

# Dataset Documentation

## Marine phytoplankton Dataset (Biosys)

17/06/2022

**This document will help you understand and use the Marine Phytoplankton dataset held in Biosys**

### Dataset description

This is for Approval for Access product Afa307.

BioSys is the Environment Agency database for storing, manipulating and reporting data from freshwater and marine biological surveys at any taxonomic level. This monitoring work is undertaken by the Environment Agency and by third parties.

All the relevant data associated with Marine benthic invertebrate site, sample 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. Note: this dataset may be referred to as 'TraC' (Transitional and Coastal water) or 'marine' phytoplankton but the terms mean the same thing for this dataset.

The data download is delivered in one CSV file within the Ecology and Fish data explorer. You will see it labelled in the download as follows:

- PHYT\_OPEN\_DATA\_TAXA (Contains marine phytoplankton site, sample and taxa information)

If you are bulk downloading the full dataset there is also a second 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 two datasets can be related to each other using the joins below:

- OPEN\_DATA\_TAXON\_INFO.TAXON\_LIST\_ITEM\_KEY = PHYT\_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 phytoplankton 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](#).

These data are 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.

## 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

customer service line  
03708 506 506

incident hotline  
0800 80 70 60

floodline  
0345 988 1188

[www.gov.uk/environment-agency](http://www.gov.uk/environment-agency)

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

Some marine sites will be listed against '**National Marine**' for the Agency Area That does not correspond to any specific 'Area' in the country. Some 'Reporting\_Area' fields may be listed as 'unknown' if the site is located a significant distance offshore.

### 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 marine phytoplankton samples that have been collected using our standard methodology and analysed in the laboratory are included in this dataset.

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.

### How are the samples analysed?

Phytoplankton samples are sent to a specialist marine contractor for species identification and abundance. There they are analysed (identified and counted) using the Utermöhl method or Sedgewick Rafter chamber if particularly turbid.

Two counting methods are used when analysing phytoplankton samples, depending on the cell density of phytoplankton. Counting strategy is listed against each taxon:

- The **whole chamber counting strategy** is the preferred method. The entire base of the chamber is viewed for the presence of phytoplankton cells. This method is used when the phytoplankton that has been identified appears in low cell concentrations, for example less than approximately 4 cells per field of view.
- The **random field counting strategy** is used when cells of a particular species or genus are too numerous to ensure accurate counting over the whole base of the chamber, for example greater than approximately 4 cells per field of view. For this strategy, a random 10 fields of view are counted.

customer service line  
03708 506 506

incident hotline  
0800 80 70 60

floodline  
0345 988 1188

[www.gov.uk/environment-agency](http://www.gov.uk/environment-agency)

It is possible for both techniques to be used when counting the same sample if different species or genera are present in different cell concentrations. In the data you will see an analysis\_count\_strategy\_desc field which shows the default strategy applied to all/the majority of the sample. Each taxon however will also display a taxon count strategy, which could be different from the analysis\_count\_strategy\_desc where a different strategy was required for that taxon.

Phytoplankton cells are identified to the lowest practical taxonomic level possible. UK WFD authorities use a reduced (truncated) phytoplankton taxon list.

The resultant data is processed through a phytoplankton tool to provide a phytoplankton assessment that can be used for WFD classification of estuarine and coastal water bodies.

Further information is available from the [UK-TAG WFD website](#).

### Why are there both site and sample grid references?

In Biosys we store multiple samples against a site. For many of our marine survey techniques we record an NGR at the site because this acts as a central/representative location around which we collect lots of samples. An NGR is then also recorded against each of the samples to give a more accurate sampling location. In the map interface of the Ecology and Fish data explorer, if a sample has a sample NGR then that NGR is used to display the point on the map, otherwise the default is to use the site NGR.

### 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:

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.

## Dataset content

### Description of column headings

Field name	Description	Marine phytoplankton taxa	Taxon information
<b>AGENCY_AREA</b>	EA Area name as given in Biosys for site (e.g. NORTH WEST - NORTH)	Yes	
<b>REPORTING_AREA</b>	EA Area name based on more-up-to-date Areas e.g. Cumbria and Lancashire.	Yes	
<b>SEA AREA</b>	Sea Area in Biosys, free-text field (e.g., LIVERPOOL BAY)	Yes	
<b>WATERBODY_TYPE_DESCRIPTION</b>	Waterbody type recorded in Biosys from a pick-list (e.g. ESTUARY: Lower section of a river between freshwater limit and the sea)	Yes	
<b>WATER_BODY</b>	Water body in Biosys, free-text field (e.g. RIBBLE OFFSHORE)	Yes	
<b>SITE_ID</b>	Unique numeric Site identifier generated when the site was created in Biosys (e.g. 148190).	Yes	
<b>SITE_VERSION</b>	Version number of site in Biosys. Together with the SITE_ID this denotes a unique site record.	Yes	



<b>SITE_NGR_PREFIX</b>	Two letter prefix of National Grid Reference (e.g. SS) for the site.	Yes	
<b>SITE_EASTING</b>	Five-digit Easting of NGR (e.g. 58450) for the site.	Yes	
<b>SITE_NORTHING</b>	Five-digit Northing of NGR (e.g. 03430) for the site.	Yes	
<b>SITE_NGR_10_FIG</b>	Ten-digit National Grid Reference (e.g. SS5845003430) for the site.	Yes	
<b>SITE_FULL_EASTING</b>	Full Easting of the site (e.g. 258450) for the site.	Yes	
<b>SITE_FULL_NORTHING</b>	Full Northing of the site (e.g. 103430) for the site.	Yes	
<b>WIMS_SITE_ID</b>	This is the site reference for our WIMS water quality database e.g. LC544405. Data from WIMS can be found on the <a href="#">Water Quality Archive</a> pages	Yes	
<b>WFD_WATERBODY_ID</b>	WFD ID code given in Biosys (e.g. GB112072065980).	Yes	
<b>SAMPLE_ID</b>	Sample ID generated when the sample is added to Biosys (e.g. 567642)	Yes	
<b>SAMPLE_VERSION</b>	Version number of sample in Biosys. Together with the SAMPLE_ID this denotes a unique sample record.	Yes	
<b>REPLICATE_CODE</b>	A free-text field used to differentiate between replicate samples taken at the same site, date and time.	Yes	

customer service line  
03708 506 506

incident hotline  
0800 80 70 60

floodline  
0345 988 1188

[www.gov.uk/environment-agency](http://www.gov.uk/environment-agency)

<b>SURVEY_CODE</b>	Unique survey code that links a subset of surveys.	Yes	
<b>SAMPLE_NGR_PREFIX</b>	Two letter prefix of National Grid Reference (e.g. SS) for the <b>sample</b> NGR	Yes	
<b>SAMPLE_EASTING</b>	Five-digit Easting of <b>sample</b> NGR (e.g. 58450)	Yes	
<b>SAMPLE_NORTHING</b>	Five-digit Northing of <b>sample</b> NGR (e.g. 03430)	Yes	
<b>SAMPLE_NGR_10_FIG</b>	Ten-digit National Grid Reference of the <b>sample</b> location (e.g. SS5845003430)	Yes	
<b>SAMPLE_FULL_EASTING</b>	Full Easting of the <b>sample</b> location (e.g. 258450)	Yes	
<b>SAMPLE_FULL_NORTHING</b>	Full Northing of the <b>sample</b> location (e.g. 103430)	Yes	
<b>SAMPLE_DATE</b>	Date the sample was taken (e.g. 05/05/1994)	Yes	
<b>SAMPLE_TYPE_DESCRIPTION</b>	The type of sample collected (e.g. MARINE: PHYTOPLANKTON - Truncated List)	Yes	
<b>SAMPLE_METHOD_DESCRIPTION</b>	The type of sample method used (e.g. MARINE: Surface water sample)	Yes	
<b>SAMPLE_REASON</b>	Displays a generic sample reason e.g. National/Investigative/Local purposes etc.	Yes	
<b>ANALYSIS_ID</b>	Analysis ID created when analysis has been added against the sample (e.g. 562376)	Yes	

customer service line  
03708 506 506

incident hotline  
0800 80 70 60

floodline  
0345 988 1188

[www.gov.uk/environment-agency](http://www.gov.uk/environment-agency)

<b>DATE_OF_ANALYSIS</b>	Date the sample was analysed.	Yes	
<b>ANALYSIS_TYPE_DESCRIPTION</b>	Analysis type description as recorded in Biosys e.g. LABORATORY PRIMARY: Analysed in laboratory by primary analyst	Yes	
<b>ANALYSIS_METHOD_DESCRIPTION</b>	Analysis method description as recorded in Biosys e.g. SEDIMENTATION CHAMBER: Sedimentation chamber (algae counts)	Yes	
<b>SAMPLE_VOLUME</b>	The volume of liquid in a marine phytoplankton sample/chamber (ml)	Yes	
<b>ANALYSIS_COUNT_STRATEGY_DESC</b>	The counting strategy used to analyse the sample e.g. Whole Chamber	Yes	
<b>ID_STRATEGY_DESC</b>	The identification strategy used to analyse the sample e.g. Full Specification	Yes	
<b>TAXON_LIST_ITEM_KEY</b>	This is a unique ID, assigned by the NHM, to each name in the EA checklist. This code is required by Biosys.	Yes	
<b>TAXON_NAME</b>	Name of the taxon unit e.g. <i>Ditylum brightwellii</i>		Yes
<b>TAXON_QUALIFIER_DESC</b>	Additional information describing an attribute of the taxon such as the size or structure of the cell. E.g. <5um_cluster	Yes	
<b>SACFOR_ABUNDANCE</b>	Selected from a picklist (see tab) of <a href="#">SACFOR</a> categories e.g. Rare to denote taxon abundance. This	Yes	

customer service line  
03708 506 506

incident hotline  
0800 80 70 60

floodline  
0345 988 1188

[www.gov.uk/environment-agency](http://www.gov.uk/environment-agency)



	is a known scale on which marine taxa can be recorded. This field is not routinely populated.		
<b>SAMPLE_VOLUME_TAXON</b>	The volume of liquid analysed at the taxon level (ml)	Yes	
<b>TAXON_SETTLING_TIME</b>	Number of hours that the sample has been left to settle ready for taxon analysis e.g. 24	Yes	
<b>TAXON_COUNTING_STRATEGY_DESC</b>	The counting strategy employed when analysing a marine phytoplankton taxon	Yes	
<b>CELLS_LITRE</b>	Number of cells per Litre	Yes	
<b>COLS_LITRE</b>	Number of colonies per Litre	Yes	
<b>IS_THIRD_PARTY_DATA</b>	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	
<b>TAXON_VERSION_KEY</b>	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
<b>AUTHORITY</b>	Naming authority e.g. (L.) Gaertn.; Linnaeus, 1758.		Yes
<b>SORT_CODE</b>	This is a code (also known as the maitland code) defined by CEH to enable us to sort taxa in a		Yes

customer service line  
03708 506 506

incident hotline  
0800 80 70 60

floodline  
0345 988 1188

[www.gov.uk/environment-agency](http://www.gov.uk/environment-agency)

	hierarchy. NOTE: our list is not the definitive list and contains omissions/errors. The latest list can be downloaded from the CEH website.		
<b>TAXON_NAME_CURRENT</b>	Yes/No field to denote where the taxon is the most current name in use on our system.		Yes
<b>PREFERRED_TAXON_NAME</b>	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
<b>PREFERRED_NAME_TVK</b>	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
<b>NBN_RECOMMENDED_TVK</b>	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
<b>PARENT_TLIK</b>	The parent Taxon_List_Item_Key for the parent taxon in Biosys		Yes
<b>PARENT_TAXON_NAME</b>	The Parent taxon name.		Yes
<b>TAXON_RANK</b>	Taxon rank e.g. Family, species etc.		Yes

customer service line  
03708 506 506

incident hotline  
0800 80 70 60

floodline  
0345 988 1188

[www.gov.uk/environment-agency](http://www.gov.uk/environment-agency)

<b>TAXON_TYPE</b>	Lists the taxon type in Biosys e.g. Algae; Other Macroinvertebrate; Macrophyte.		Yes
<b>TAXON_GROUP_NAME</b>	Lists the taxon group as defined by the NHM e.g. alga; crustacean.		Yes
<b>NON_NATIVE_SP</b>	Lists whether the taxon is a non-native species to England.		Yes
<b>PROTECTED_TAXA</b>	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