

Dataset Documentation

Freshwater river diatom surveys (Biosys)

17/06/2022

This document will help you to understand and use the Freshwater river diatom survey dataset held in Biosys

Dataset description

This is for Approval for Access product Afa307.

Biosys is a database for storing, manipulating and reporting data from freshwater and marine biological surveys at any taxonomic level. These samples are collected and analysed by the Environment Agency and by third parties.

All the relevant data associated with Biosys freshwater macrophyte surveys, such as site information, sample information and analysis information can be obtained from the [Ecology and Fish data explorer](#). The data can be accessed using the interactive map to produce small downloadable files or they can be downloaded in bulk as a set of relational CSVs that can be linked via unique codes in each file.

The data downloads are split into three different files within the Ecology and Fish data explorer: Sites, metrics and taxa. You will see them labelled in the downloads as follows:

- DIAT_OPEN_DATA_SITE (Contains freshwater diatom site information)
- DIAT_OPEN_DATA_METRICS (Contains freshwater diatom sample and metric information)
- DIAT_OPEN_DATA_TAXA (Contains freshwater diatom analysis and taxa information)

If you are bulk downloading the full dataset there is also a fourth file available which contains additional taxonomic information for all of the taxa listed in our database:

- OPEN_DATA_TAXON_INFO (Contains taxonomic information for all taxon names available in Biosys)

The tables can be related to each other using the joins below:

- DIAT_OPEN_DATA_SITE.SITE_ID = DIAT_OPEN_DATA_METRICS.SITE_ID
- DIAT_OPEN_DATA_METRICS.ANALYSIS_ID = DIAT_OPEN_DATA_TAXA.ANALYSIS_ID
- OPEN_DATA_TAXON_INFO.TAXON_LIST_ITEM_KEY = DIAT_OPEN_DATA_TAXA.TAXON_LIST_ITEM_KEY

Please refer to the dataset content table, at the end of this document, for an explanation of the fields/columns in each file.

API

The diatom dataset can also be accessed using an API (Application Programming Interface): [Ecology and Fish Data API](#). For more information regarding the service please visit the dedicated Frequently Asked Questions section of the Defra Data Services Portal (DSP) Forum: [Ecology and Fish Data FAQs](#).

This dataset includes data from monitoring locations in England only. For Welsh survey data please contact Natural Resources Wales.

Update frequency

This dataset is extracted monthly on the first Sunday of each month. It can then take a week or so for the data to appear on the explorer. Data entered onto Biosys immediately prior to extraction may not have been quality assured and so may not reflect the final version of the data. Recent data can be identified by viewing the DATE_OF_ANALYSIS field.

Related datasets

Sites in this dataset may also be listed in the Freshwater river macroinvertebrate (Biosys) or Freshwater river macrophyte (Biosys) dataset.

Common questions & known issues

Known issues:

Reporting Area and Agency Area names:

The Reporting Area field was added to the Explorer as part of the 2022 update. The Reporting Area identifies the team who report on that data. It is more closely aligned to the [Environment Agency public face Area names](#) than the older, Agency Area field. The table below shows the comparison between all the fields.

Please be aware that the geographical boundaries for the old Southern and Thames Agency Areas do not match the geographical boundaries for the current Areas replacing them. Therefore, a small proportion of sites located in the old Area may not be located in the matched new Area:

Agency Area (Old EA Region and Area Name)	Reporting Area	EA Public Face Area Name
Anglian - Central	East Anglia - West	East Anglia
Anglian - Eastern	East Anglia - East	East Anglia
Anglian - Northern	Lincolnshire and Northamptonshire	Lincolnshire and Northamptonshire
North West - North	Cumbria and Lancashire	Cumbria and Lancashire
North West - South	Greater Manchester Merseyside and Cheshire	Greater Manchester Merseyside and Cheshire
South West - Devon	Devon and Cornwall	Devon and Cornwall
South West - Cornwall	Devon and Cornwall	Devon and Cornwall
South West - North Wessex	Wessex	Wessex
South West - South Wessex	Wessex	Wessex
Midlands - Central	West Midlands	West Midlands
Midlands - East	East Midlands	East Midlands

Midlands - West	West Midlands	West Midlands
North East - Yorkshire	Yorkshire	Yorkshire
North East - North East	North East	North East
Thames - North East	Hertfordshire and North London	Hertfordshire and North London
Thames - South East	Kent South London and East Sussex	Kent South London and East Sussex
Thames - West	Thames	Thames
Southern - Kent & E. Sussex	Kent South London and East Sussex	Kent South London and East Sussex
Southern - Solent & South Downs	Solent and South Downs	Solent and South Downs

WFD Waterbody ID:

Care should be taken when using this field to search for data. The WFD Waterbody ID field is populated on Biosys with the Water Framework Directive waterbody ID that the site *sits* in. This may be different to the Waterbody that the site is used to produce a WFD status assessment for. This field is often null for historic sites and may also be null where site sits in a waterbody that is too small to have a WFD ID. The ID may also relate to the WFD cycle 1 set of waterbodies or cycle 2 waterbody IDs. If you are interested in a particular waterbody ID then it is also worth using this field in conjunction with a search on the interface map and/or an extract of the data using the Waterbody name.

The [catchment data explorer](#) can also be used to give further high level information on a particular waterbody and could be used to compliment/support interpretation of this dataset.

Common questions:

What data are included in this dataset?

All freshwater (river) diatom samples that have been collected using our standard sampling methods and analysed in the laboratory are included. The sampling methods included in this dataset are:

- STANDARD DIATOM method for lakes and rivers (WFD Compliant)
- DIATOM TILE SCRAPE: Collection of epilithic diatoms from tiles
- DIATOM ROPE SCRAPE: Collection of epilithic diatoms from rope
- DIATOM STONE SCRAPE: Collection of epilithic diatoms from stones
- DIATOM MACROPHYTE SCRAPE: Collection of epilithic diatoms from macrophytes
- ARTIFICIAL SUBSTRATE: Any artificial substrate deployed for a fixed period of time

Please note that we do not include pollution sample data or protected species where releasing details of their location could have a detrimental effect e.g. freshwater pearl mussels.

Site name is excluded from this dataset. This is because it is a free-text field and may contain information or data that cannot be released under an Open Government Licence.

The sample reason field in this dataset displays a generic primary sample reason such as 'National monitoring' or 'Local investigative monitoring'. It does not include detail on all of the reasons for which the sample has been collected.

customer service line
03708 506 506

incident hotline
0800 80 70 60

floodline
0345 988 1188

www.gov.uk/environment-agency

All samples are analysed in a laboratory using one of two methods. Older samples collected prior to 2017 were analysed primarily using light microscopy with the analyst recording the number of diatom valves of each species observed. From 2017 onwards, analysis was primarily completed using a DNA based method known as Next Generation Sequencing (NGS). This analysis method records the diatom species present in the sample based on the relative abundances of different RbcI gene variants in the sample. The raw data from these two analysis methods are not comparable and should be separated by filtering on the different ANALYSIS_TYPE_DESCRIPTION:

- LABORATORY PRIMARY: Analysed in laboratory by primary analyst
- DNA PRIMARY: Analysed using DNA techniques

In some situations some replicates may have been collected to analyse the sample using both these analysis techniques.

Can I run these data through DARLEQ (Diatom Classification Tool)?

Yes, all the data required by DARLEQ, can be obtained from these datasets. To aid in running the data through DARLEQ ensure you use the TAXON_LIST_ITEM_KEY as this corresponds to the 'NBS code' in DARLEQ.

- **Data and DARLEQ versions**

There has been some development recently in the use of DARLEQ and the analysis of diatom samples. A briefing note with further information, released by UKtag, can be found [here](#).

DARLEQ 2 and the user guide, can be downloaded from the [UKtag Phytobenthos pages](#).

Diatom samples that have a DATE_OF_ANALYSIS after 07/03/2017 should not be used in DARLEQ 2 because of the changes we made to our taxonomy as we moved over to the TDI5/ DARLEQ 3 taxonomy. If you wish to run data through DARLEQ 3 then the tool and the user guide can be downloaded [here](#). DARLEQ 3 can be used to calculate indices and classifications on both historic and current data. This allows comparisons to be made over time where there have been changes in the taxa and indices. Please note, however that historic data will be based on a much-reduced taxa list.

- **Alkalinity**

Not all sites have alkalinity data listed. If this is the case it is still possible to run data through DARLEQ but the expected index scores may not be accurate for the site you are interested in which means any resulting Status classifications may also be incorrect. In these situations DARLEQ uses a default value for alkalinity of 150 mg/l CaCO₃. The Observed index calculations produced by DARLEQ will still be accurate however and reflective of the true trophic status of the diatom community at that site.

To help fill in any gaps, alkalinity data are available on the [Water Quality Archive](#) from our water quality monitoring points. Alkalinity does not vary that much from site to site or year to year, so it's possible to interpolate a value based on monitoring sites nearby, particularly on the same geology and/or watercourse. However please be aware that alkalinity used to generate expected index values must be at reference status to produce an accurate classification. Therefore measured values of alkalinity, especially from waterbodies that are heavily impacted by anthropogenic discharges may not be suitable for use in producing classifications and should be used with caution.

Have sample & analysis methods changed over time?

Yes, prior to the introduction of the WFD we utilised a number of sampling methods, such as the collection of diatoms from artificial substrates, particularly when carrying out surveys for the Urban WasteWater Treatment directive. Since 2006 the methodology has become standardised, as per WFD requirements. Guidance, training and quality assurance of Diatom survey methods and identification has also improved; bringing greater consistency, precision and accuracy within our datasets.

Since 2016 we have moved to using DNA based analysis methods to analyse the majority of our diatom samples. The method uses species specific variation in the RbcL gene to identify all the diatom species within a sample. The genetic sequence (or barcode), of each RbcL gene in the sample, is identified using automated high-throughput DNA sequencing technology known as Next Generation Sequencing (NGS). NGS can isolate thousands or even millions of DNA sequences in one reaction from a sample. These isolated gene sequences are then analysed and compared to a DNA barcode reference database to identify which species they belong to. This information is used to determine the species composition and relative abundance of diatom taxa in a sample. This process of matching high quantities of sequenced gene barcodes to a species database with NGS technologies is known as 'metabarcoding'.

Over the years there have also been changes to the diatom taxonomy to stay up-to-date with the different analysis methods and the various iterations of DARLEQ and the Trophic Diatom Index (TDI). This means there have been changes over the years to the taxa that we record as part of a diatom analysis.

What are the abundance values in the dataset?

The abundance values are displayed in the field 'UNITS_FOUND_OR_SEQUENCE_READS'. For samples analysed using light microscopy the values will be the number of valves counted in the sample, per taxon name. For NGS analysis the value will be the number of sequence reads per sample, per Operational Taxon Unit (OTU). Care should be taken when using the dataset to understand what analysis type has been used as absolute abundance values across different analysis types are not comparable. However relative abundances (i.e. proportional occurrence) are comparable between the two analysis methods.

Some samples will have abundance data listed against the following taxa:

- 'Unknown diatom' - this taxon is used infrequently in light microscopy where taxa are unidentified.
- 'No blast hit (Diatom DNA analysis)' - this taxon unit is used in NGS analysis where the sequence recorded does not match a reference DNA barcode (for a match to be made it must be at least 95% similar). This means for that particular gene variant we do not have a match to species name in the diatom barcode reference database. The database currently contains circa 10% of the 2500+ known UK freshwater diatom taxa (although this percentage is increasing over time). In some samples there may be a high number of sequence reads listed against this 'taxon'. In these situations, caution should be used when interpreting observed TDI scores as they will be based on a small proportion of the diatom community only.

What Diatom indices are supplied in this dataset?

A brief overview of the indices are given below. Further information on Water Framework Directive (WFD) classifications and TDI can be found on the [UKtag Phytobenthos pages](#)

The Trophic Diatom index (TDI): this biological index uses information about the composition of the diatom assemblage to indicate the trophic status of the diatom community present. The TDI is calculated using information about the environmental preferences of the individual diatom taxa present in the sample. There are a number of different versions of the river TDI index that have been updated over time as understanding diatom assemblage has improved and NGS analysis has been developed. This dataset includes TDI4 and TDI5 LM (light microscope) indices.

TDI scores can be used to infer the trophic status of a waterbody. These scores can be compared either between sites or over time to understand how the trophic status of a catchment is changing through space and time.

We can also compare the actual (observed) TDI scores with a predicted TDI score that we would expect to find if the waterbody were in pristine condition. These predictions are generated from a model built from reference data collected at a range of pristine sites. The model uses this reference information to infer what we would expect to find at a site based on its physical characteristics. By looking at how the observed TDI scores differ from the predicted scores we can identify the degree of impact occurring in the diatom assemblage due to environmental pressures (primarily eutrophication). The comparison of the observed to expected scores is known as an Ecological Quality Ratio (EQR). EQR values can then be translated to a WFD ecological status classification result.

DARLEQ is a stand-alone tool, which we can use to generate observed and expected TDI scores, EQR values and ecological status classifications. It does this based on data from diatom sampling that is inputted into the tool.

TDI4 - This index is only produced for samples analysed using light microscopy and with a DATE_OF_ANALYSIS < 07/03/2017. This is because these TDI4 calculations are based on the taxonomic list used by DARLEQ 2. After this date our taxonomy was updated to incorporate taxonomic changes and allow TDI5 values to be calculated. Samples analysed after this date will have TDI5 values calculated only.

Additional indices associated with the sample are also provided:

- TDI4_PERCENT_PLANKTONIC - This is the percentage of planktonic taxa recorded in the sample.

customer service line
03708 506 506

incident hotline
0800 80 70 60

floodline
0345 988 1188

www.gov.uk/environment-agency

- TDI4_PERCENT_MOTILE - This is the percentage of motile taxa recorded in the sample.
- TDI4_PERCENT_PTV - This is the percentage of pollution (organic) tolerant values recorded in the sample.
- TDI4_PERCENT_SALINITY - This is the percentage of saline taxa recorded in the sample.

TDI5LM - This is the latest version of the TDI index and is produced for all samples analysed using a light microscope.

Additional indices associated with the sample are also provided:

- TDI5LM_PERCENT_PLANKTONIC - This is the percentage of planktonic taxa recorded in the sample.
- TDI5LM_PERCENT_MOTILE - This is the percentage of motile taxa recorded in the sample.
- TDI5LM_PERCENT_PTV - This is the percentage of pollution (organic) tolerant values recorded in the sample.
- TDI5LM_PERCENT_SALINITY - This is the percentage of saline taxa recorded in the sample.

We do not calculate the TDI5LM index on samples using DNA techniques. Our system is not able to calculate the TDI5NGS index either. We do however provide the associated calculations e.g. % motile proportions (TDI5_NGS_PERCENT_MOTILE) etc. To generate a TDI5NGS observed and expected score, the data can be run through DARLEQ 3.

Please note that Biosys will calculate indices for all diatom samples as long as there is at least 1 'scoring' taxon present. Care should be taken when interpreting indices, and the values produced, to ensure that there are sufficient taxonomic data present to have confidence in the output. For example we would have low confidence in the TDI calculations on samples with a TOTAL_TDI4_CELLS or TOTAL_TDI5LM_CELLS value of less than 300. For TDI5NGS calculations we should treat any TDI values based on less than 3000 Sequence reads with low confidence.

How do I use the taxonomic information?

The Natural History Museum maintain the [UK Species Inventory \(UKSI\)](#) and we use that to update our 'taxon dictionary' in Biosys. They maintain a bespoke checklist for us that limits the taxon names we can record against and also ensures we have the taxon hierarchy and names required by our calculations. This does mean that sometimes we have bespoke names or names that are slightly out-of-date. You can use our dataset to understand the latest name that we use (as this may be different in older samples), the type of taxon (e.g. macrophyte) but also the recommended NBN Taxon_Version_Key. If you wish to link to other datasets then it is the taxon_version_key (TVK) that you need.

Disclaimers:

Note: We do our best to avoid quality problems but this dataset reflects the data we hold. Our datasets may contain errors.

External websites:

We have provided links to external websites as a convenience and to provide additional information. We are not responsible for the reliability or content of those external websites. We are not liable for any loss or damage that may come from using the external website links in this document.

Sort code:

customer service line
03708 506 506
www.gov.uk/environment-agency

incident hotline
0800 80 70 60

floodline
0345 988 1188

We list the Sort code in Biosys which is derived from the Maitland list, owned and maintained by [CEH](#). Our sort code list contains omissions, additions and errors and should *not* be used as a definitive list. It can be used as an aid to ordering taxa in a taxonomic hierarchy.

customer service line
03708 506 506

www.gov.uk/environment-agency

incident hotline
0800 80 70 60

floodline
0345 988 1188

Dataset content

Description of column headings

Field name	Description	Freshwater Diatom Site	Freshwater Diatom Metrics	Freshwater Diatom Taxa	Taxon information
AGENCY_AREA	EA Area name as given in Biosys for site (e.g. NORTH WEST - NORTH)	Yes		Yes	
REPORTING_AREA	EA Area name based on more-up-to-date Areas e.g. Cumbria and Lancashire.	Yes		Yes	
CATCHMENT	Catchment in Biosys, free-text field (e.g. LUNE)	Yes			
WATERBODY_TYPE_DESCRIPTION	Waterbody type recorded in Biosys from a pick-list (e.g. RIVER: Natural/semi-natural flowing fresh watercourse)	Yes			
WATER_BODY	Water body in Biosys, free-text field (e.g. LUNE (NW CENTRAL))	Yes			
SITE_ID	Unique numeric Site identifier generated when the site was created in Biosys (e.g. 66155)	Yes	Yes		
SITE_VERSION	Version number of site in Biosys. Together with the SITE_ID this denotes a unique site record.	Yes			

NGR_PREFIX	Two letter prefix of National Grid Reference (e.g. SS)	Yes			
EASTING	Five-digit Easting of NGR (e.g. 58450)	Yes			
NORTHING	Five-digit Northing of NGR (e.g. 03430)	Yes			
NGR_10_FIG	Ten-digit National Grid Reference (e.g. SS5845003430)	Yes			
FULL_EASTING	Full Easting of the site (e.g. 258450)	Yes			
FULL_NORTHING	Full Northing of the site (e.g. 103430)	Yes			
WFD_WATERBODY_ID	WFD ID code given in Biosys (e.g. GB112072065980).	Yes			
ALKALINITY	Alkalinity (mg/l)	Yes			
BASE_DATA_DATE	Date the site Base Data (e.g. Alkalinity) was entered against the site in Biosys (e.g. 01/01/1995)	Yes			
MIN_SAMPLE_DATE	The first Diatom sample date at the site.	Yes			
MAX_SAMPLE_DATE	The latest Diatom sample date at the site.	Yes			
COUNT_OF_SAMPLES	The number of Diatom samples taken at the site.	Yes			
ECN_SITE_INV	This field identifies if a site is sampled for diatoms as part of the Environmental Change Network.	Yes			
SAMPLE_ID	Sample ID generated when the sample is added to Biosys (e.g. 282922)		Yes		

customer service line
03708 506 506

incident hotline
0800 80 70 60

floodline
0345 988 1188

www.gov.uk/environment-agency

SAMPLE_VERSION	Version number of sample in Biosys. Together with the SAMPLE_ID this denotes a unique sample record.		Yes		
REPLICATE_CODE	A free-text field used to differentiate between replicate samples taken at the same site, date and time.		Yes		
SAMPLE_DATE	Date the sample was taken (e.g. 05/05/1994)		Yes		
SAMPLE_TYPE_DESC RIPTION	The type of sample collected (e.g. FRESHWATER: ALGAE All algal samples (including phytoplankton and diatom).)		Yes		
SAMPLE_METHOD_D ESCRIPTION	The type of sample method used (STANDARD DIATOM method for lakes and rivers (WFD Compliant))	Yes	Yes		
SAMPLE_REASON	Displays a generic sample reason e.g. National/Investigative/Local purposes etc.		Yes		
ANALYSIS_ID	Analysis ID created when analysis has been added against the sample (e.g. 264143)		Yes	Yes	
ANALYSIS_TYPE_DE Scription	Analysis type description as recorded in Biosys e.g. LABORATORY PRIMARY: Analysed in laboratory by primary analyst		Yes		
ANALYSIS_METHOD_ DESCRIPTION	Analysis method description as recorded in Biosys e.g. SULPHURIC/OXALIC ACIDS: Sulphuric/oxalic acids		Yes		

customer service line
03708 506 506

incident hotline
0800 80 70 60

floodline
0345 988 1188

www.gov.uk/environment-agency

TOTAL_UNITS_FOUN D_LM	The total number of valves in the analysis counted using light microscopy		Yes		
TDI4	The TDI4 (Trophic Diatom Index). This is only calculated for samples analysed before 07/03/2017 using a light microscope.		Yes		
TOTAL_TDI4_CELLS	The total number of TDI4 diatom scoring taxa valves in the sample.		Yes		
TDI4_PERCENT_PLA NKTONIC	The % of planktonic valves in the sample.		Yes		
TDI4_PERCENT_MOTI LE	The % of motile valves in the sample.		Yes		
TDI4_PERCENT_PTV	The % of organic/pollution tolerant valves in the sample.		Yes		
TDI4_PERCENT_SALI NITY	The % of Saline tolerant valves in the sample.		Yes		
TDI5LM	The TDI5 (Trophic Diatom Index). This is only calculated for samples analysed under a light microscope.		Yes		
TOTAL_TDI5LM_CELLS	The total number of TDI5 diatom scoring taxa valves in the sample.		Yes		
TDI5_LM_PERCENT_ PLANKTONIC	The % of planktonic valves in the sample.		Yes		
TDI5_LM_PERCENT_ MOTILE	The % of motile valves in the sample.		Yes		
TDI5_LM_PERCENT_ PTV	The % of organic/pollution tolerant valves in the sample.		Yes		

customer service line
03708 506 506

www.gov.uk/environment-agency

incident hotline
0800 80 70 60

floodline
0345 988 1188

TDI5_LM_PERCENT_SALINITY	The % of Saline tolerant valves in the sample.		Yes		
TOTAL_NGS_SEQUENCE_READS	The total number of sequence reads per taxon in the sample.		Yes		
TOTAL_TDI5_NGS_SEQUENCE_READS	The total number of sequence reads per TDI5-Scoring taxa in the sample.		Yes		
TDI5_NGS_PERCENT_PLANKTONIC	The % of planktonic valves in the NGS sample.		Yes		
TDI5_NGS_PERCENT_MOTILE	The % of motile valves in the NGS sample.		Yes		
TDI5_NGS_PERCENT_PTV	The % of organic/pollution tolerant valves in the NGS sample.		Yes		
TDI5_NGS_PERCENT_SALINITY	The % of Saline tolerant valves in the NGS sample.		Yes		
IS_THIRD_PARTY_DATA	Lists whether the data is from a third-party organisation. Note: a 'No' entry may still mean data has been provided by a third-party (particularly for historic data) as this is a relatively new field in Biosys.		Yes		
TAXON_LIST_ITEM_KEY	This is a unique ID, assigned by the NHM, to each name in the EA checklist. This code is required by Biosys.			Yes	Yes
TAXON_NAME	Name of the taxon unit e.g. <i>Achnanthidium minutissimum</i>				Yes
UNITS_FOUND_OR_SEQUENCE_READS	Units found (Analysis under light microscope) or sequence reads (NGS analysis) per taxon.			Yes	

customer service line
03708 506 506

incident hotline
0800 80 70 60

floodline
0345 988 1188

www.gov.uk/environment-agency

TAXON_VERSION_KEY	This is a unique ID, assigned by the NHM to different versions of a taxon name e.g. where authority or context (e.g. <i>sensu stricto</i>) is different. This is the NBN code that can be used to link to data with other organisations & the National Biodiversity Network.				Yes
AUTHORITY	Naming authority e.g. (L.) Gaertn.; Linnaeus, 1758.				Yes
SORT_CODE	This is a code (also known as the maitland code) defined by CEH to enable us to sort taxa in a hierarchy. NOTE: our list is not the definitive list and contains omissions/errors. The latest list can be downloaded from the CEH website.				Yes
TAXON_NAME_CURRENT	Yes/No field to denote where the taxon is the most current name in use on our system.				Yes
PREFERRED_TAXON_NAME	This is the most current and preferred name for the taxon on Biosys. If the TAXON_NAME_CURRENT field is 'No' then you will see the latest name listed here.				Yes
PREFERRED_NAME_TVK	The Taxon Version Key for the Biosys preferred taxon name. This is the NBN code that can be used to link to data with other organisations & the National Biodiversity Network.				Yes

customer service line
03708 506 506

www.gov.uk/environment-agency

incident hotline
0800 80 70 60

floodline
0345 988 1188

NBN_RECOMMENDED_TV_K	This is the NHM recommended Taxon Version Key (taken from their Nameserver table). This is the NBN code that can be used to link to data with other organisations & the National Biodiversity Network.				Yes
PARENT_TLIK	The parent Taxon_List_Item_Key for the parent taxon in Biosys				Yes
PARENT_TAXON_NAME	The Parent taxon name.				Yes
TAXON_RANK	Taxon rank e.g. Family, species etc.				Yes
TAXON_TYPE	Lists the taxon type in Biosys e.g. Algae; Other Macroinvertebrate; Macrophyte.				Yes
TAXON_GROUP_NAME	Lists the taxon group as defined by the NHM e.g. alga; crustacean.				Yes
NON_NATIVE_SP	Lists whether the taxon is a non-native species to England.				Yes
PROTECTED_TAXA	Lists whether the taxon is a protected species. Note we use JNCC information to help us list protected species but we only flag taxa that have the potential to be impacted by activities we permit or activities we carry out.				Yes