ICES Year of the Stomach Dataset – securing and making the data available

Background to request

In a response to the call for WGDIM[1] to work closely to address the needs of users and empower the ICES Data Centre to deliver meaningful services, WGSAM[2] suggested the following ToR for WGDIM in 2009:

WGSAM requests that WGDIM work towards making the ICES 'Year of the Stomach' datasets for North Sea and Baltic more readily available to the ICES community. This will require the creation of a standardized and quality-controlled version of the data including an updated look-up key for prey codes.

Year of the Stomach history

The first year of the stomach ran in **1981** and covered a handful of species in the North Sea. There was some follow up data collection performed in 1985-1986 and a decade later in **1991** the 2nd year of the stomach ran. The data were collected as part of an ICES initiative and the results and analysis were presented in two cooperative reports (CRR 164 and CRR 219).

End product

WGDIM should work with ICES Data Centre and the various contributors to year of the stomach to:

- Collate the dataset(s) into one location
- Standardize references and format
- Create an online, downloadable dataset
- Make the research reports available online (with the data)
- Secure the dataset for future use by the scientific community

Chronological action of the stomach data

October 2008 - ICES DC made a call to gather copies of the stomach data

November 2008 – Morten Vinther (DTU-Aqua, Denmark) answered the call with a treated dataset of the North Sea

March 2009 – The first database of Stomach data was established and some summaries were made.

May 2009 – The draft summary of the dataset was presented in WGDIM meeting

May 2009 – Ingeborg De Boois highlighted that IMARES had part of the stomach dataset that was not present in the one presented

1 ICES Working Group on Data and Information Management

2 ICES Working Group on Multispecies Assessment Methods

September 2009 – Both datasets were compared and it was concluded that the datasets had data in common but also exclusive data

March 2010 – Both datasets were merged to have the most complete dataset

May of 2010 – The stomach data were developed and the data put online

Stomach database

The current stomach database is stored in a relational database and it is available at the address: http://ecosystemdata.ices.dk/stomachdata

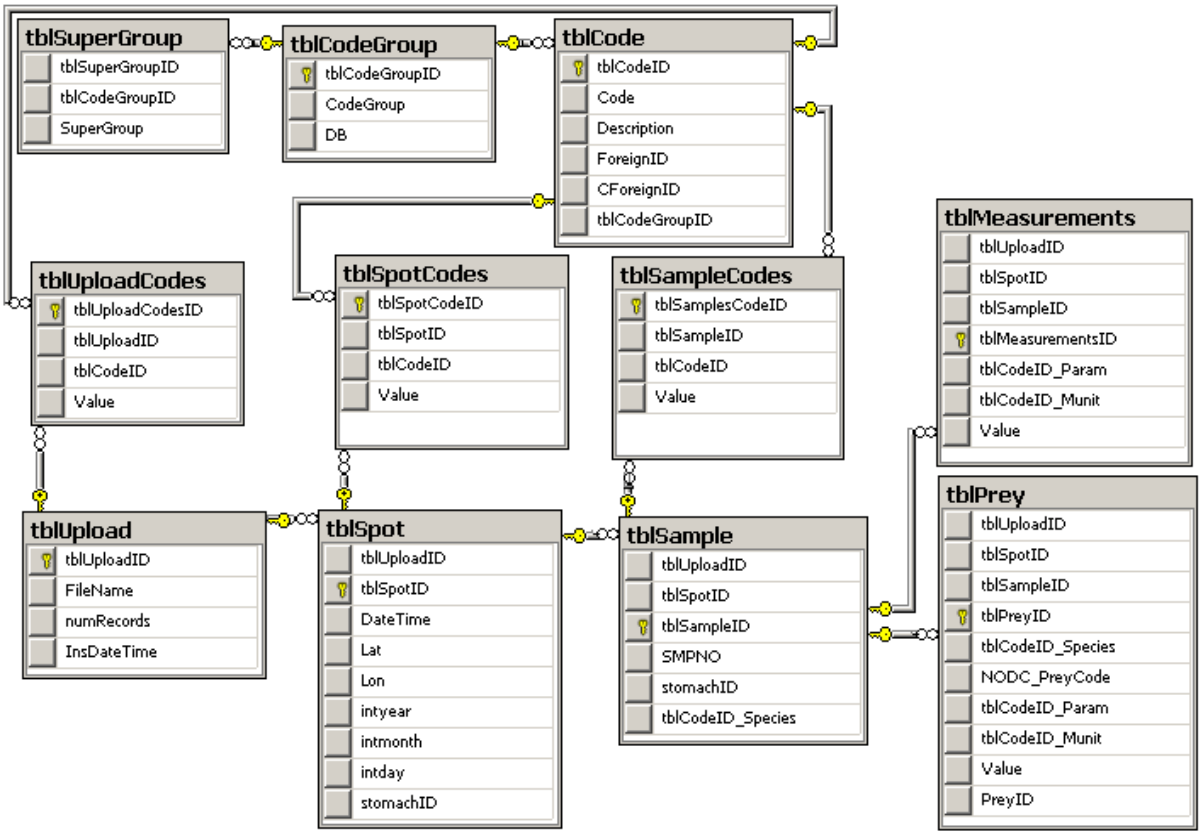


Figure 1 – Entity relation diagram of the stomach database

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

To accommodate the data (that was on a plain table) on a database, it was necessary to perform data transformations. Now the data is normalized. 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 describes the mapping from the original table, to the new field in the relational database.

Original field name	Table name	Field Name [CodeGroup]
NOT IN THE ORIGINAL DATA	tblSample	stomachID
FileName	tblUpload	FileName
Year	tblSpot	intyear
Quarter	tblSpotCodes	Value [Quarter]
Square	tblSpotCodes	Value [Square]
Predator code	tblSampleCodes	NODC
Country	tblUploadCodes	Value [Contry]
Ship	tblSpotCodes	Value [Ship]
Sampling method	tblSampleCodes	Value [SamplingMethod]
Station/Haul	tblSpotCodes	Value [Station]
Month	tblSpot	intmonth
Day	tblSpot	intday
Time of the day	tblSpotCodes	Value [Time]
Quadrant	tblSpotCodes	Value [Quadrant]
Latitude	tblSpot	Lat
Longitude	tblSpot	Lon
Depth	tblSampleCodes	Depth
Temperature	tblSampleCodes	Temperature
Predatorlengh	tblMeasurement	Value [Lengh]
PedatorWeight	tblMeasurement	Value [Weight]
PredatorAge	tblMeasurement	Value [Age]
PredatorLowerLenghBound	tblMeasurement	Value [LowerLengh]
PredatorUpperLengthBound	tblMeasurement	Value [UpperLengh]
CPUE	tblMeasurement	Value [CPUE]
NumberWithFood	tblMeasurement	Value [NumStomachsWithFood]
NumberRegurgitated	tblMeasurement	Value [NumStomachsRegurgitated]
NumberWithSkeletalRemains	tblMeasurement	Value [NumStomachsWithSkeletalRemains]
NumberEmpty	tblMeasurement	Value [NumStomachsEmpty]
PreySpeciesCode	tblPrey	NODC_PreyCode
PreyPreyLowerLengthBound	tblPrey	Value [LowerLengh]
PreyUpperLengthBound	tblPrey	Value [UpperLengh]
PreyWeight	tblPrey	Value [Weight]
PreyNumber	tblPrey	Value [Lengh]
DigestionStage	tblSampleCodes	Value [DigestionStage]
NumberStomachsAnalized	tblMeasurement	Value [NumberStomachAnalized]

Table 1 – Table describing the original filed names and where they are currently stored in the database.

Acknowledgments

Thanks to IMARES, in particular to Ingeborg for the help and support she gave during all the process of compiling and treating the data.

Thanks also to Niels Daan, he was available when needed to help and clear any doubts.
Thanks also to Morten Vinther for making the treated stomach dataset available.

Data Acknowledgement

Please acknowledge the following data sources when using the data:

ICES Stomach Dataset 2010, ICES, Copenhagen

ICES Cooperative Research Report No. 164
(http://www.ices.dk/sites/pub/Publication%20Reports/Cooperative%20Research%20Report%20(CRR)/crr164/CRR164.pdf)

ICES Cooperative Research Report No. 219
(http://www.ices.dk/sites/pub/Publication%20Reports/Cooperative%20Research%20Report%20(CRR)/crr219/CRR219.pdf)

IMARES (Institute for Marine Resources & Ecosystem Studies)

Contact:

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

Web : http://www.ices.dk/marine-data/Pages/default.aspx
Email: accessions@ices.dk
Tel : +45 3338 6700

Appendix A

Transformations to the data

Prey weight:
```
set preyweight = preyweight/1000 where filename <> 'imaresdataset'
```

Prey weight:
```
set [PreyWeight] = null where [PreyWeight] = 99999.999
```

Ship:
```
set ship = 'TRI' where ship = 'TRID'
```

Country:
```
set country = 'NED' where country = 'NET'
```

Country:
```
set country = null where country = 'XXX'
```

Country:
```
set country = null where country = 'ALL'
```

Country:
```
set country = 'GFR' where country = 'GER'
```

Country:
```
set country = null where country = 'M'
```

Day:
```
set [day] = null where [month] is null
```

Time of the day:
```
set [Time of the day] = NULL where left([Time of the day],2) = 99
```

Time of the day:
```
set [Time of the day] = null where cast(right([Time of the day],2) as int) > 60
```

Time of the day:
```
set [Time of the Day] = null where cast(right([Time of the Day],2) as int)>60
```

Time of the day:
```
set [Time of the day] = null where [Time of the day] like '99%' or [Time of the day] like '%99'
```

Time of the day:
```
set [Time of the day] = null where len([Time of the day]) <3
```

Temperature:
```
set Temperature = Temperature/10 where filename = 'IMARESdataset' and temperature > 18
```

Temperature:
```
set Temperature = Temperature where filename = 'IMARESdataset' and temperature < 18
```

Temperature:
```
set Temperature = Temperature/10 where  filename <> 'IMARESdataset'
```

Ship:
```
set ship = null where  ship = '9999' or  ship = 'XXXX' or ship = 'unk'
```

Sampling method:
```
set [Sampling method] = null where [Sampling method] <> 'DEM' and [Sampling method] <> 'PEL'
```

Month:
```
set [month] = null where [month] = 0
```
Square:
```
set [square] = NULL where [square] = '9999'
```

Digestion Stage:
```
set DigestionStage = null where DigestionStage = 9
```

Day:
```
set [Day] = null where [Month] = 6 and [day] = 31
```

Predator upper length bound:
```
set [PredatorUpperLengthBound] = null where [PredatorUpperLengthBound] = '9999' or [PredatorUpperLengthBound] = '-1'
```

Predator lower length bound:
```
set [PredatorLowerLenghBound] = null where [PredatorLowerLenghBound] = '9999'
```

Station:
```
set [Station/Haul] = null where [Station/Haul] = '   999'
```

CPUE:
```
set [CPUE] = null where [CPUE] = '9999999'
```

Predator weight:
```
set [PredatorWeight] = null where [PredatorWeight] = 0
```

Predator age:
```
set PredatorAge = null where [PredatorAge] = -1
```

Prey lower length bound:
```
set [PreyLowerLengthBound] = null where [PreyPreyLowerLengthBound] = '-1' or [PreyPreyLowerLengthBound] = '9999'
```

Prey upper length bound:
```
set [PreyUpperLengthBound] = null where [PreyUpperLengthBound] = '-1'
```

Appendix B

Data Policy

In October 2005, the ICES Council adopted a new Data Policy having considered that

A) ICES is committed to openness for the scientific process and to free access to scientific data;

B) ICES recognises that proper data interpretation requires insight into the sampling design, compilation, and analysis;

C) To ensure proper interpretation of the data, data sources may define access procedures either as a general policy or in relation to specific datasets;

D) By making the data publicly available, data contributors and users continue to make ICES the focal point for data in the NE Atlantic area and serve the scientific community.

ICES Data Policy 2006 (http://www.ices.dk/marine-data/guidelines-and-policy/Pages/default.aspx) applies to data submitted from May 1, 2006.

For data submitted prior to May 1st, 2006, data sources will be contacted individually and may specify access restrictions in agreement with ICES. The former Data Policy of 1994 can be found http://www.ices.dk/marine-data/guidelines-and-policy/Pages/default.aspx.

ICES Data Policy 2006 conforms to the IOC Oceanographic Data Exchange Policy (http://www.iode.org/index.php?option=com_content&task=view&id=51&Itemid=95).

Appendix C

Changes to the IMARES DATA

1. stomach ids added to the IMARES data

2. changing ship names:

```
ship IN('9','21','9999','XXX','','2 3','unkn') --> ship='unk'
ship='DAN' --> ship='DAN2'
ship='WAL2' --> ship='WAH2' (consistency with DATRAS)
ship='AND' --> ship='AND2';

ship='unk' --> ship=station_haul for records where station_haul
IN('EXP','TRI','ISI','COM')
```

3. changing station_haul codes:

```
station_haul='' --> station_haul=ship for records where
original ship in ('9','21','9999','XXX','','2 3','unkn')

station_haul
IN('unk','unknwn','999','9999','999909','0','00','X','') -->
station_haul='unk'
```

4. adding filename to IMARES records:

```
filename='IMARESdataset'
```

5. changing missing variables:

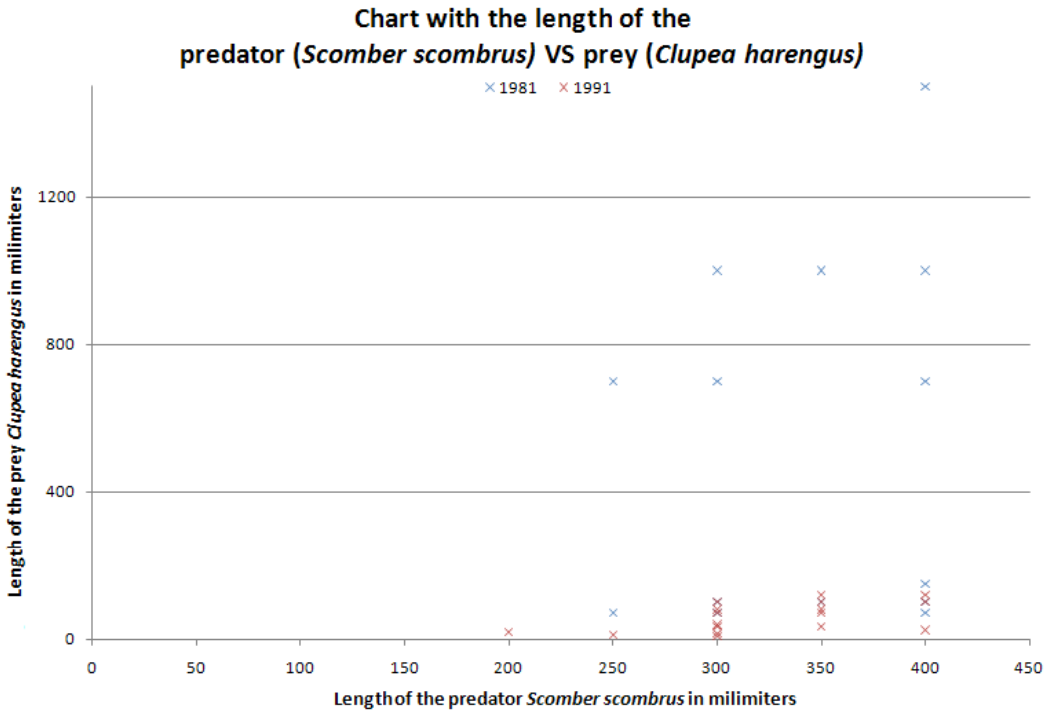
```
month IN(9999,0) --> month=.
day IN(99,9999) --> day=.
time IN(999,9999) --> time=.
depth=999 --> depth=.
temperature IN(99,9999,0) --> temperature=.
square='' --> square=quadrant for records where quadrant was
filled in
```

6. changing species variables:

```
IF nodc GT 10000000000 --> nodc=nodc/100 (IMARES data have a 12
number nodc, ICES has 10 numbers in nodc. This is because IMARES
someties incorporates subspecies, but that is something seldomly
happening)
scientific_name='Gadus morhua' --> nodc=8791030402 (IMARES
dataset did not have an nodc for Cod)
predator_code=. --> AND predator_code=pred for records where
predator_code was absend and pred was present
```

Appendix D

Problems found in the data



We have found records where the length of the prey is longer than the length of the predator. In some cases this can be normal, in another cases it is very unlikely. We dug a little bit more and by limiting to one species of prey and predator we can see the problem only happened in 1981 and it seems to be random.

There are 345 records where the length of the prey is longer than the length of the predator.

In the following table the summary of the cases where the prey is longer than the predator:

Predator	Prey	Prey code	Year	Number of records
Eutrigla gurnardus		6179220199	1990	1
Eutrigla gurnardus	Mysidae	6153010000	1990	1
Eutrigla gurnardus	Steiracrangon orientalislongicauda	6179220118	1990	1
Melanogrammus aeglefinus	Anaitides	5001130100	1991	2

Melanogrammus aeglefinus	Spionidae	5001430000	1991	5
Merlangius merlangus		6179220199	1987	1
Merlangius merlangus	Crangon	6179220100	1981	3
Scomber scombrus	Cheilostomata	7814000000	1981	1
Scomber scombrus	Clupea harengus	8747010201	1981	32
Scomber scombrus	Clupeidae	8747010000	1981	9
Scomber scombrus	Crangonidae	6179220000	1981	6
Scomber scombrus	Gadidae	8791030000	1981	1
Scomber scombrus	Limanda limanda	8857040904	1981	2
Scomber scombrus	Maurolicus muelleri	8759010501	1981	4
Scomber scombrus	Meganyctiphanes norvegica	6174020201	1981	6
Scomber scombrus	Nereididae	5001240000	1981	3
Scomber scombrus	Sprattus sprattus	8747011701	1981	30
Scomber scombrus	Trisopterus esmarkii	8791031703	1981	16
Eutrigla gurnardus	Pontophilus echinulatus	6179220602	1990	1
Melanogrammus aeglefinus	Goniada emerita	5001280207	1991	1
Melanogrammus aeglefinus	Nereis	5001240400	1991	1
Merlangius merlangus	Crangon allmanni	6179220119	1981	1
Merlangius merlangus	Hippoglossoides platessoides	8857040603	1981	1
Merlangius merlangus	Melanogrammus aeglefinus	8791031301	1981	1
Merlangius merlangus	Paguridae	6183060000	1991	1
Scomber scombrus	Alloteuthis subulata	5706010401	1981	22
Scomber scombrus	Ammodytidae	8845010000	1981	113
Scomber scombrus	Callionymus lyra	8846010106	1981	5
Scomber scombrus	Cephalopoda	5700000000	1981	1
Scomber scombrus	Gadus morhua	8791030402	1981	1
Scomber scombrus	Merlangius merlangus	8791031801	1981	1
Scomber scombrus	Nereis pelagica	5001240403	1981	1
Scomber scombrus	Phaeophyceae	1501000000	1981	1
Scomber scombrus	Polychaeta	5001000000	1981	2
Scomber scombrus	Syngnathus rostellatus	8820020119	1981	3
Scomber scombrus	Teleostei	8735000000	1981	37
Scomber scombrus	Trachinus vipera	8840060101	1981	1
Scomber scombrus	Trachurus trachurus	8835280103	1981	26