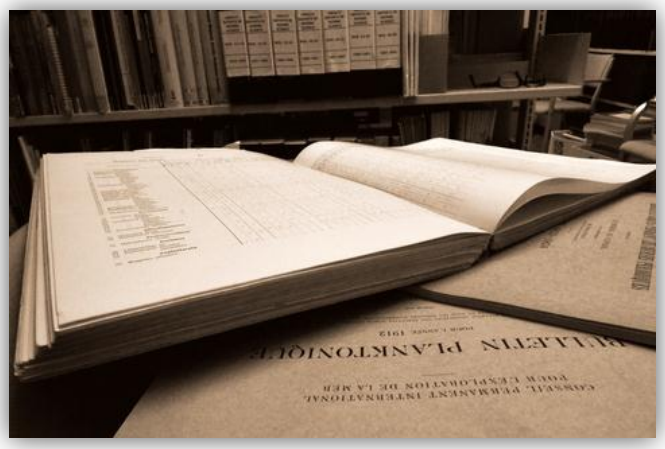
ICES Data Centre

Historical Plankton Fact Sheet

February 2013

ICES Historical Plankton Dataset - Securing and making the data available



1 Background

This project was split over 3 stages, the digitization of the data and the production of a quality controlled dataset were undertaken by SAHFOS and the University of Plymouth under a contract from ICES. The final stage of producing an operational database and online system were carried out by the ICES Data Centre. Please refer to the acknowledgements section for the details of the leaders of the project phases.

2 Dataset Overview

This database consists of quality checked data digitized from seven historical ICES volumes:

- ICES Bulletin Resultats Acquis Pendant les Courses Periodiques 1901-1902
- ICES Bulletin Resultats Acquis Pendant les Courses Periodiques 1902-1903
- Resultats Acquis Pendant les Courses Periodiques 1903-1904
- Resultats Acquis Pendant les Courses Periodiques 1904-1905
- Resultats Acquis Pendant les Courses Periodiques 1905-1906
- Resultats Acquis Pendant les Courses Periodiques 1906-1907
- Bulletin Planktonique 1907-1912.

The dataset includes 13,379 samples in the North-East Atlantic, North Sea, Irish Sea, Baltic Sea, and Arctic Sea.

2.1 Quality assurance

Data entry and quality checking were conducted from April 2008 to March 2011 by the University of Plymouth and in conjunction with the Sir Alister Hardy Foundation for Ocean Science (SAHFOS). Data tables from the historical volumes were photocopied and then entered into an online database by a team of data enterers assembled by the University of Plymouth. The digitized versions were then printed out and compared against the original

photocopies to locate any discrepancies. If mistakes were discovered, they were corrected by the database manager at the University.

3 Dataset and Database notes

3.1 Species name resolution and STAGE information

Approximately 2015 plankton taxa (with 1375 resolved to species level) have been encountered in the historical ICES data tables; this number has been greatly reduced during the course of the project due to the considerable effort invested in resolving taxa names and identifying duplicates and synonyms. As the data are a century old, many of the taxa have undergone taxonomic reclassification or changes in name. Additionally, duplications of taxa due to spelling ambiguities in the original data tables have also been identified. Resolving these names so that the historical data may be compared with contemporary data is a time consuming but crucial part of this project and as this work progresses the number of separate plankton taxa in the database will decrease. As species were resolved, the ICES historical database was updated with complete taxonomic information, including current species name. Concurrently, the WoRMS (World Register of Marine Species: http://www.marinespecies.org) AphiaID number was added to the ICES historical database. WoRMS is the taxonomic backbone of OBIS (Oceanbiogeographic Information System: http://www.iobis.org) and the Aphia IDs provide a mechanism through which to join the ICES historical data to other worldwide marine databases. Many of the taxa recorded in the historical ICES volumes did not previously exist in WoRMS and have been added (and new Aphia IDs created) during this project. If a taxa could not be resolved to species level, it was resolved to the finest taxonomic level possible based on the taxonomic group information found in the historical volumes (for example, 'Macroura' is a non-taxonomic grouping of decapods and so can only be resolved to 'Decapoda', which is an Order). Species name resolution required literature and online database searches as well as input from taxonomic experts at SAHFOS.

Appendix B provides a table of resolved species names that conflicted with entries in the WoRMS database during importing.

STAGE information has also been extracted from the original Species Names. Appendix A provides details of the transformation rules applied.

3.2 Database description

The online dataset can be found at http://ecosystemdata.ices.dk/historicalplankton

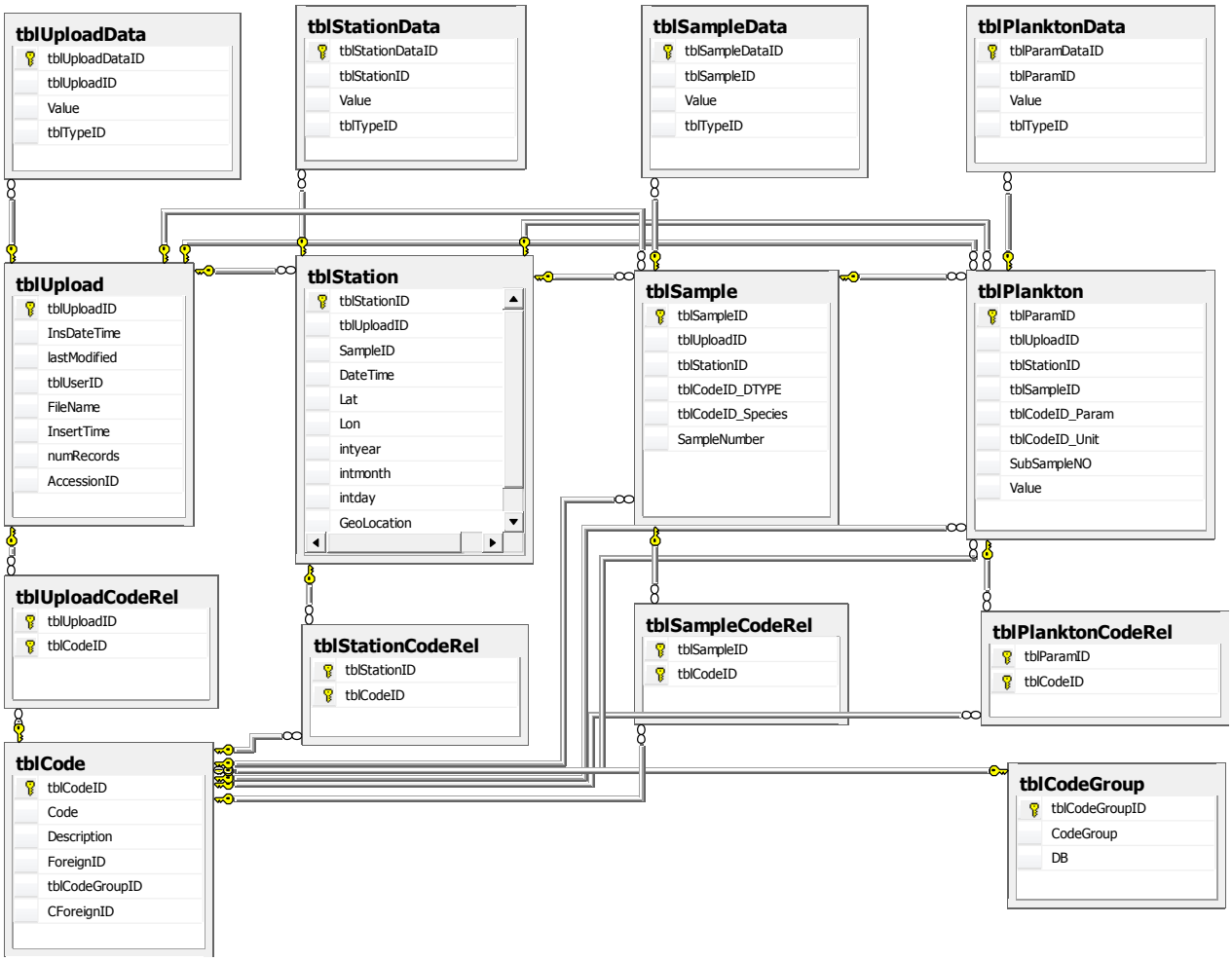


Figure 1 – Entity relation diagram of the ICES Historical Plankton database

The 'appendix A' describes the transformations applied to the data for the quality assurance of the data.

To accommodate the database provided by SAHFOS, it was necessary to perform data transformations. Normalization is a systematic way of ensuring that a database structure is suitable for general-purpose querying. The structure we can see in figure one has the advantage of allowing the data to have extra fields stored in the codes table.

Table 1 below describes the mapping from the original table, to the new field in the relational database.

Orirginal table name	Original field name	Table name	Field Name [CodeGroup/Type]
NA	NA	tblUpload	tblUploadID
OriginalTable		tblUpload	FileName
NA	NA	tblUpload	InsDateTime
OriginalTable	Book	tblUploadData	Value [Book]

OriginalTable	BookSection	tblUploadData	Value [BookSection]
OriginalTable	PageTableNumber	tblUploadData	Value [PageTableNumber]
OriginalTable	Month	tblUploadData	Value [Month]
OriginalTable	Year	tblUploadData	Value [Year]
OriginalTable	SampleArea	tblUploadData	Value [SampleArea]
OriginalTable	Ship	tblUploadData	Value [SHIP]
OriginalTable	Notes	tblUploadData	Value [Notes]
NA	NA	tblStation	tblStationID
NA	NA	tblStation	tblUploadID (FK)
StationSample	StationSampleID	tblStation	SampleID
StationSample	SamplingDate	tblStation	DateTime
Station	LatitudeDegrees + LatitudeMinutes + LatitudeDirection	tblStation	Lat
Station	LongitudeDegrees + LongitudeMinutes + LongitudeDirection	tblStation	Lon
StationSample	SamplingDate(year)	tblStation	intyear
StationSample	SamplingDate(month)	tblStation	intmonth
StationSample	SamplingDate(day)	tblStation	intday
NA	NA	tblStation	GeoLocation (geometry::STPointFromText)
Station	StationNumber	tblStationData	Value [STATN]
Station	SourceForCoordinates	tblStationData	Value [METDC]
Station	TimeStart	tblStationData	Value [STIME]
Station	TimeEnd	tblStationData	Value [ETIME]
StationSample	VerticalDepth	tblStationData	Value [WADEP]
StationSample	VerticalDepthRange	tblStationData	Value [VDRange]
StationSample	HorizontalDistance	tblStationData	Value [HDist]
StationSample	LineOut	tblStationData	Value [LineOut]
StationSample	NetType	tblStationData	Value [GEART]
StationSample	Temperature	tblStationData	Value [Temp]
StationSample	Salinity	tblStationData	Value [Salinity]
StationSample	Density	tblStationData	Value [Density]
NA	NA	tblSample	tblSampleID
NA	NA	tblSample	tblUploadID (FK)
NA	NA	tblSample	tblStationID (FK)
NA	NA	tblSample	tblCodeID_DTYPE
SampleData	SpeciesID (AphiaID)	tblSample	tblCodeID_Species
SampleData	SampleID	tblSample	SampleNumber
Species	SpeciesName	tblSampleData	Value [OgSpeciName]
Species	Extracted from Species Name	tblSampleCodeRel	Value [STAGE]
NA	NA	tblPlankton	tblParamID
NA	NA	tblPlankton	tblUploadID (FK)
NA	NA	tblPlankton	tblStationID (FK)
NA	NA	tblPlankton	tblSampleID (FK)
NA	NA	tblPlankton	tblCodeID_Param [PARAM]
NA	NA	tblPlankton	tblCodeID_Unit [MUNIT]
NA	NA	tblPlankton	SubSampleNo
SampleData	SpeciesCount	tblPlankton	Value
SampleData	OtherCount / AlsoObserved	tblPlanktonData	Value [NOTES]

The original table “SampleData” contains the plankton abundance data and is linked to the “Species” table. The two tables are joined by the field “SpeciesID”. Abundance data are stored two ways in the historical volumes: as a quantitative abundance (field “SpeciesCount” In “SampleData”), or as a non-numerical representation or present, absent or very rare to very common (field “OtherCount” In “SampleData”). The historical data was stored in the original volumes in two sections per cruise: 1) the main data table consisting of the majority of the data, usually frequently occurring taxa and 2) an ‘Also Observed’ list which was reserved for rarely occurring species in a cruise. Therefore, in the database, the field “AlsoObserved” field identifies plankton sample data that occurred in the supplementary cruise data, rather than in the main data table per cruise.

The “Species” table (Table 3) contains taxonomic information on the plankton. The fields “TaxonomicGroup” and “SpeciesName” pertain to taxonomic information found in the historic volumes while the field “NewName” contains the current (as of March 2011) taxonomic names, as identified using WoRMS, the literature, and expert advice. The “AphiaID” field links the current taxonomic information to the full taxonomic description in WoRMS (table “WoRMSSpecies”; Table 4).

The taxonomic information in “WoRMSSpecies” may change as taxonomic research progresses. However, the AphiaID values provide unique identifiers which can help update this information through the WoRMS web service.

The numerical abundance data has been assigned to the ICES reference parameter ABUNDNR (Abundance Number). The non-numerical sample data has also been mapped to ABUNDNR with a default value of 0. The original values held in the OtherCount and AlsoObserved columns have been stored in the tblPlanktonData data, in case they need to be retrieved.

3.3 Quality and potential issues with the use of the dataset

The historical ICES volumes are a compilation of data from sampling programmes conducted by multiple countries. Sampling methodologies vary widely between datasets. For example, some programmes recorded plankton species as present (X) or absent; others recorded species as very rare (rr), rare (r), common (c), very common (cc); and still others recorded species quantitatively. Even these general groupings of data recording contain inconsistencies – the quantity of organisms present to be noted as ‘rare’ is not uniform between data tables and at times is not numerically defined in the historical records at all. Additionally, the volume of water (Table: “StationSample”, field “QuantityExamined”) assessed in sampling programmes which quantitatively recorded abundances varies as does the mesh size of nets used to collect samples. Furthermore this information is often not present in the historical volumes.

Due to these vagrancies in the data, particularly lack of quantitative data and missing sample sizes, it may be difficult to use the data in a quantitative manner. It is, however, possible to analyze the ICES historical database at the semi-quantitative presence/absence

level. Many possibilities for analysis exist using semi-quantitative techniques and these include, for example, exploration of changes in range, distribution and frequency of occurrence.

4 GIS Mapping Facility

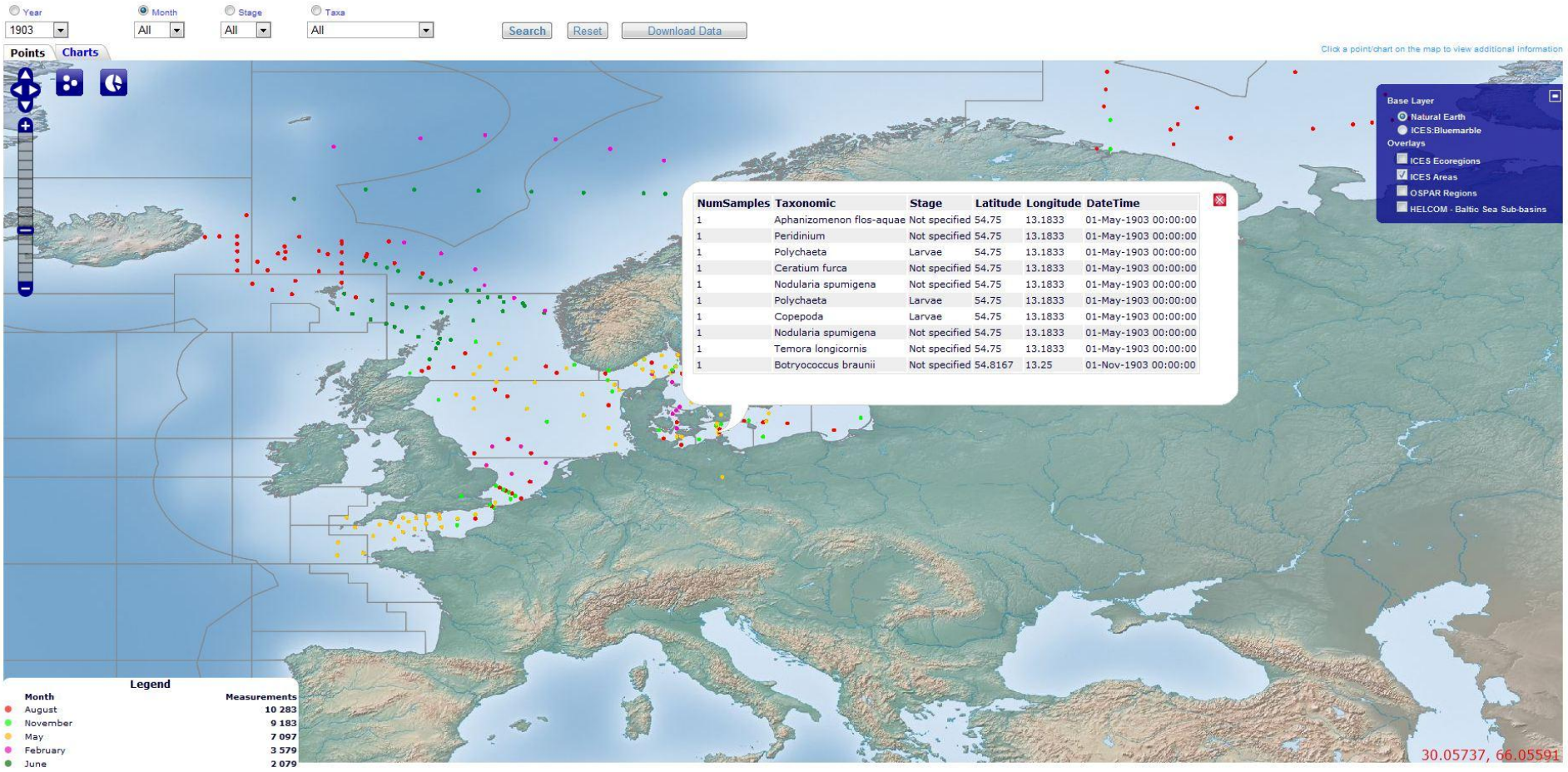


Figure 2 – A screenshot showing measurements collected during 1903 as points (Grouped by Month).

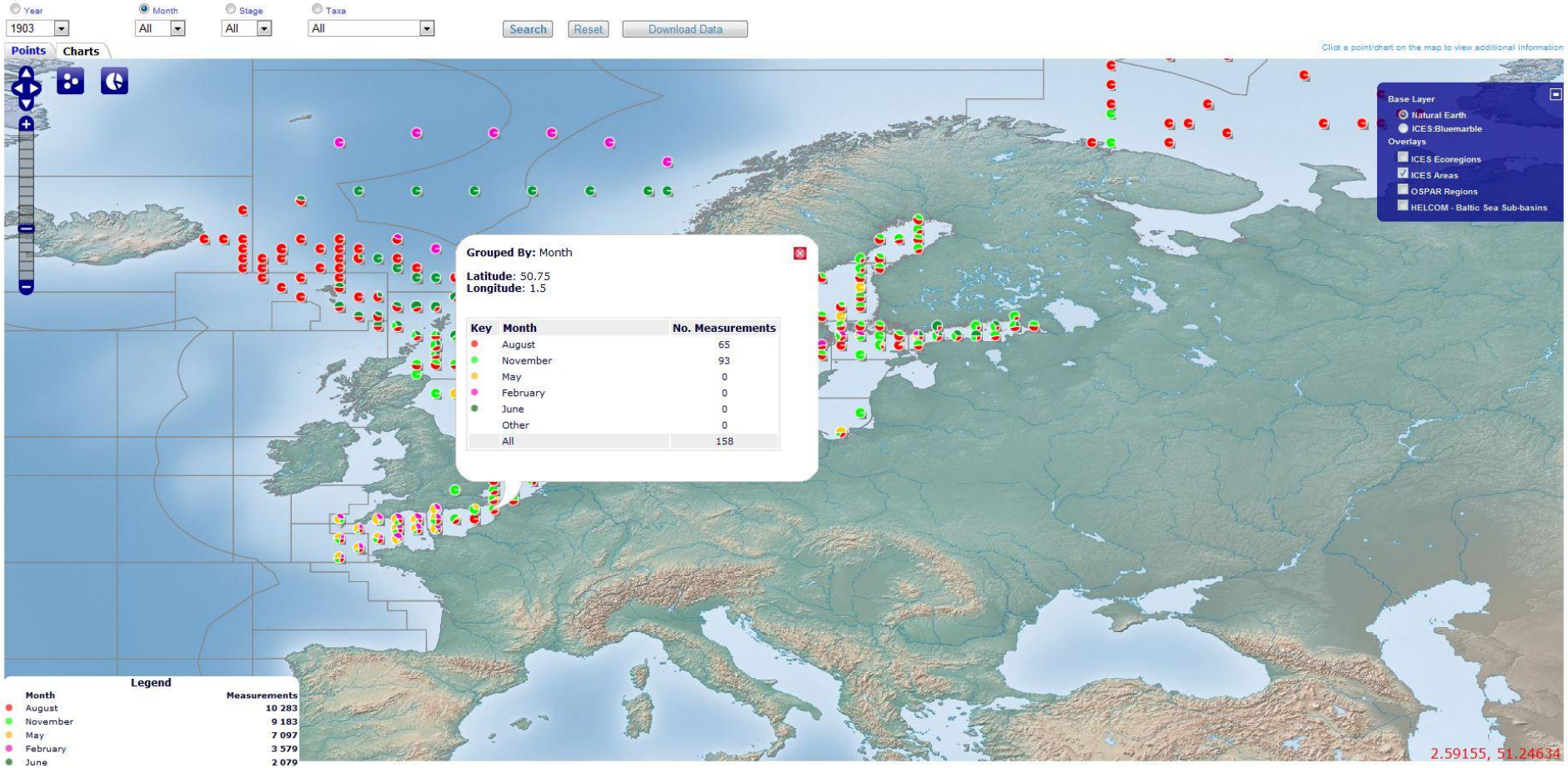


Figure 3 – A screenshot showing measurements collected during 1903 as aggregated Charts (Grouped by Month).

5 Acknowledgments, citation and publications

5.1 Publications

McQuatters-Gollop A (2010) Biological indicators across heterogeneous datasets. Invited speaker at EC Marine Observation and Data Expert Group - 10th meeting, Brussels, Belgium, September 2010.

5.2 Data Citation

Please acknowledge the following data source when using the data:

ICES Historical plankton dataset 2011, ICES, Copenhagen

Web : http://ecosystemdata.ices.dk/historicalplankton
Email: accessions@ices.dk
Tel : +45 3338 6700

Please refer to the ICES Data policy for terms and conditions of data usage and rights.
http://info.ices.dk/datacentre/datapolicy.asp

5.3 Acknowledgements

ICES is very thankful to the following institutes and individuals for their participation in this project:

Digitization and production of the quality controlled dataset

Dr Abigail McQuatters-Gollop, Science and Policy Researcher
SAHFOS
The Laboratory, Citadel Hill
Plymouth, UK
PL1 2PB

Paul Dowland, Associate Professor in Information Systems Security
School of Computing and Mathematics
University of Plymouth, Portland Square, Drake Circus,
Plymouth, UK
PL4 8AA

Production of the online map and download system

Lee Martin, Data Analyst
Carlos Pinto, Data Analyst
ICES
H. C. Andersens - Boulevard 44-46
DK-1553
Copenhagen V
Denmark

Provision of taxonomic online services

Appeltans W, Bouchet P, Boxshall GA, Fauchald K, Gordon DP, Hoeksema BW, Poore GCB, van Soest RWM, Stöhr S, Walter TC, Costello MJ. (eds) (2011). World Register of Marine Species.

http://www.marinespecies.org

6 Appendix

Appendix A

Transformations to the data

FileName:

```
SET FileName_MyField =

CAST([OriginalTableID] AS VARCHAR(5)) + '^' + REPLACE ( REPLACE( REPLACE ([Book],' ',''),
'BulletinPlanktonique','BP_'), 'ICESBulletinResultatsAcquisPendantlesCoursesPeriodiques', 'IB_')
+ '^' + REPLACE( REPLACE( REPLACE( REPLACE( REPLACE( REPLACE( REPLACE( [BookSection],',',''),'
',''), 'Pourl''anee','PL_'), 'Pourl''anNe','PL_'), 'PourlesAnnees', 'PLA_'), 'Bulletin','B_'),
'Supplement','S_') + '^p' + REPLACE(ISNULL([PageTableNumber],''),' ','_') + '^' +
REPLACE(ISNULL([Month],''),' ','_') + '^' + REPLACE( ISNULL([Year],''),' ','_') + '^' + REPLACE(
REPLACE( ISNULL([SampleArea],''),': ',':'), ' ','_') + '^' + REPLACE(ISNULL([Ship],''),' ','_')
```

STAGE:

```
SET STAGE = CASE

WHEN SpeciesName like '%larva%' OR SpeciesName like '%larve%' THEN 'LV'

WHEN SpeciesName like '%juv%' OR SpeciesName like '%juven%' THEN 'JV'

WHEN SpeciesName like '%naupli%' THEN 'NP'

WHEN SpeciesName like '%Eggs%' THEN 'EG'

ELSE 'NS'
```

Lat:

```
CASE [LatitudeDirection]
WHEN 'N' THEN
        '+' + CAST(ROUND(CAST([LatitudeDegrees] + ([LatitudeMinutes] / 60) AS REAL), 4) AS
        VARCHAR(20))
ELSE
        '-' + CAST(ROUND(CAST([LatitudeDegrees] + ([LatitudeMinutes] / 60) AS REAL), 4) AS
        VARCHAR(20))
END AS Latitude
```

Lon:

```
CASE [LongitudeDirection]
WHEN 'E'  THEN
        '+' + CAST(ROUND(CAST([LongitudeDegrees] + ([LongitudeMinutes] / 60) AS REAL), 4) AS
        VARCHAR(20))
ELSE
```

```
        '-' + CAST(ROUND(CAST([LongitudeDegrees] + ([LongitudeMinutes] / 60) AS REAL), 4) AS
        VARCHAR(20))
END AS Longitude,
```

GeoLocation:

```
SET GeoLocation = geometry::STPointFromText('POINT(' + CAST([Lon] AS VARCHAR(20)) + ' ' +
CAST([Lat] AS VARCHAR(20)) + ')', 4326)
```

Intyear: `DATEPART(Year, CONVERT(datetime, SS.SamplingDate, 103)) AS intyear`

Intmonth: `DATEPART([Month], CONVERT(datetime, SS.SamplingDate, 103)) AS intmonth`

Intday: `DATEPART(DAY, CONVERT(datetime, SS.SamplingDate, 103)) AS intday`

Appendix B

Problems found in the data

Bathymetric Checks of the Station / Positions provided

There are currently:

Number of stations/positions with an elevation above 0m = 91
Number of stations/positions with an elevation below 0m = 1328

Those stations that have an elevation above 0m have been flagged as suspicious and the elevation value has been stored in the tblStationData table. The samples taken at these stations are still provided in the download and mapping facilities.

Stations without Coordinates

Original Stations without Coordinates = 127 (Station)
Imported Station (Samples) records without coordinates = 273 (tblStation)

This means that these samples cannot be displayed on the mapping facility. They will however be included in the downloadable data.

Species Names that do not map to WoRMS

AphiaID	SAHFOS Species NewName	ERMS/WoRMS Species Name (in data and lookup table)	Modified and Inserted
102137	Guernea coalita	Guernea (Guernea) coalita	Yes
113987	Leptohyalis scottii	Leptohyalis scotti	Yes
393179	Pulvinulina menardi	Pulvinulina menardii	Yes
102111	Siphonoecetes kroyeranus	Siphonoecetes (Centraloecetes) kroyeranus	Yes
522406	Spiroloculina fragilissima **	*NO MATCH IN ERMS, maybe in WoRMS?*	*No*