

Natural England

Long Term Monitoring Network

Guidance for entry and collation process of vegetation survey data and the quality assurance protocol.



Updated: 27/05/2020

Purpose

The purpose of this document is to instruct the user on:

- How the vegetation data collected in the field on the long term monitoring network (LTMN) in-house surveys and contracted surveys, are collated.
- How the quality assurance (QA) process is applied once the data are collated.

This is to prepare the vegetation data for analysis and ensure the quality assurance process is applied to all data, in the same manner and to the same level of scrutiny. The vegetation data is compiled and readied for analysis, as part of the wider LTMN project to see if climate change is having any effect on vegetation within England, and, if so, identify any habitats and plant communities more at risk than others.



One of the surveyors at a plot at The Lizard LTMN survey in 2017.

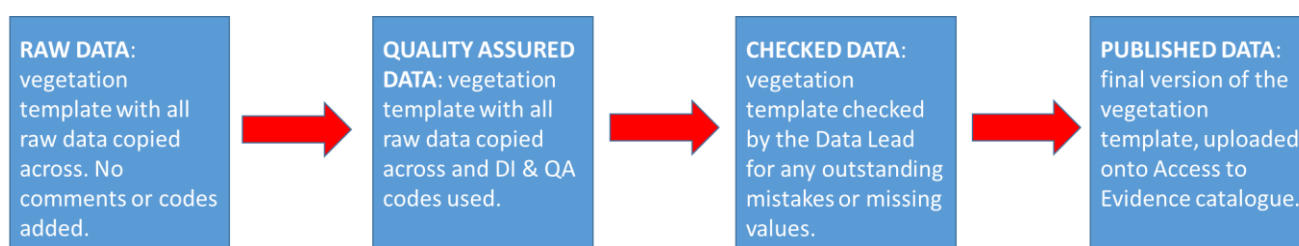
This document may be useful for both data managers (as a step-by-step guide to processing), and to analysts (to understand the history of the data). The table below gives a brief overview of the sections.

Section	Description	Useful for:		Page
		Data managers	Data analysts	
Section 1: Collation of Plot Data	How to assemble the individual survey forms into a single dataset for a site.	✓		2
Section 2: Quality Assurance (QA) Process	How to apply the QA process to the dataset, and associated decision-making.	✓	✓	6
Section 3: Ready for Publication	How to make final changes before the data can be published as open data.	✓		20
Appendix	Lookup tables.	✓	✓	22

Section 1: Collation of Plot Data

The field data from each survey need to be collated into one dataset, so it can be quality assured and then analysed. During LTMN in-house surveys raw data are entered onto iPads using eForms, instead of on paper. These are then emailed during the survey at the end of each day to the LTMN email address (LTMN@Naturalengland.org.uk). After the survey these forms in the mailbox are then saved onto TRIM (Natural England's records management system) before being collated into one vegetation data spreadsheet for each survey.

This diagram shows the process of our data pipeline.



Versions and File Names

The data spreadsheet holding only the raw data that has been copied across from the survey forms is saved as:

Sitename_LTMN_Vegetation_Yearsurveywascompleted_raw_alldata.

Before it undergoes the quality assurance process, it is saved as

Sitename_LTMN_Vegetation_Yearsurveywascompleted_QA_alldata.

Once the dataset has been thoroughly checked by yourself, then it is passed to the Data Lead to complete a final check. After this it can be saved as **Sitename_LTMN_Vegetation_Yearsurveywascompleted_checked_alldata**.

After the data template has been signed off by the Data Lead you can then go on to publish the data. When publishing the data however there should be no traces of personal information associated with the dataset. The final version intended for publication should be saved as **Sitename_LTMN_Vegetation_Yearsurveywascompleted_published_alldata**.

All of these versions should be saved onto the working data folder on TRIM (see link below).



Research and evidence - LTMN Data - Vegetation Data Working.tr5

Collation Process for in-house LTMN Surveys

1. Open the vegetation data input template and save as the correct file name for the location and year of the survey. For example, **Ainsdale_LTMN_Vegetation_2016_raw_alldata**. Make sure it is saved as an excel workbook and you are using the **without macro** version.



Vegetation_Data_Input_Template_FINAL_updated_June_2020.tr5

This spreadsheet will hold all of the raw data from one survey, collated together from the individual survey forms.

2. Fill in the **Site**, **Year**, **Creator** and **Organisation** fields on the Home tab. In the **Change Log**, add the date of creation and “Document created by <your name>”.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2		Long Term Monitoring Network Vegetation Survey											
3													
4		Site	<input type="text" value="Ennerdale & Scoat Fell"/>										
5													
6		Year	<input type="text" value="2019"/>										
7													
8		Creator	<input type="text" value="Joe Bloggs"/>										
9													
10		Organisation	<input type="text" value="Natural England"/>										
11													
12		<u>Additional Information</u>											
13		All methodologies for the data collection, processing and quality assurance can be found here:											
14		LTMN@ Access to Evidence Catalogue											
15		Units for measurements can be found below:											
16													
17													
18													
19													
20		Sheet	Field	Unit									
21		Whole Plot Data	ALTITUDE	Metres									
22		Ground Features	VEG_HEIGHT	Centimetres									
23		Tree Measurement	GIRTH	Centimetres									

For example, this is the Home tab for the Ennerdale and Scoat Fell survey that was conducted in 2019.

3. Download the survey forms from TRIM and open them one at a time.

4. On the survey form, right click the spreadsheet tab and select **move or copy**. In 'To Book' dropdown, **select the vegetation data input template**. Tick the button 'Create a copy' and click ok. Place the copied tab **after** the other tabs in the input template.
5. As you are copying the survey tabs across, **rename each of the tabs** after its quadrat number, for example, if it holds data for quadrat 1 then call it 1. If a quadrat is missing then that number will be missed off. For example if quadrat number 4 was not surveyed then you would not name any tab 4. Just go straight onto quadrat 5 and name the tab 5. If it is an extra species tab then it is named <the quadrat number>x. For example, 1x. This is to differentiate between the two types of data you'll have. **Do steps 3 and 4 for every survey form in order**, for both the main tab and the extra species tab. Once you have done this they will all be appended sequentially to your input template, and you will pull data across from these tabs.
6. After you have done this **unmerge** all of the survey form and extra species tabs. You can do this all at once. Right click on one of the sheet tabs and click select all sheets. On the active sheet (the one you are working on) click the select all button at the top left of the worksheet. Click the dropdown arrow for merge & centre, click unmerge cells.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
7		EVIDENCE OF HUMAN DISTURBANCE			burnt within last 10-15 years, not far from track														
8		ADDITIONAL COMMENTS e.g. anything interesting to know about the plot or area			some young birch in general area.														
9																			
10	CELL	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
11	VEGETATION HEIGHT (CM)	17.0	38.0	31.0	15.0	16.0	13.0	25.0	34.0	14.0	19.0	15.0	10.0	10.0	12.0	11.0	13.0	16.0	14.0
12	OPEN WATER																		
13	BARE ROCK																		
14	BARE SOIL																		
15	BARE PEAT																		
16	LITTER			Y				Y	Y										Y
17																			
18																			
19	SPECIES	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
20	Calluna vulgaris-Heather		Y	Y	Y			Y	Y										Y
21	Vaccinium myrtillus-Bilberry	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
22	Vaccinium vitis-idaea-Cowberry	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
23	Deschampsia flexuosa-Wavy Hair-grass	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
24	Pleurozium schreberi-Red-stemmed Feather-moss	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

This is an example of the raw data that you will copy across to the data spreadsheet and extract the data from.

7. You can now **start copying data** across from the raw data tabs (survey forms and extra species tabs) into the collation tabs of the spreadsheet, e.g. whole plot data and species template.
 - a. Copy and paste data from the survey forms that contain whole plot data e.g. date, names, altitude, land use code etc. into the **whole plot data tab**. You can see each column header on the whole plot data tab and use this as a guide.
 - b. Copy and paste the species data from each survey form tab into the **species template tab**. This includes the plot ID (number), species name, presence and absence (blank cell for absence) and percent cover. Frequency does not need to be copied as it works automatically against a formula. Common names will also be removed later.
 - c. Copy and paste the ground features from each plot including: vegetation height, litter, open water, bare soil and bare peat into the **ground features tab**.

- d. The **extra species** recorded in the extra species tabs of the recording forms (now named e.g. 1x) are copied into the species template tab, below the main chunk of species data. Once they are identified (or not), these lines of data are integrated with the rest of the species data in this tab.
8. Make sure you **copy across the plot number** for each line of data as you do not want to lose it.
9. **Alternatively to copy and pasting, you can pull the data into the collation template using a formula.** This works on the principle that the same sort of data is in the same cell on every survey form, e.g. altitude data is always in cell B5 of the survey form. The Excel formula is =**'tab name'!**\$column\$**row**, e.g. =**'1'!**\$B\$5 will pull through altitude from the tab named 1 (which should be plot 1). The dollar signs tell Excel that those parts of the formula shouldn't change when you drag the formula down to fill a column. Unfortunately, however, Excel isn't clever enough to populate the tab names in sequential cells, so you will need to do this manually. **If you do this, make sure to copy-paste your filled-in columns as values before you delete the field survey tabs, or all your data will disappear.**
10. In the empty cells where the species or ground feature does not have a 1, because it was not identified in that cell **change the empty cell to a 0.** To do this quickly **highlight the area** between cell 1 and cell 25 covering all the species, **press F5**, click **all blanks**, it will then highlight all of the blank cells within the range you have highlighted. Type in **0 and ctrl enter** and it will fill them all in with 0. This is needed because the column that totals frequency works with 1 and 0.
11. Once the species have been copied across you also need to **separate the Latin from the common name** and delete the common so you are just left with the Latin name. This is because the species codes relate to Latin names only. This is done by creating a **new column** to the right of the species names. In this new column click on the first cell and enter in =**LEFT(F2,FIND(" -",F2)-1)** and click enter. This should give you the Latin name only, and you can drag the formula down to trim all the names. Then select the whole column and **paste it back into the species name column as a value** (i.e. just as text, not as the formula). Use this link for guidance: <https://www.extendoffice.com/documents/excel/2738-excel-extract-text-before-after-dash.html>
12. If any **paper forms** were filled out during the survey, that data will need to be manually entered. They will also need to be scanned and copied and saved into the appropriate TRIM site containers.
13. Keep all of the raw data tabs in the RAW version of the spreadsheet. Save onto TRIM working data folders.



Research and evidence - LTMN Data - Vegetation Data Working.tr5

14. You have now collated all of the data.

Section 2: Quality Assurance (QA) Process

The QA process is used to track changes, and highlight any errors or missing data in the compiled data, and flag them in readiness for analysis. Some codes highlight errors like missing percentage covers, while other codes are used to show that the data cannot be used in the analysis process. For example, they may not have a species name or the frequency data may not have been entered on the survey, so it cannot be used.

There are two different types of codes that are used:

- **Data Issues (DI) codes:** describe concerns or failures regarding the recording process.
- **Quality Assurance (QA) codes:** describe issues of data quality and what steps have been taken to address them during or following data entry.

All of the DI and QA codes are listed in the appendix at the end of this document, and they can also be found in the Lookup tab in the vegetation input template.

QA forms the bulk of the data processing, and the vegetation data input spreadsheet can then be checked by the Data Lead, as well as having its species list checked by a habitat specialist. After a few final changes (see *Readying for Publication*, on page 20) the dataset can then be published on the [Access to Evidence catalogue](#).

The following steps will guide you through the quality assurance process.

Home Tab

Before you start the QA process resave the document as **Sitename_LTMN_Vegetation_Yearsurveywascompleted_QA_alldata**.

This is to mark the change between raw data and quality assured data. You must keep the previous raw_alldata version as well, so don't overwrite it.

1. **Delete all of the raw data tabs** from the spreadsheet. All of this data should now be copied across to the appropriate tabs, and they are saved in the raw spreadsheet version if you need to refer to them. **If you used formulas to pull data through from the raw data tabs (instead of copy-pasting) make sure you have copied and pasted these back into the columns as values.** Otherwise they will disappear when you delete the raw data tabs.
2. On the home tab, add a second entry to the **change log**: the date, and "QA process by <your name>".
3. There is also a **survey notes** box to summarise the data that was collected. In this box mention: many plots were surveyed, if any were woodland plots and how many, if any were moved, if high accuracy GPS was used on the survey, any problems that arose, and how many plant species were found (though you won't be able to add this last detail until you've done the Species List).

Survey Notes:

Third survey on this site (baseline survey in 2010). 50 plots surveyed, of which 20 woodland. Plot L survey not completed due to wasp nest. 4 woodland plots surveyed at a later date (G, H, S, T). Note woodland co-ordinates (and therefore trees chosen) differ to previous surveys, which were quite incomplete. In terms of woodland survey, this may be considered the baseline survey. Species richness: 145.

For example, the survey box for the Bure Marshes survey in 2019.

- 4. There is also a hyperlink to the LTMN vegetation protocol: make sure it is called [LTMN@ Access to Evidence Catalogue](#) and hyperlinked to the **correct page** (<http://publications.naturalengland.org.uk/publication/6331137473904640?category=5895911152025600>).

Whole Plot Data Tab

This tab is used to record grid references, surveyor names, dates and plot measurements. In this tab any remarks about the location of individual plots also need to be recorded, for example, if the plot was moved or if it was near any particular features which would make it easier to find.

- 5. Make sure that all the dates are formatted as **dd/mm/yyyy**. Names should be **full names** e.g. Joe Bloggs or J. Bloggs.

An example of the whole plot data tab after all of the data has been entered.

- 6. Check that the **site code** is the correct one for each site (this can be checked in the Lookup Table, the last tab in the spreadsheet). **VC** is a vegetation plot and **VW** a woodland plot in the MCODE column. The **year** is the year the survey was conducted in (not the year it is being QA'd, if this is different).
- 7. **If any data is missing**, enter the appropriate DI code in the DI code columns, e.g. they have not entered an altitude value use the DI code **DI900**. If they have entered a 0 into a plot data column like Aspect take this as meaning that the aspect for that plot is 0 (North), not that they have not entered it in. **All of the DI and QA codes are listed in**

the appendix at the end of this document, and they can also be found in the Lookup tab in the vegetation input template.

8. **Check that all values are appropriate** for the site and the variable, and that the surveyors have not entered something which is not a true value. E.g. land use codes only go up to 22 so anything above that is wrong. Aspect can only go up to 360 degrees. Units that these measurements were taken in do not need to be recorded in the cells, as they are specified on the Home tab.
9. A common error is entering 0 for aspect to indicate “none” (where slope is 0), but in fact 0 degrees would be due North. **Where slope is 0, enter N/A into aspect.** You must also enter DI990 and QA901 into the DI/QA columns to indicate that there was a problem with the plot data and you modified it. Also enter a comment into the Comments column describing what you did.
10. If aspect has been recorded using compass points (e.g. SW), you will need to **convert to degrees.**
11. **Check that broad and priority habitats** are correctly spelt and that they are from the list in the lookup table. After the QA process is complete you send the completed document to the national habitat specialist that was present for the survey (not done for contracted surveys). They then check the species list and the habitat types to make sure they are correct for the site.
12. The **Field Sheet Notes** column holds all the information from the Evidence of Grazing/Human Disturbance and Additional Comments cells in the survey form. The **Score** codes are used to describe disturbance. The Score 1 column is to describe any animal grazing or disturbance e.g. 141 sheep grazing. Score 2 column is human disturbance e.g. code 152 would mean a human path was present. Score 3 column is used if there is more than one code required for animal or human disturbance. For example, if there is more than one grazing animal you would put the second code for the animal grazing into the Score 3 column.
13. **Remember that anywhere you find a problem you must use a DI code, and/or a QA code, and any time you make a change, you must use a QA code.** If a QA code is used then this also needs to be re-entered into the **final QA column.** If more than one QA code has been used then the code that has changed the data more significantly needs to be entered into the final QA code column.
14. You should also **enter a comment** in the Data Entry Comments column briefly describing the problem and/or what change you made. E.g. if they entered the slope code as ‘flat’, change to 2, which represents a flat (not necessarily level) surface, then put in data comments that it was originally entered as ‘flat’, and add QA codes to flag the change you made: **QA903** (slope form does not comply with survey methodology) and also **QA902** (slope form modified/ inferred).
15. Use the site plot checker (used on each individual survey) to **update the Location Remarks** column. E.g. “plot not found, new feno marker installed as close to original co-ordinates as possible”. **If a plot cannot be found and a new feno marker is placed as close to the original coordinates as possible the plot identification changes.** It is now called the number of the plot and ‘a’ (e.g. 2a). This is to show the

plot is a replacement for the original, sited as near to the original as possible but the original feno marker was not found. Throughout the whole vegetation template the plot needs to be called the correct name e.g. 2a not 2, in all tabs containing data for that plot.

16. The **Master Plot Checker** and any **site plot tracker** that was in use during the survey should show the same set of **coordinates** and can be used to copy and paste the most updated grid references and BNG grid reference into the veg data sheet - as long as it is the most up-to-date version.
17. For any **contracted data** that comes through make sure they have entered in **NVC values** for the 1st, 2nd and 3rd column as well as made their own judgement in the NVC class column. As well as the NVC class in the three columns there should also be a percentage value. There should be a dash between the NVC value and the percentage. E.g. H12a-30.4. An example can be seen in the appendix.

Species Template Tab

This is where the bulk of the botanical data is found – which species are in which plot, and their percentage cover. It should be ordered numerically by plot number. **Be aware that if your dataset contains any plots with ‘a’ appended, Excel’s sort function groups them at the end. You will need to remedy this manually.**

18. The species code formula is
=IF(F2="", "", VLOOKUP(F2, All_LTMN_Lookups!\$J\$1:\$K\$1223, 2, FALSE))
This should already be in all the cells of the Species_No column, but make sure that it's copied down so it covers all the lines of data. If some species do not have a species code (N/A comes up) then this normally just means that they are not in the species lookup table, and this is fine. 'Unidentified' species will also not have a species code, while some that are identified to genus may do.
19. **Double-check lines that have N/A as species code**, as sometimes it is because a synonym was used and the actual name is in the species lookup table. Or it could just require adding (g) to a tree species (growth stage). Manually enter the code from the lookup table if necessary.
 - a. If you are adding a growth stage for a tree or shrub that lacks it, so changing the species name slightly so that it fits with the name and growth stage in the lookup table, **you will need to use a DI and/or a QA code**. For example, just adding (g), then this needs the code, **DI814** (tree species: stage not recorded on field form, assumed in the ground layer).
20. If the species **does not have a percentage cover** (i.e. the surveyor has forgotten to enter one) then this needs a code, **DI800** (abundance % cover missing).
21. If there is **no frequency cover**, i.e. no presence is indicated in any of the cells (but it has a percentage cover) then we can just assume that it was found within the plot but the surveyor has forgotten to enter in presence. In this case you can **enter presence in cell 1**, use the codes **DI801** (abundance, cell data missing), **QA601** (cell data modified) and **QA777** (abundance data unreliable for analysis). **Add a comment** to the data entry comments column saying that you have added 1 to cell 1.

22. If a species % cover is under 0.1% then this needs to be rounded up to 0.1 and the **original value entered** into the data comments. It then needs codes **DI805** (cell data % cover does not comply with survey methodology) and then **QA600** (% cover modified/inferred).
23. If you have **changed the spelling** of any of the species names because it was slightly incorrectly spelt then enter into data comments on that line **'spelt slightly incorrectly'**. If the correct spelling was easy to find then this does not need any codes. However, if it took a while to find the correct spelling or correct name then enter the code **DI806** (taxon name unreadable/incorrect) and **QA606** (taxon confirmed post survey by checking protocol). If you are unable to read someone's writing then return to the surveyor and ask if they can read what they have written.
24. If taxa are entered as groups higher than species or genera e.g. bryophytes or algae then this needs to be changed to **unidentified**, **QA766** and the **original value entered** in data entry comments.
25. The frequency values in cell_1 to cell_25 indicate presence/absence, so **only values of 1 or 0 are correct**. Sometimes there are obvious typos, for example, if it says 11, we can assume they meant to type 1 (indicating presence in that cell).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	
1	SITECODE	MCODE	YEAR	PLOT_ID	SPECIES_NO	DESC_LATIN	CELL_1	CELL_2	CELL_3	CELL_4	CELL_5	CELL_6	CELL_7	CELL_8	CELL_9	CELL_10	CELL_11	CELL_12	CELL_13	CELL_14	CELL_15	CELL_16	
2	B31	VC	2017	1	278	<i>Calluna vulgaris</i>	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	
3	B31	VC	2017	1	541	<i>Erica cinerea</i>	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	
4	B31	VC	2017	1	1066	<i>Pteridium aquilinum</i>	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1
5	B31	VC	2017	1	1891	<i>Polytrichum commune</i>	1	1	0	1	1	1	0	1	1	1	0	1	0	1	0	1	1
6	B31	VC	2017	1	2695	<i>Hypnum jutlandicum</i>	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	
7	B31	VC	2017	1	3079	<i>Agrostis curtisii</i>	1	0	0	1	1	0	0	1	0	1	0	0	1	1	1	1	
8	B31	VC	2017	1	2939	<i>Quercus seedling</i>	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
9	B31	VC	2017	1	2366	<i>Cladonia portentosa</i>	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
10	B31	VC	2017	1	576	<i>Festuca rubra</i>	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	
11	B31	VC	2017	1	876	<i>Molinia caerulea</i>	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
12	B31	VC	2017	1	2619	<i>Pinus sylvestris (s)</i>	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
13	B31	VC	2017	1	2346	<i>Cladonia chlorophaea</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	B31	VC	2017	1	2694	<i>Cladonia</i>	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
15	B31	VC	2017	2	876	<i>Molinia caerulea</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16	B31	VC	2017	2	541	<i>Erica cinerea</i>	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0
17	B31	VC	2017	2	#N/A	<i>Ulex minor</i>	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	0	0
18	B31	VC	2017	2	541	<i>Erica cinerea</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

An example of the Species Template tab.

Species template and extra species	Check the extra species for any spelling mistakes, check with the Stace book and NBN atlas for correct name. Move checked species into species template by cutting and pasting.	If the taxon name is unreadable/ very incorrect use code DI806 . Then QA606 (taxon confirmed by checking protocol), if able to identify. If only identified to genus QA607 . If there is a ? After name use DI807 .	If no percentage cover use DI800 . If no cell frequency, enter DI801 , and add cell presence in cell 1. Put codes QA601 (presence modified) & QA777 (abundance unreliable for analysis).	If percentage is too low and has to be changed then use DI805 (% cover does not comply with methodology) and QA600 (% cover modified).
	Also use experts' returned species lists to see if they have identified any of the extra species. Where found, use correct codes and move into species template.	If a species is identified by an expert than you can use code QA608 (added post survey from sample return). If it is confirmed by an expert then use QA605 . Confirmed post survey by surveyor QA604 .	If it can only be identified to genus use QA607 . Leave genus in species names.	If it cannot be identified then use code QA766 (taxon unknown, unreliable for analysis). Rename 'unidentified'. Put named recorded on survey form in comments.
	Multiple species identified in one sample.	DI804 - Abundance: two or more records share cell/% cover data DI810 - Sample returned: more than one sample in bag	QA605 - Taxon: confirmed post survey by sample QA777 - Abundance: data unreliable for analysis	
	Species recorded twice. Merge cell frequency data and % covers into single entry.	DI809 - Taxon/Abundance: species recorded more than once on field form	QA602 - Abundance: % cover and cell presence modified / inferred. Different code if just merged cell data or just %.	
	Species entered as sample on recording form, but no matching sample reaches expert.	DI808 - Taxon: labelled as 'sample' but no sample returned/supplied	QA607 - If entry on recording form includes genus. QA766 – If entry on recording form didn't include genus.	If genus is recorded, leave that in species name. If genus isn't recorded, change name to Unidentified and move original name record (e.g. Bryophyte A) to Data Entry Comments.

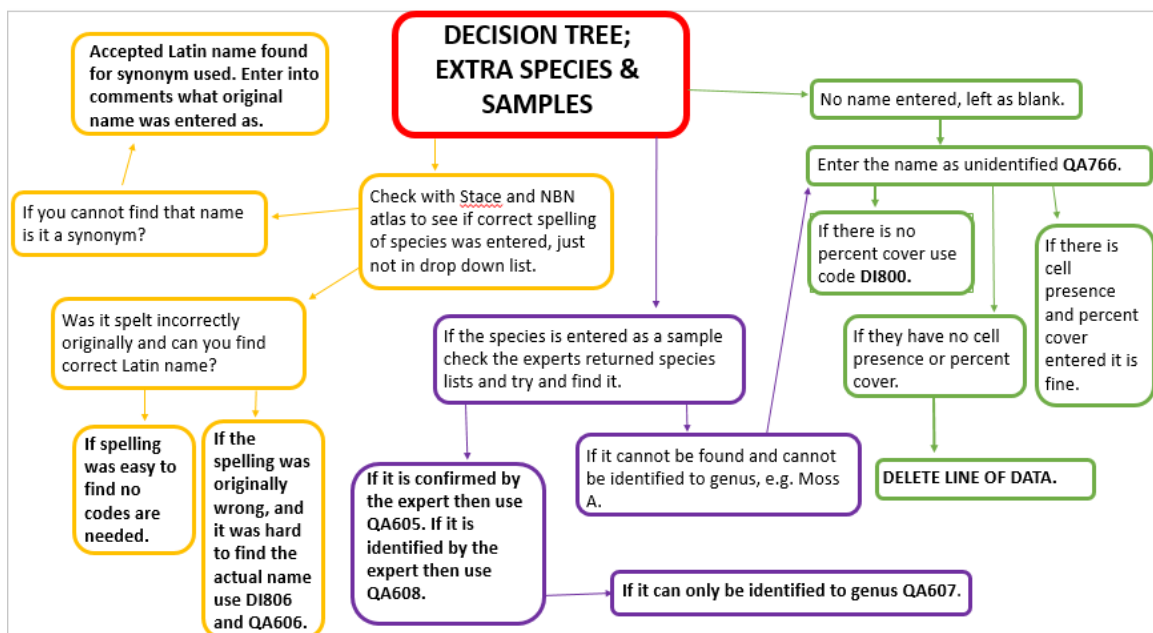
Table 1, showing the different problems that may arise during the quality assurance process and the codes that need to be attached to different lines of data.

26. Now you should address the **extra species**.

- a. Check all of the extra species to see if any of them are **named in full**. The species name might be correct but might not have been on the drop down list. Check the

names against the [NBN database online](#) and the book, *New Flora of the British Isles* by Clive Stace, 2019 edition. If you change the species name because it was a synonym or incorrect spelling, **enter into data comments the original entry and why it was changed**. However, if a species name has changed recently, use the version that appears in the species lookup table on the Lookup tab, so that the species code will pull through.

- b. The extra species that are left should now be either illegible, or species that have been given to an expert as samples for identification. **You should receive a list from the expert(s)** in the weeks post-survey identifying these. As what was entered on the recording form should be the same as what was written on the sample bag (e.g. "Bryophyte A") you can then use this list to **match up species** that they have identified.
- c. When changing a sample name to the identified name, you should also **put the original name the surveyor entered** (e.g. moss A) into data entry comments. This is to keep a record of any changes after the surveying process. There is then different **QA codes** to use for a species identified after surveying or from a sample bag (see below).
- d. If there are any samples in extra species that were **not identified by the expert** after the survey and you have no idea what they are e.g. Bryophyte B then put the **original name into data entry comments** and change the species name to **Unidentified**.
- e. For any lines of data where the species is **unidentified** enter the code **QA766**. This code means the taxon is unknown and unreliable for analysis. If however, the taxon was entered in as 'Sphagnum' for example, but was not identified to species level then you can leave it and use the code **QA607 (only confirmed to genus)**. This data can still be used. Just make sure to enter into data comments 'only confirmed to genus'. This is the lowest level the taxon can be identified to.



A decision tree diagram, use this to show how decisions should be made relating to extra species and sample bag species.

27. Once you have dealt with all of the extra species you can **integrate them into the main bulk of the species data**, according to plot number. Make sure to copy any

codes and comments you have entered, and that they are entered in numerical plot order.

28. Obviously, the usual checks and codes refer to these lines of data as regards **issues with frequency or percentage cover**.

29. Before moving onto the next tab make sure any lines of data with a QA code also have a **final QA code**.

Ground Features Tab

This tab is used to describe any ground features that were present within the plot: litter, bare soil, bare soil and open water. There are 25 cells in each plot, and their presence is marked for each. The vegetation height is also measured in each cell, and integrated into the ground feature data so that the plots remain in numerical order

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
1	SITECODE	MCODE	YEAR	PLOT_ID	FEATURE	CELL_1	CELL_2	CELL_3	CELL_4	CELL_5	CELL_6	CELL_7	CELL_8	CELL_9	CELL_10	CELL_11	CELL_12	CELL_13	CELL_14	CELL_15	
2	B31	VC	2017	1	Vegetation Height	18.5	44.5	40.5	38.5	12	36.5	25.5	18	24.5	10	29.5	25.5	26	12.5	33	
3	B31	VC	2017	1	Litter	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	B31	VC	2017	2	Vegetation Height	25.5	31.5	32.5	27.5	19.5	26.5	39.5	39.5	25.5	17.5	35.5	41.5	29.5	26.5	21	
5	B31	VC	2017	2	Litter	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
6	B31	VC	2017	3	Vegetation Height	18	29.5	25	26.5	19.5	30.5	24.5	19	29	30.5	42.5	23	15.5	47	32	
7	B31	VC	2017	3	Litter	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
8	B31	VC	2017	4	Vegetation Height	15.5	21.5	24.5	32.5	18	17.5	19	14.5	30	18.5	14.5	22.5	12.5	22.5	26	
9	B31	VC	2017	4	Litter	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
10	B31	VC	2017	5	Vegetation Height	28.5	24.5	35.5	26	28.5	19	34	41.5	28.5	40.5	19	29	39	38	36	
11	B31	VC	2017	5	Litter	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
12	B31	VC	2017	6	Vegetation Height	6.5	26.5	14.5	19.5	17	9	26	22	19.5	16.5	10.5	13.5	25	49		
13	B31	VC	2017	6	Litter	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
14	B31	VC	2017	7	Vegetation Height	22	36	42	13	15	12	31	19	13	26	22	11	8	5		
15	B31	VC	2017	7	Bare Soil	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
16	B31	VC	2017	7	Litter	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
17	B31	VC	2017	8	Vegetation Height	39.5	116	170	150	96	39.5	91	109	85	35	39.5	11	4.5	6	12	

An example of what the ground features and extra species tab should look like.

30. On this tab the same details need to be filled in for **site code, mcode, year and plot ID**.

31. Data stays in plot numeric order.

32. Make sure that in the columns cell 1 to cell 25 there is either a **1 or 0 for all features** apart from vegetation height. All the ground features apart from vegetation height should also have a **percentage cover and frequency**.

33. Vegetation height should never be more than 100cm. If the surveyors have entered anything more than this, reduce it 100, make a note in the data entry comments of the original entry, and use the codes **DI910** (vegetation height does not comply with survey methodology) and **QA901** (plot data modified/inferred).

34. Once more enter any missing details into **data entry comments** e.g. missing percentage cover and put in appropriate **DI and/or QA code**. Wherever you use a QA code in the whole spreadsheet there should also be a **final QA code**. If there is only one QA code for one line of data then you put this in final QA code. However, if there is more than one QA code on a line of data you have to make the decision which is more

important to be entered in the final QA code (which has made a bigger impact or change on the data).

Woodland Data Tab

For some of the LTMN sites there are plots that are surveyed within a woodland. This involves additional measurements in the field, and means that they will have extra data filled out compared to sites with no woodland plots. The tabs for the woodland data only need to be filled in for surveys that have woodland plots, and these are: **woodland coordinates, seedlings template, tree measurement and dominance template**. In some surveys there will not be very much woodland data as there may only be one or two woodland plots, such as Stiperstones, which only has one woodland plot. But for some surveys nearly all of the plots will be woodland plots, for example, Epping Forest.

Woodland data is recorded on the same field survey form but in different tabs. During the collation process these tabs should also be appended to the vegetation template, and the data from these forms should be copied across to the relevant tabs on the template.

- **For all of the woodland data tabs site code, mcode, year and plot ID all need to be filled in.**
- **If any new trees were marked on the survey this needs to be mentioned in the survey notes on the home tab.**

Woodland Coordinates Tab

The woodland coordinates tab is used to record the positioning of 10 trees per plot, and if there were any new trees marked that year.

- 35.** Like the other tabs **sitecode, mcode, year** and **plot ID** all need to be filled in.
- 36.** The woodland subcell ID should have been filled in during the collation process from any paper forms, along with subcell X and subcell Y. These are taken from the diagram that is filled out when first surveying the random trees, and refer to the coordinate system (1-25) used within the 10m plot.
- 37.** Bearing of stem from centre of random cell and distance of stem from centre of random cell should have been filled out during the survey and copied across during the collation process. The distance of stem column was only added into the survey method in 2018 so previous surveys are unlikely to have this column filled in. Occasionally surveyors incorrectly enter this distance measurement into the distance column on the Tree Measurements tab, which is meant to hold the distance of the surveyor from the tree (for calculating tree height). You will need to use your judgement as to whether this has occurred on the survey you are working on.
- 38.** If **bearing has been entered as compass points** (e.g. SW), you will need to convert to degrees.
- 39.** If there is any missing information this needs to have the correct DI code added to that line. Before moving onto the next tab make sure any lines of data with a QA code also have a **final QA code**.

	A	B	C	D	E	F	G	H	I	J
1	SITECODE	MCODE	YEAR	PLOT_ID	WOODLAND_SUBCELL_ID	WOODLAND_SUBCELL_X	WOODLAND_SUBCELL_Y	BEARING_OF_STEM_FROM_CENTRE_OF_RANDOM_CELL	DISTANCE_OF_STEM_FROM_CENTRE_OF_RANDOM_CELL	DATA_ENTRY_COMMENT
2	B34	VW	2018	1a	1	8	25	80	1.33	New tree 2018.
3	B34	VW	2018	1a	2	5	8	300	1.62	New tree 2018.
4	B34	VW	2018	1a	3	6	12	123	3.15	New tree 2018.
5	B34	VW	2018	1a	4	20	21	338	1.39	New tree 2018.
6	B34	VW	2018	1a	5	6	7	121	1.57	New tree 2018.
7	B34	VW	2018	1a	6	11	23	75	1.23	New tree 2018.
8	B34	VW	2018	1a	7	13	3	246	1.6	New tree 2018.
9	B34	VW	2018	1a	8	10	1	260	4.93	New tree 2018.
10	B34	VW	2018	1a	9	4	13	242	2.05	New tree 2018.
11	B34	VW	2018	1a	10	17	12	17	3.44	New tree 2018.
12	B34	VW	2018	2a	1	2	25	166	2.78	New tree 2018.
13	B34	VW	2018	2a	2	25	24	200	1.45	New tree 2018.
14	B34	VW	2018	2a	3	2	20	196	1.23	New tree 2018.
15	B34	VW	2018	2a	4	16	20	20	1	New tree 2018.
16	B34	VW	2018	2a	5	9	22	110	4.72	New tree 2018.

This is an example of what the woodland coordinates tab should be like once all details are entered.

Seedlings Template Tab

This tab is used to show a record of any seedlings found in the 10 random cells in the woodland plot. The seedlings names are in Latin so that they have a species code.

- 40.** As with the species template tab, **species names should be checked**, especially where they have N/A in the species code column. QA codes such as QA607 (identified to genus only) can also be used here.
- 41.** Within the columns cell 1 to cell 10 there should just be a number present to show the number of seedlings of that species found within that cell (**it can be more than 1**). If the cell is blank we can presume it is 0, and you need to enter this if you haven't already (see step 9 of collation instructions).
- 42.** Enter in any data comments that needs a **DI or QA code**, for example they have not used a species name or it is only confirmed to genus.
- 43.** If **no seedlings** were found at the plot leave it empty and then write in data comments 'no seedlings'.
- 44.** Before moving onto the next tab make sure any lines of data with a QA code also have a **final QA code**.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	SITECODE	MCODE	YEAR	PLOT_ID	VEG_SPEC	DESC_LATIN	CELL_1	CELL_2	CELL_3	CELL_4	CELL_5	CELL_6	CELL_7	CELL_8	CELL_9
2	B34	VW	2018	1a	2616	<i>Ilex aquifolium (g)</i>	2	0	0	0	0	0	0	0	0
3	B34	VW	2018	2a	2616	<i>Ilex aquifolium (g)</i>	1	0	1	0	0	0	0	0	0
4	B34	VW	2018	6	2616	<i>Ilex aquifolium (g)</i>	1	0	0	1	0	0	0	0	0
5	B34	VW	2018	7	2616	<i>Ilex aquifolium (g)</i>	0	0	1	0	0	0	0	0	0
6	B34	VW	2018	8											
7	B34	VW	2018	9											
8	B34	VW	2018	12											
9	B34	VW	2018	13	2616	<i>Ilex aquifolium (g)</i>	0	1	0	0	0	0	0	1	0
10	B34	VW	2018	14											
11	B34	VW	2018	15	2616	<i>Ilex aquifolium (g)</i>	0	0	1	0	0	0	0	0	0
12	B34	VW	2018	16											
13	B34	VW	2018	17											
14	B34	VW	2018	19	2616	<i>Ilex aquifolium (g)</i>	0	0	0	0	0	0	0	1	1
15	B34	VW	2018	20a											
16	B34	VW	2018	21a	2607	<i>Betula pendula (n)</i>	0	0	0	0	0	1	0	0	0

This is an example of the seedlings template tab.

Tree Measurement Tab

This tab holds the measurements of the 10 trees per plot identified in the woodland coordinates tab. This should include species name (Latin), girth of tree, diameter, height, and number of stems are entered. Depending on the method used for calculating tree height, details about the surveyor may also be entered (their distance from the tree, their height to eye level, and the angle to the top of the tree).

45. Make sure that **cell ID** is filled in, and this can then be **copied into tree ID** as they are the same thing. Stem ID is not always measured so can be left blank in the template and no codes are needed.
46. As with the species template tab, **species names should be checked**, especially where they have N/A in the species code column. QA codes such as QA607 (identified to genus only) can also be used here.
47. If any **data is missing** e.g. the surveyor height, then the appropriate code needs to be used to show there is missing data. **DI900** Plot data: some data not recorded/missing.
48. If **girth is missing** this can be worked out using the diameter, and vice versa. To work out:

Girth = diameter x pi	in Excel =<cell with diameter>*Pi()
Diameter = girth/pi	in Excel =<cell with girth>/Pi()
49. Before moving onto the next tab make sure any lines of data with a QA code also have a **final QA code**.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	SITECODE	MCODE	YEAR	PLOT_ID	CELL_ID	TREE_ID	STEM_ID	VEG_SPEC	DESC_LATIN	GIRTH	DIAMETER	HEIGHT	NUM_STEMS	DISTANCE	SURVEYOR HEIGHT	ANGLE_degrees
2	B03	VW	2017	1	1	1		971	<i>Pinus sylvestris (c)</i>	232	73.8	28.9		15	1.8	61
3	B03	VW	2017	1	2	2		1078	<i>Quercus robur (c)</i>	91	29.0	57.8	2	15	1.8	75
4	B03	VW	2017	1	3	3		1078	<i>Quercus robur (c)</i>	149	47.4	15.3	2	15	1.8	42
5	B03	VW	2017	1	4	4		1078	<i>Quercus robur (c)</i>	135	43.0	19.1	3	15	1.8	49
6	B03	VW	2017	1	5	5		1078	<i>Quercus robur (c)</i>	166	52.8	22.4		15	1.8	54
7	B03	VW	2017	1	6	6		1078	<i>Quercus robur (c)</i>	84	26.7	19.1	2	15	1.8	49
8	B03	VW	2017	1	7	7		570	<i>Fagus sylvatica (c)</i>	10	3.2	8.5		15	1.8	24
9	B03	VW	2017	1	8	8		1078	<i>Quercus robur (c)</i>	95	30.2	15.3		15	1.8	42
10	B03	VW	2017	1	9	9		1078	<i>Quercus robur (c)</i>	67	21.3	14.4		15	1.8	40
11	B03	VW	2017	1	10	10		1078	<i>Quercus robur (c)</i>	68	21.6	14.4	2	15	1.8	40
12	B03	VW	2017	2	1	1		570	<i>Fagus sylvatica (c)</i>	121	38.5	21.7	2	15	1.8	53
13	B03	VW	2017	2	2	2		570	<i>Fagus sylvatica (c)</i>	120	38.2	14.4		15	1.8	40
14	B03	VW	2017	2	3	3		570	<i>Fagus sylvatica (c)</i>	93	29.6	19.7		15	1.8	50
15	B03	VW	2017	2	4	4		570	<i>Fagus sylvatica (c)</i>	88	21.6	17.3		15	1.8	46
16	B03	VW	2017	2	5	5		1078	<i>Quercus robur (c)</i>	189	60.2	20.3		15	1.8	51
17	B03	VW	2017	2	6	6		1078	<i>Quercus robur (c)</i>	128	40.7	21.7		15	1.8	53
18	B03	VW	2017	2	7	7		1078	<i>Quercus robur (c)</i>	123	39.2	20.3		15	1.8	51
19	B03	VW	2017	2	8	8		1078	<i>Quercus robur (c)</i>	76	24.2	15.8		15	1.8	43

This is an example of how the tree measurement tab should look after all of the data is entered.

Dominance Template Tab

This tab lists all the tree species present in the woodland plot and their level of dominance in the canopy. The species code again should work if the species name is correctly spelt and there are no blank spaces at the end of the species name.

50. As with the species template tab, **species names should be checked**, especially where they have N/A in the species code column. QA codes such as QA607 (identified to genus only) can also be used here.
51. If any **data is missing** then the appropriate code needs to be used to show there is missing data. **DI900** Plot data: some data not recorded/missing.
52. The dominance code column just requires the single letter code. These can be found in the **dominance lookup table** on the Lookup tab, and also in the appendix to these instructions.
53. Before moving onto the next tab make sure any lines of data with a QA code also have a **final QA code**.

Species List Tab

A species list also needs to be compiled at the end of the QA process and checked by the national habitat specialist that attended the survey. This is to make sure that all the correct spellings are used, that species are placed into the correct categories, and that there are no unlikely species for that habitat or region. The species list is created from all the species found during that survey, taken from the species template and any woodland plots tabs. The species list is split into the categories: Bryophytes, Ferns & Horsetails, Grasses, Sedges & Rushes, Herbs, Lichens, Trees & Shrubs.

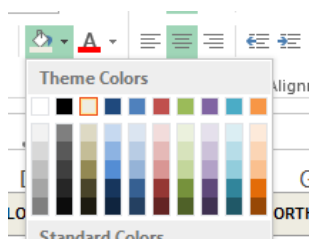
	A	B	C	D	E	F	G
1	Bryophytes	Ferns & Horsetails	Grasses, Sedges & Rushes	Herbs	Lichens	Trees & Shrubs	
2	Brachythecium albicans	Pteridium aquilinum	Agrostis canina	Cerastium fontanum	Cladonia	Acer pseudoplatanus (g) (c) (s)	
3	Brachythecium rutabulum		Agrostis capillaris	Cirsium palustre	Cladonia fimbriata	Betula pendula (s) (g) (c)	
4	Campylopus pyriformis		Aira praecox	Epilobium		Betula pubescens (s) (c)	
5	Ceratodon purpureus		Alopecurus geniculatus	Galium aparine		Calluna vulgaris	
6	Dicranella heteromalla		Alopecurus pratensis	Galium saxatile		Carpinus betulus (g) (s) (c)	
7	Dicranum scoparium		Anthoxanthum odoratum	Geum urbanum		Crataegus monogyna (g) (s) (c)	
8	Hypnum cupressiforme		Carex flacca	Hypericum pulchrum		Erica tetralix	
9	Hypnum jutlandicum		Carex hirta	Lonicera periclymenum (g)		Fagus sylvatica (s) (g) (c)	
10	Kindbergia praelonga		Carex pilulifera	Lotus pedunculatus		Fraxinus excelsior (g)	
11	Leucobryum glaucum		Carex remota	Ranunculus repens		Hedera helix (g)	
12	Mnium hornum		Deschampsia flexuosa	Rumex acetosa		Ilex aquifolium (g) (s) (c)	
13	Polytrichastrum formosum		Festuca ovina	Rumex acetosella		Malus sylvestris (g)	
14	Polytrichum commune		Holcus lanatus	Rumex obtusifolius		Populus tremula (c)	
15	Polytrichum formosum		Holcus mollis	Senecio jacobaea		Prunus spinosa (g) (c)	
16	Pseudoscleropodium purum		Juncus acutiflorus	Stellaria graminea		Quercus cerris (c)	
17	Rhytidiadelphus squarrosus		Juncus effusus	Stellaria media		Quercus robur (g) (c)	
18	Sphagnum fallax		Juncus inflexus	Teucrium scorodonia		Quercus seedling	
19	Sphagnum palustre		Luzula campestris	Trifolium repens		Rosa canina	

This is an example of a species list that was compiled after the quality assurance process.

54. **Copy across all species** from the species template and any woodland plot tabs into one column.
55. **Remove duplicates:** highlight the list, and go to Data tab > Data Tools panel > Remove Duplicates. Now you should have only one entry per species.
56. Highlight remaining data and **sort A-Z**. Remove any “Unidentified”, and remove any species that are only identified to genus **unless** it is the only entry for that genus (in which case retain).
57. Move species into the appropriate columns, keeping them alphabetical. Check them on [NBN database online](#) if you need to. If there are no examples of that type of plant (e.g. no lichens were found) still retain that column.
58. Select all the species by clicking and dragging over the area. In the bottom right hand corner, Excel will show you how many entries you have selected, and this is the **number of plant species found on that survey**. Make a note of that in the survey notes section on the home page.

Final Checks before Sign Off

- Make sure **all text is the same font, size and colour** (Arial, 8, automatic black) with the exception of the Species List tab, which can be size 10, and the Home tab, which has fonts of various sizes.
- **Column headers should be in bold, uppercase, and centralised** both horizontally and vertically. The column header cells should be filled with this colour:



- Make sure there are **no highlighted cells and that none of them are filled with colour** apart from the first column with the titles in it. There should only be a **border** around the column header cells.
- Only the Latin species names should be in **italics**.
- Try to keep the **heights of the rows the same**, but also try and make sure all of the **text in a cell is visible** – widen a column as much as you need to. To adjust all rows, select the whole tab, go to Home tab > Cells panel > Format > Row height. 15.00 is a suitable height. You will need to adjust the header row manually after doing this (this row should be taller than other rows).
- Make sure all **extra species have been moved** into the main body of the species data on the Species Template tab.
- On the home tab, in the **survey notes** make sure to include:
 - If this is the baseline, 1st, 2nd, or 3rd (etc) survey;
 - How many plots were completed, and whether they were vegetation or woodland plots;
 - If any plots were missed and why, and if any plots were relocated.
 - If new trees were marked;
 - If high accuracy GPS was used on the survey to relocate plots;
 - Any other big issues with the survey;
 - How many plant species were found on the survey.
- On whole plot data make sure that the **eastings and northings are entered to 3 decimal points** even if they are not high accuracy GPS'd, just so that they are all in the same format. Also, it makes it easier to see if the plots were located using high accuracy GPS or not.
- There should be **no units** in the columns of data. Double check that percentages do not have % in the column.
- Make sure that on all tabs the **plot numbers go from lowest to highest**.
- **Check that the mcode (VC or VW)** listed for each plot on each tab matches that listed for that plot on the Whole Plot Data tab.
- Where they hold the same variable, make sure all cell columns **are roughly the same width** across different tabs, e.g. plot ID.
- Some species codes may not come up because there is a **space after the species name**, this needs to be rectified in order to gain the species code. If this has affected a lot of your species names, you can use **Excel's trim function** to help: you will need to make a new column next to the species list, and enter **=TRIM(A1)**, where A1 is the cell you need to trim. This will give you the names without a leading or trailing space. You then need to copy and paste them back into the original species names column **as a value** (not as formula).

- At this stage **don't remove the names** of surveyors, creator, etc.
- Always have a **final QA code** if a QA code is used on a line of data.
- **Delete any empty lines and empty columns** on the bottom and on the end of the data vegetation spreadsheet. This is to reduce the size of the dataset. To do this, go to the first empty row and select it. Then press Ctrl + shift + ↓ and all the unused rows should be highlighted. Go to the Home tab > Cells panel > Delete > Delete Sheet Rows. Nothing will appear to have happened, but it reduces the size of the file. To delete extra columns, highlight the first empty column and press Ctrl + shift + →. Then go to the Home tab > Cells panel > Delete > Delete Sheet Columns.
- Make sure that **any data that uses a formula (including species code) is copy and pasted back into its cells as a value**. No formulas should remain in the dataset by the end of this stage.
- For the contractors, when filling in the columns for NVC first, second and third you need to enter the NVC value that Mavis returns **and the percentage value**, with a dash separating them. The NVC class column is for yourself to make a judgement on what the NVC class actually was:

NVCCLASS	NVC_FIRST	NVC_SECOND	NVC_THIRD
W10d	W21a-15.9	W21-15	W8d-14.8
W15a	W15a-25	W15-24.9	W12c-23.5
W16a	W16a-36.2	W10d-32.7	W16-32.2

- Three QA codes mean that the data is **unreliable for analysis**. Firstly, **QA766** means that the species can't be identified, so cannot be used for analysis. Secondly, the code **QA777** is used to show that the abundance data is unreliable for analysis. Lastly, the code **QA788** and the rarest code to use is to describe data that is unreliable in another way.
- Always, **write in data entry comments** column if you have changed any of the data, or what data is missing. Include the original value that was entered if you change anything.
- All tabs should be **zoomed into 100%**.
- Make sure that once you have finished working with the data that you leave each tab with the **cell in the top left hand corner of selected**.
- Make your final save with the **home tab showing**.
- This should now be **sent to the Data Lead and to the species experts** for that site so they can check the QA, and to make sure that all the correct spellings are used, that species are placed into the correct categories, and that there are no unlikely species for that habitat or region.

Section 3: Ready for Publication

Once the Data Lead and experts have sent feedback and signed off on the dataset, a little further processing is needed before publication.

1. Resave the dataset as
Sitename_LTMN_Vegetation_Yearsurveywascompleted_checked_alldata.
2. Working with the Data Lead, **address any concerns** that have been raised by them or the experts. For instance, if species seems unlikely for that site, it needs the code **QA999**. Sometimes the experts or Data Lead feel the Broad or Priority Habitats have been misidentified, in which case add a comment to that effect and the code **DI990**. If the Lead or experts are confident about identifying the correct habitat, make the change, along with a comment, and the codes **DI990** and **QA901**.
3. All tabs should be **zoomed into 100%**.
4. Make sure that once you have finished working with the data that you leave each tab with the **cell in the top left hand corner selected**.
5. Make your final save with the **home tab showing**.
6. Resave the dataset as
Sitename_LTMN_Vegetation_Yearsurveywascompleted_published_alldata.
7. **Remove any sensitive species**, in order to protect their locations, using this NBN Atlas list to search <https://lists.nbnatlas.org/speciesListItem/list/dr804>. The entire line of data should be removed from the Species Template tab, so there was no sign that they were found. Remember to remove them from the Species List tab also.
8. **Remove all surveyor and creator names from the dataset**. This includes obviously the Surveyor column on the Whole Plot Data, and the creator name on the Home tab (just leave these fields blank). Also remove your/your colleague's names from the Change log on the Home tab, and check the Field Sheets Notes, Data Entry Comments and Location Remarks column on the Whole Plot Data tab, as sometimes these comments mention people by name. If you find a name, change it to something like "Surveyor".
9. Open the raw_alldata version of the dataset and **perform 10 random cross-checks** of data from each tab between the two datasets. This should highlight any gross changes that have occurred accidentally during the process.
10. **Compare the Eastings, Northings and BNG grid reference** with those listed on the Master Plot Checker. If they do not match, you will need to investigate with the Data Lead to work out which is right and what the source of error might be.
11. **Compare the survey date range** with that recorded on the Master Audit Spreadsheet. If they disagree, the Master Audit should be changed, as the dataset is more likely to be correct.
12. As with the previous versions,

- a. All tabs should be **zoomed into 100%**.
- b. Make sure that once you have finished working with the data that you leave each tab with the **cell in the top left hand corner selected**.
- c. Make your final save with the **home tab showing**.

13. The dataset is now ready for publication.

APPENDIX

DATA_ISSUE	DESCRIPTION	QA_CODE	DESCRIPTION
999	Free-text information is associated with this data record	999	Free-text information is associated with this data record
DI800	Abundance: % cover missing	QA600	Abundance: % cover modified / inferred
DI801	Abundance: cell data missing	QA601	Abundance: cell data modified / inferred
DI802	Abundance: % cover and cell data missing	QA602	Abundance: % cover and cell presence modified / inferred
DI803	Abundance: unclear/uncertain tick, cross or % cover	QA603	Abundance: uncertain match with taxon
DI804	Abundance: two or more records share cell/% cover data	QA604	Taxon: confirmed post survey by surveyor
DI805	Abundance: Cell data / %cover does not comply with survey methodology	QA605	Taxon: confirmed post survey by sample
DI806	Taxon: name unreadable/incorrect	QA606	Taxon: confirmed post survey by checking protocol
DI807	Taxon: '?' after species' name	QA607	Taxon confirmed to genus only
DI808	Taxon: labelled as 'sample' but no sample returned/supplied	QA608	Taxon: added post-survey from sample return
DI809	Taxon/Abundance: species recorded more than once on field form	QA610	Taxon presence recorded as 'dead'
DI810	Sample returned: more than one sample in bag	QA766	Taxon unconfirmed/unknown: unreliable for analysis
DI813	Tree species: stage recorded on field form as seedling	QA777	Abundance: data unreliable for analysis
DI814	Tree species: stage not recorded on field form, assumed in ground layer	QA788	Other: data line unreliable for analysis
DI815	Taxon: not specified if seedling	QA900	Last day of survey only recorded
DI900	Plot data: some data not recorded/missing	QA901	Plot data: modified / inferred
DI990	Plot data: other issue	QA902	Plot data: slope form modified / inferred
DI910	Data issue: vegetation height does not comply with survey methodology	QA903	Plot data: slope form does not comply with survey methodology
DI911	Data issue: vegetation height not clear	QA999	Uncommon species for site/likely errors: needs checking
DI912	Data issue: tree measurements appear incorrect		
100	No information available - data lost.		
101	No sample/reading taken - equipment out of action/unable to visit equipment		
102	Sample lost or inadvertently discarded		
103	Partial loss of sample		
200	Adverse weather conditions affected sampling/recording		
201	Biting insects affected sampling/recording		
202	Failing light affected sampling/recording		
230	Anthill present		
231	Rabbit scrape present		
235	Mole-hill present		
236	Approximate or estimated count		
237	Tree dead		
238	Stem dead		
239	Taxon identified with high confidence		
240	Taxon identified with medium confidence		
241	Taxon identified with low confidence		

This table from the lookup shows all of the data issue (DI) and quality assurance (QA) codes that can be attached to lines of data.

DOM_CODE	DOM_DESC
C	canopy dominant
S	subdominant
I	intermediate
U	suppressed
H	shrub layer
A	sapling
E	seedling

The types of dominance codes that are decided by surveyors out in the field, entered on the dominance template tab.

CODE	DESCRIPTION	CODE	DESCRIPTION
999	Free-text information is associated with this data record	551	Quadrat feature - other
100	No information available - data lost.	131	Trampling during sampling period
101	No sample/reading taken - equipment out of action/unable to visit equipment	141	Grazing by sheep
102	Sample lost or inadvertently discarded	142	Grazing by cattle
103	Partial loss of sample	143	Grazing/browsing by deer
200	Adverse weather conditions affected sampling/recording	144	Grazing - other
201	Biting insects affected sampling/recording	114	Bonfire in vicinity during sampling period
202	Failing light affected sampling/recording	115	Heather burning in vicinity during sampling period
230	Anthill present	116	Forest fire in vicinity during sampling period
231	Rabbit scrape present	117	Straw burning in vicinity during sampling period
235	Mole-hill present	118	Crop spraying in vicinity during sampling period
236	Approximate or estimated count	119	Construction work in vicinity during sampling period
237	Tree dead	120	Liming in vicinity during sampling period
238	Stem dead	121	Change of land use in vicinity
239	Taxon identified with high confidence	138	Mowing of survey area
240	Taxon identified with medium confidence	139	Muck/slurry/slag application
241	Taxon identified with low confidence	140	Application of chemicals (details should be supplied)
136	Cell not surveyed due to obstruction	154	Rolling of survey area
145	Woodland Management: Coppicing	133	Evidence of disease in plot/transect section
146	Woodland Management: Thinning	134	Significant disturbance in plot/transect section
147	Woodland Management: Clear Felling	137	Flooding of survey area
148	Woodland Management: Brashing	150	Animal path present
149	Wind-throw	151	Forest ride present
130	Plot/transect section not surveyed	152	Human path present
132	Plot not accurately relocated	153	Path present (unspecified)
550	Quadrat feature - fallen dead wood		

Scode codes are used to describe animal disturbance and human disturbance. These are entered into the **whole plot data tab** from the information provided by the surveyor in the field comments.