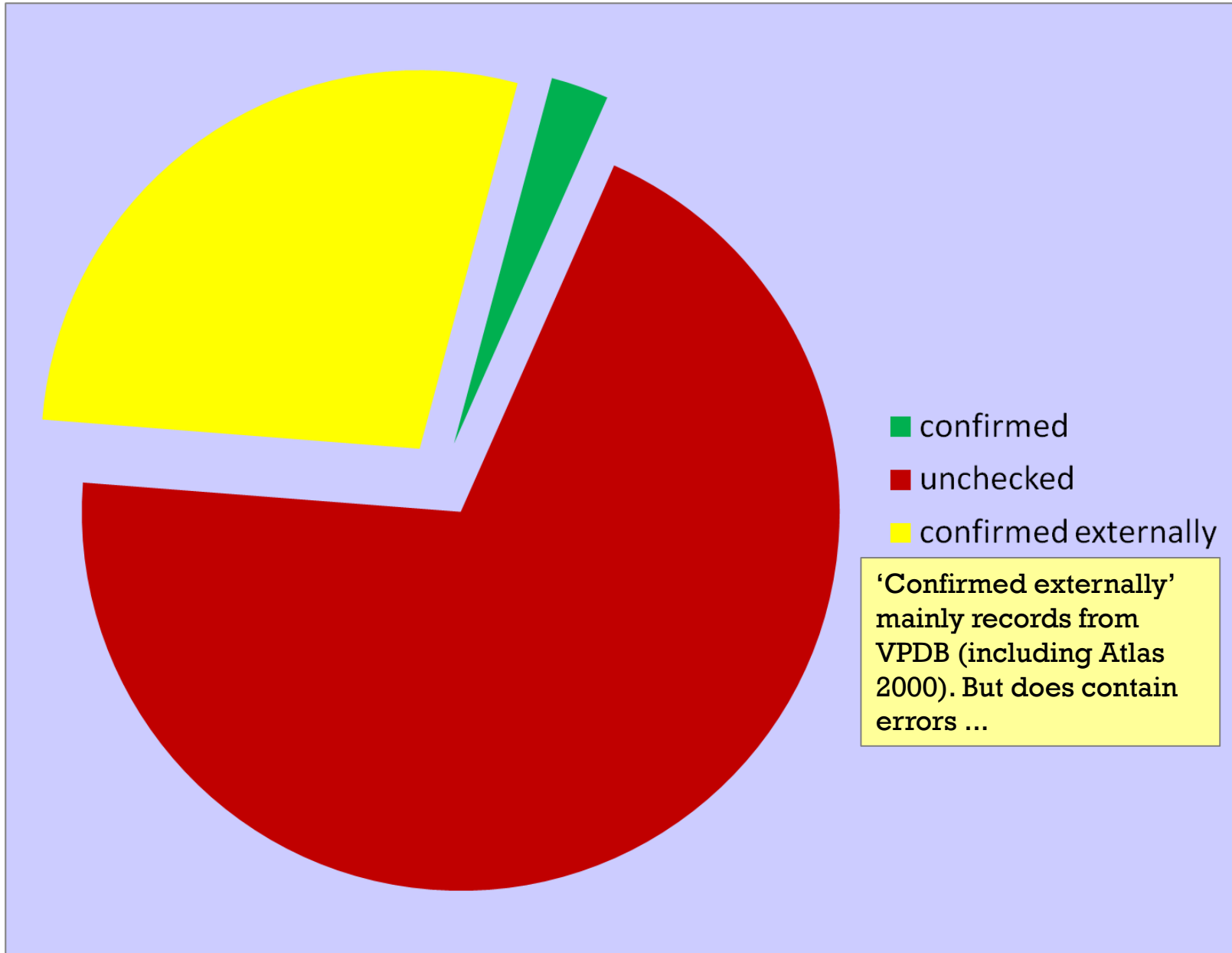
<< first < prev 1 **2** next > last >>

	record	taxon	recorder	vc	locality	grid ref	date	status
<input type="checkbox"/>		Aegopodium podagraria	Barron, Margaret	VC96	Foyers Burial Ground	NH4920	12/5/1983	
<input type="checkbox"/>		Agrostis stolonifera	Barron, Margaret	VC96	Foyers Burial Ground	NH4920	12/5/1983	
<input type="checkbox"/>		Ajuga reptans	Barron, Margaret	VC96	Foyers Burial Ground	NH4920	12/5/1983	
<input type="checkbox"/>		Alchemilla glabra	Barron, Margaret	VC96	Foyers Burial Ground	NH4920	12/5/1983	
<input type="checkbox"/>		Anemone nemorosa	Barron, Margaret	VC96	Foyers Burial Ground	NH4920	12/5/1983	
<input type="checkbox"/>		Anthoxanthum odoratum	Barron, Margaret	VC96	Foyers Burial Ground	NH4920	12/5/1983	
<input type="checkbox"/>		Arabidopsis thaliana	Barron, Margaret	VC96	Foyers Burial Ground	NH4920	12/5/1983	
<input type="checkbox"/>		Bellis perennis	Barron, Margaret	VC96	Foyers Burial Ground	NH4920	12/5/1983	
<input type="checkbox"/>		Betula pubescens	Barron, Margaret	VC96	Foyers Burial Ground	NH4920	12/5/1983	
<input type="checkbox"/>		Cardamine hirsuta	Barron, Margaret	VC96	Foyers Burial Ground	NH4920	12/5/1983	
<input type="checkbox"/>		Cerastium fontanum	Barron, Margaret	VC96	Foyers Burial Ground	NH4920	12/5/1983	
<input type="checkbox"/>		Cerastium glomeratum	Barron, Margaret	VC96	Foyers Burial Ground	NH4920	12/5/1983	

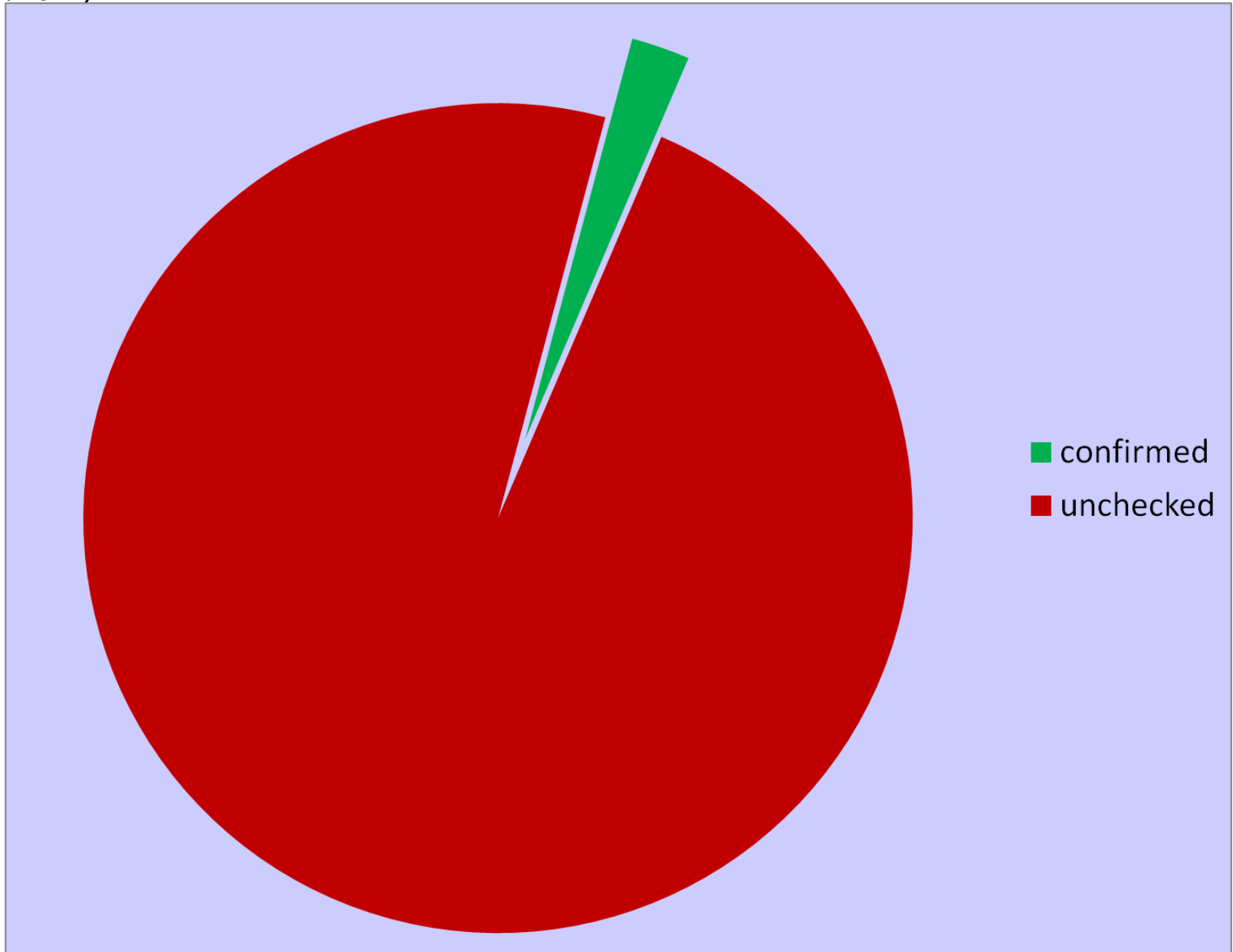


Validation state of records on DDb at (24/2/2017):

Showing that only a very small proportion of BSBI records have been confirmed.



4.8 million records added to the DDb in 2016 ... BUT only 2.3% confirmed so far
(24/2/2017)



PRIORITIES FOR VALIDATION

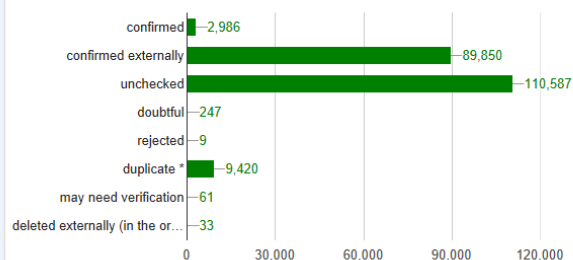
Can be seen in the *My County Data Validation* tab:

County report for VC88 Mid Perthshire

Summary **Data validation** Atlas 2020 survey coverage

Validation state

Validation state assigned to records from the county. (unchecked is the default state for most incoming data)



[\(view data\)](#)

* In addition to any explicitly marked duplicates listed above, [48645 records](#) have been inferred to be duplicates.

Grid-reference - vicecounty boundary matching

Records are hidden on public maps (and would be excluded from the Atlas) if a grid-reference is further than 2km from the specified county. Exceptions will be made for boundary errors and to allow marine species to occur off-shore.

	number of records
good	203,370
≤ 1km	13
1 - 2km	0
2 - 10km	0
10 - 50km	0
> 50km	0
≤ 2km offshore	0
> 2km offshore	0
not testable	0

Records to target for validation for Atlas2020

Suggestions for starting points for detailed checking of records from your county.

Singletons and anomalies

One-off or isolated occurrences records should have their status positively confirmed or rejected.

	current validation state	confirmed	needs verification
	unchecked		
county singletons	80	55	0
100km radius singletons	11	15	0
hectad singletons	1555	2372	15
first county record	419	774	2
records with no Atlas hectad status assigned	2847	1033	9

☒ Exclude neophytes and unchecked taxa.

Rarities

Records of taxa that appear on lists of rarities (either regionally or nationally) may be worth closer scrutiny for Atlas2020.

	current validation state	confirmed	needs verification
	unchecked		
GB National Status	195	91	2
UK Bap Species	56	38	0
GB Red list	91	63	2
Scottish Biodiversity List	114	100	2

☒ Exclude occurrences that are consistent with the distribution in Atlas2000.

Other possible oddities

Please check occurrence records that the database system has highlighted (only a small proportion of these are likely to be erroneous).

	current validation state	confirmed	needs verification
	unchecked		
records near the maximum of their (inferred) altitudinal range	5350	4570	1
records near the minimum of their (inferred) altitudinal range	34	147	0

[749](#) records used site names that may be inconsistently applied to grid-references (i.e. the site name is associated with two or more unrelated locations).

Based on known or inferred dates of death of recorders or determiners, [6 records](#) may have incorrect dates (in the wrong BSBI date class).

IT'S REALLY USEFUL TO UNDERSTAND WHAT'S GONE WRONG

Validation (examples of sources of errors)	Verification (examples of errors)
Grid ref estimated by eye from map	Mis-identification
GPS and smart phone grid ref inaccuracy	Correct ID but recorded wrong taxon, e.g. crossed off wrong name on card or data entry error
Spurious grid ref precision, eg. precise grid ref applied to an ancient record; monad used for tetrad; monad used for quadrant	Taxonomic changes
Grid ref = centroid (especially where 'site' extends to more than one hectad or tetrad)	
Grid ref = where parked car (but recorded over wider area)	
100km grid ref error	
Misreading 3 and 8 in grid ref on GPS screen	
Transcription error	
Site grid ref not compatible with grid ref in Comments	
Site name includes grid ref, but not compatible with site grid ref	
Inaccurate vc boundary used	
MapMate assigned wrong vc	
Date of publication used as date of record	
Incorrect or misspelt site name	
Re-use of site names for different sites	

HOW TO APPLY A 'VALIDATION STATE' TO A RECORD

- DEMO



Validation state change

Please confirm records that you are happy to accept as likely to be correct. Prioritize validation of records for rare taxa and other more interesting or unusual occurrences. **There is no need to validate all data, most noncontroversial records can be left as unchecked.**

Use the 'reasons' below to distinguish between data that you have high confidence in versus records that are assumed to be valid.

Set 1 record as 'confirmed'?

reason:

- Considered correct
- Readily identifiable and uncontroversial
- Positive identification
- Consistent with accepted distribution
- Conforms with literature report

Validation

Mark as confirmed

Cancel



Validation state change

Please only reject records that you know to be *wrong* and where it would be *impossible to safely correct the record*.

Data that cannot be confirmed and which is unlikely to be valid should normally be marked as '*doubtful*'. Records that are unacceptable, but which it might be possible to correct later should be marked as '*needs checking*'.

Set 1 record as 'rejected'?

reason:

Outside known range

Record has internal inconsistencies (date/gridref/locality/other)

Not compatible with local conditions

Similar occurrences have been previously rejected.

Record is misidentified.

Associated specimen is inadequate or corrupted

Mark as rejected

Cancel



EDITING RECORDS

After validating your records, you will have a subset marked 'needs checking', often because they require some form of edit.

When editing a record there are a few basic rules to follow:

1. Edits to a record should be made in the original source; e.g. if you have MapMate, edit in MapMate then sync with the DDb



EDITING RECORDS

After validating your records, you will have a subset marked 'needs checking', often because they require some form of edit.

When editing a record there are a few basic rules to follow:

2. Where mistakes are unambiguous then editing of the record is the best course of action. Otherwise, change the validation state and add an appropriate comment



EDITING RECORDS

After validating your records, you will have a subset marked 'needs checking', often because they require some form of edit.

When editing a record there are a few basic rules to follow:

3. Editing should be used only for changes that are consistent with the intent of the original recorder. Don't add extra detail or changes to make a record fit your preconceptions!



WHAT TO DO IF YOU FIND A PROBLEM

edit or reject?

Record (field observation)	map	validation	status	attributes	version history	admin
Record (field observation)						
taxon	Calystegia soldanella (L.) R.Br. "Sea Bindweed"					
date	23/5/1973					
locality	VC50 Denbighshire VC50 SH976809 					
recorder	I.R. Bonner and J.M. Brummitt					
validity	confirmed externally, assessed by external VPDB user 					
	rejected: "Incorrect grid ref. Correct duplicate exists" (Record has internal inconsistencies (date/gridref/locality/other)) 2017-03-04 by Delyth Williams 					
source data set	Atlas2000 - updated records from December 1995 onwards (VPDB:dataset:158)					
status	native					
external id	BSBI:VPDB:record:198834					
ddb row id	2cd4p9h.1e2xxs #2cd4p9h.4hzcth					
data source	Vascular plants database (via provider The Botanical Society of the British Isles)					
last modified	Wed, 16 Mar 11 14:44:00 +0000					

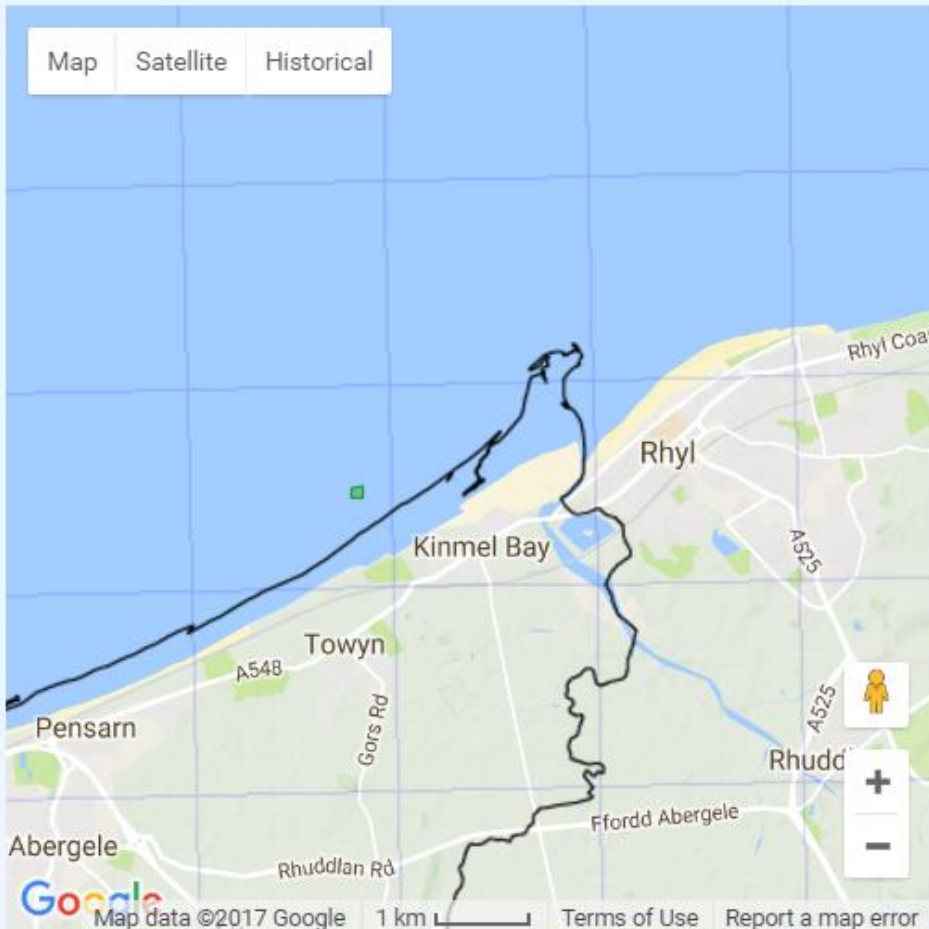
Please cite this database record as <https://database.bsbi.org/record/2cd4p9h.1e2xxs>

WHAT TO DO IF YOU FIND A PROBLEM

edit or reject?

Record (field observation) **map** validation status attributes version history admin

Map Satellite Historical



Current search-form filter

records where validation state any and recorded between 23/5/1973 and 23/5/1973 (using intersection of ranges) and taxon [Calystegia soldanella](#) (L.) R.Br. "Sea Bindweed" and grid-reference [SH976809](#) (radius 2000m)

[display results](#)[email results](#)[download results](#)[save query](#)[reset search form](#)[help...](#)[results table](#)[download](#)[printable map](#)[zoomable map](#)[help](#)

4 records matched your query.

	record	taxon	recorder	vc	locality	grid ref	date	status
<input type="checkbox"/>		Calystegia soldanella	Green, J.A.	VC50	Kinmel Bay	SH996809	1973	
<input type="checkbox"/>		Calystegia soldanella		VC50	Bae Cinmel	SH996809	1973	
<input type="checkbox"/>		Calystegia soldanella	Bonner, I.R. Brummitt, J.M.	VC50 *	VC50 Denbighshire	SH976809	23/5/1973	
<input type="checkbox"/>		Calystegia soldanella	{ Green, J.A. }	VC50	Foryd	SH98	1970-1986	

[select all](#)[select none](#)

selected records action

[change record validation state ▼](#)mark validation state as [apply change to selected records](#)☒ [hide duplicates](#)

Current search-form filter

records where validation state any and recorded between 23/5/1973 and 23/5/1973 (using intersection of ranges) and taxon [Calystegia soldanella](#) (L.) R.Br. "Sea Bindweed" and grid-reference [SH976809](#) (radius 2000m)

display results

email results

download results

save query

[help...](#)

results table

download

print

4 records matched your query.

	record	taxon
<input type="checkbox"/>		Calystegia soldanella
<input type="checkbox"/>		Calystegia soldanella
<input checked="" type="checkbox"/>		Calystegia soldanella
<input type="checkbox"/>		Calystegia soldanella

select all

edit:

Only records to which you have access

apply change to selected records

Bulk edit

Edit 1 record?

Any changes to grid references should reflect the intent of the original recorder.

Remember that the record's site name or county may be wrong rather than the gridref and that a place name might legitimately refer to a wide area. If the gridreference is spuriously precise then it may be best to change it to a centroid instead of editing it using this tool.

new grid reference

comment:

Corrected probable typo in off-shore grid-reference to match duplicate records, grid-ref was SH976809

Edit

Cancel

grid ref	date	status
SH996809	1973	
SH996809	1973	
SH976809	23/5/1973	
SH98	1970-1986	

