

**Development of a central database for European trawl survey data**

**DATRAS  
Database TRAwl Surveys**

**Final report**

Project number QLRT-2001-00025

## Participating Institutes



Animal Sciences Group, Wageningen UR  
Netherlands Institute for Fisheries Research (RIVO)  
Dept. Fisheries and Marine Ecology  
P.O. Box 68, Haringkade 1, 1970 AB IJmuiden,  
The Netherlands



International Council for the Exploration of the Sea (ICES)  
Palaegade 2-4  
1261 Copenhagen K  
Denmark



Fisheries Research Services (FRS)  
Marine Laboratory  
PO Box 101 Victoria Road Aberdeen AB11 9DB  
United Kingdom



Institut Française de Recherche pour l'Exploitation de la Mer (IFREMER)  
8, rue François Toullec  
F-56100 Lorient  
France

Participants (in alphabetical order)

Daan, N.	RIVO	
de Boois, I.	RIVO	
Emslie, D.C.	FRS	
Fletcher, N.	ICES	
Heessen, H.	RIVO	
Larsen, L.I.	ICES	
Lassen, H.	ICES	
Mahe, J.C.	IFREMER	
Newton, A.W.	FRS	
Patberg, W.	RIVO	
Peach, K.J.	FRS	
Piet, G.J.	RIVO	Coordinator
Sparholt, H.	ICES	
Star, B.	RIVO	
Storbeck, F.,	RIVO	

# Contents

<b>1</b>	<b>INTRODUCTION</b> .....	<b>7</b>
<b>2</b>	<b>SURVEYS</b> .....	<b>7</b>
2.1	IBTS .....	7
2.1.1	<i>IBTS North Sea</i> .....	7
2.1.2	<i>IBTS Western and Southern areas</i> .....	8
2.2	BITS .....	14
2.3	BEAM TRAWL SURVEYS .....	15
<b>3</b>	<b>DATA</b> .....	<b>16</b>
3.1	DATA COLLECTION PROCEDURES .....	16
3.2	DESCRIPTION OF THE DATA .....	17
3.2.1	<i>Existing databases</i> .....	17
3.2.2	<i>Exchange format</i> .....	17
3.3	AVAILABILITY OF DATA WITHIN DATRAS .....	17
3.4	INDICES .....	18
3.4.1	<i>Beam trawl survey</i> .....	19
3.4.2	<i>BITS Indices</i> .....	20
3.4.3	<i>IBTS indices</i> .....	20
3.5	EXPERT ADVICE .....	21
<b>4</b>	<b>THE DATRAS SYSTEM</b> .....	<b>24</b>
4.1	SOFTWARE .....	25
4.2	DATABASE STRUCTURE .....	25
4.2.1	<i>Reference code system (RECO)</i> .....	26
4.3	SPECIES CLASSIFICATION AND CODING (ITIS) .....	26
4.4	DATA SCREENING UTILITY (DATSU) .....	27
4.4.1	<i>Checks</i> .....	29
4.4.2	<i>DATSU Database</i> .....	29
4.4.3	<i>Internal front-end</i> .....	31
4.4.4	<i>Web Interface and running the program</i> .....	32
4.5	DATRAS DATABASE .....	34
4.5.1	<i>Structure DATRAS database</i> .....	34
4.5.2	<i>Aggregation of data and calculations of indices</i> .....	36
4.5.3	<i>Data warehouse</i> .....	40
<b>5</b>	<b>ACCESS POLICY</b> .....	<b>44</b>
<b>6</b>	<b>OUTPUT</b> .....	<b>45</b>
6.1	PUBLIC DATA .....	45
6.2	AGGREGATED DATA .....	47
6.3	RAW DATA .....	48
6.4	'HEALTH' WARNING .....	48
<b>7</b>	<b>REFERENCES</b> .....	<b>49</b>

<b>APPENDIX I: EXCHANGE FORMAT .....</b>	<b>50</b>
<b>APPENDIX II: RECO DATABASE DIAGRAM.....</b>	<b>52</b>
<b>APPENDIX III: DATSU DATABASE DIAGRAM .....</b>	<b>54</b>
<b>APPENDIKS IV: DATRAS DATABASE DIAGRAM .....</b>	<b>55</b>
<b>APPENDIX V: DATA IMPORT .....</b>	<b>56</b>
<b>APPENDIX VI: THE IBTS ALK SUBSTITUTION PROCEDURE.....</b>	<b>68</b>
<b>APPENDIX VII: DATA WAREHOUSE DATABASE DIAGRAM.....</b>	<b>69</b>

## Summary

As part of the assessment of the status of the fish stocks in the North Sea, Baltic Sea and western European waters (i.e. Celtic Sea, Bay of Biscay and Eastern Atlantic from Shetlands to Gibraltar), fisheries institutes of the bordering countries have for many years carried out research vessel surveys. Besides information on commercial species, also information on all other fish species in the catch was collected on a routine basis. At present three major international trawl surveys can be distinguished: the International Bottom Trawl Survey (IBTS), the Baltic International Trawl Surveys (BITS) and the Beam Trawl Survey (BTS). Although initially these surveys were coordinated at the national level, over the years a number of these trawl surveys have evaluated into standardised international research programmes, most often coordinated by the International Council for the Exploration of the Sea (ICES).

The aim of this project is to establish an international database of trawl survey data at the ICES HQ including:

- IBTS North Sea, Skagerrak, Kattegat
- IBTS western and southern divisions
- BTS North Sea, Channel and Irish Sea
- BITS Baltic Sea

One of the benefits of such a central database is that both input and quality assurance of all survey data can be standardized, thereby securing the same level of quality control not just among participating institutes within a survey but also between surveys. To that end software was developed that requires data quality checks before data can be loaded into the database. This imposes a certain level of quality assurance on all institutes delivering data to that central database. Another benefit is that the access to survey data can be greatly improved through web facilities that allow access through the same database front-end. The easy access to survey data was also a great cause for concern among the institutes participating in the surveys and required the development of a formal protocol that regulates the level of data access depending on the status of the requesting party. The general public has access to the information in the database through a suite of standard tables and figures that are annually updated, some parties may have access to aggregated data while only few may have access to the raw, un-aggregated data.

At this point DATRAS, the database of bottom trawl survey data, contains most of the data from the four surveys. All these data have passed through a checking program and errors have been corrected. Both the uploading of new data and the downloading of existing data can be done through the web. Background information on each of these surveys as well as their most recent manual can also be found on the DATRAS website (<http://www.ices.dk/datacentre/datras/datras.asp>). Although most of the background as well as the work done to develop and create DATRAS is described in this report, the reader is encouraged to visit the website for a better appreciation of its contents.

# 1 Introduction

As part of the assessment of the status of the fish stocks in the North Sea, Baltic Sea and western European waters (i.e. Celtic Sea, Bay of Biscay and Eastern Atlantic from Shetlands to Gibraltar), fisheries institutes of the bordering countries have for many years carried out research vessel surveys to estimate the recruitment of the youngest age groups of a number of commercial fish species. Besides information on commercial species, also information on the length compositions of all other fish species in the catch was collected on a routine basis. Therefore the data collected during these surveys are not only relevant for the management of commercially important fish stocks but also provide a valuable source of information for studies into the spatial and temporal variation in the fish community, biodiversity or the status of the ecosystem.

Although initially these surveys were coordinated at the national level, over the years a number of these trawl surveys have evaluated into standardised international research programmes, most often coordinated by the International Council for the Exploration of the Sea (ICES). At present four major international trawl surveys can be distinguished: the International Bottom Trawl Survey (IBTS) in the North Sea, IBTS western and southern divisions, the Baltic International Trawl Surveys (BITS) and the Beam Trawl Survey (BTS). Some of these surveys already existed in separate databases at ICES (international Council for the Exploration of the Seas) others existed only in various national databases.

The aim of this project is to establish at ICES an international database of trawl survey data that can be accessed through the web. Web facilities should include standardised input and quality assurance as well as the possibility of standardised data extractions depending on user-specific demands.

## 2 Surveys

### 2.1 IBTS

The IBTS consists of a number of national surveys that aim at improving standardisation and collaboration between surveys. Two areas can be distinguished that differ in terms of the length of time spent on this trajectory and hence the degree to which standardisation was achieved: IBTS North Sea and IBTS Western and Southern areas. Although there are marked similarities between these two areas and both are represented in the ICES IBTS working group it was established at the 1999 meeting of this working group that due to the considerable difficulties in merging the protocols used in the North Sea with those used in the western and southern divisions, two separate manuals should be the standard.

#### 2.1.1 IBTS North Sea

In the North Sea the IBTS started in the 1960's as a survey that was directed at juvenile herring and was at that time called the International Young Herring Survey (IYHS). As it was gradually realised that the survey also yielded valuable information for other fish species, such as cod and haddock, the objectives were broadened and the survey was renamed into the International Young Fish Survey (IYFS). Besides the IYFS, which was carried out in the first quarter, a number of national surveys developed in the 1970's and 1980's that were mainly carried out in the third quarter. In 1990 ICES decided to combine the international and the national surveys

into the IBTS. The IBTS is carried out twice per year (1<sup>st</sup> and 3<sup>rd</sup> quarter) since 1997 and on a quarterly basis in the period 1991-1996.

Prior to 1977 there was no standardisation of gear although all ships used bottom trawls with a small mesh cover. In 1977 ICES recommended that all ships should use a GOV trawl as specified by the Institut des Peches Maritimes, Boulogne. A detailed description of the net is to be found in the manual (Anon. 2004). The GOV trawl has been gradually phased in, e.g. in 1979 only 3 vessels were equipped with the GOV trawl, but by 1983 all 8 nations were using this gear. It should be noted that although the gear is now standard, variations in the rigging exist between the various countries. The fishing method is also standardized and described in the manual (Anon. 2004). Fishing speed is 4 knots measured as trawl speed over the ground. In 1977 ICES also recommended that the duration of a tow should be reduced from an hour to half an hour with the catch data to be expressed in numbers per hour. All nations accepted this recommendation although it was a number of years before 30 minutes became the standard.

## 2.1.2 IBTS Western and Southern areas

In 1994, it was suggested to extend the remit of the IBTS working group to co-ordinate the surveys in the western and southern areas (i.e. Channel, Celtic Sea, Bay of Biscay, eastern Atlantic waters from the Shetlands to the strait of Gibraltar) and contacts were established with the national laboratories involved to co-ordinate and standardise these surveys. The following national bottom trawl surveys were identified in the area:

Table 2.1 National bottom trawl surveys and the quarters and areas covered

Country	Quarter	Survey
	1	Division VIa
Scotland	3	Rockall Survey (SGF6b) (every second year)
	4	Scottish Mackerel Recruit Survey
Northern Ireland	1	Irish Sea (Division VIIa)
	4	Irish Sea (Division VIIa)
Ireland	4	West coast Groundfish Survey
	4	Irish Sea-Celtic Sea Groundfish Surveys
England	1	Celtic Sea and Western Approaches Groundfish Survey
France	4	Eastern Channel (Division VIId)
	4	Celtic Sea and Bay of Biscay (Divisions VIIIf,g,h,j;VIIIa,b)
	3	Porcupine bank (Divisions VIIb,k)
Spain	4	Cantabrian Sea and Off Galicia (Divisions VIIIf,c and Northern part of IXa)
	2	Gulf of Cadiz (Southern part of division IXa)
	4	Gulf of Cadiz (Southern part of division IXa)
Portugal	3	Bottom trawl Survey (Portuguese shelf - Division IXa)
	4	Bottom trawl Survey (Portuguese shelf - Division IXa)

International coordination of surveys in this region began in 1997 and was based on a previous EU project – SESITS (South-western European Shelf International Trawl Surveys). The surveys covered in this project concentrate on the third quarter and involve France, Spain and Portugal. Each country conducts surveys in adjacent areas with no overlapping. While some attempts have been made in order to achieve a consensus on the choice of a standard gear, this was not



achieved due to the variation in bottom types, and each country uses a different gear (GOV for France, BAKA for Spain and Norwegian Campelen Trawl for Portugal). With the sampling protocols, however, a significant level of standardisation was achieved, and all three countries are using depth-stratification in their sampling strategy. Building on the success of this project, joint projects with England, Ireland and Scotland were initiated. As a consequence the level of standardisation with these countries is lower; e.g. some countries use depth strata, others ICES rectangles.

### **Scottish Surveys**

The Quarter 1 Scottish Groundfish survey started in 1981 and was initially intended to cover the fishing grounds on the continental shelf to the west of Scotland; in 1996 the survey area was extended to include the northern Irish Sea. The Scottish West Coast Surveys use a similar ICES rectangle based sampling strategy to that used in the North Sea. Trawl stations are selected at one tow per rectangle based on a library of clear tows. There is no explicit return to the same trawling position every year, although this is generally the case. Since 1999 the potential for using a depth rather than rectangle based stratification has been under investigation. To this end, and where possible, those rectangles which display substantial internal depth variation have been sampled twice at different depths. The recent inclusion of samples collected between 200 and 500m would suggest that depth stratification should be initiated as soon as possible.

The survey covers Division VIa and extends into the northern part of the Irish Sea and NW of Ireland. The depth range covered has been 20 to 500m since 2000. The survey is usually carried out in March of each year. The target species are cod, haddock, whiting, saithe and herring and age frequencies are constructed for these species. All other fish species encountered are also sampled for at least length frequencies. Indices of abundance at age are calculated for all the target species and these data are used at the Northern Shelf Assessment Working Group and also made available for the Herring Assessment Working Group.

The Scottish Quarter 3 Rockall survey began in 1985 and was carried out annually until 1997. However, in 1998 it was decided to make the survey a bi-annual event; in 1998 a new survey of deep water stocks was completed, in the future the Rockall survey and a deep water survey will occur in alternate years. The Rockall surveys will generally be in September during odd numbered years. The survey covers only a relatively small area, in the order of eight ICES rectangles. Trawl stations are on known clear tows and vary between 2 and 8 per rectangle depending on the proportion of the area within the 250m depth contour.

The Scottish Mackerel recruit Quarter 4 survey began in 1990 and has a depth range of 20 – 500m. The survey extended to the area west of the British Isles between 56B and 61B N and bounded by the 200m depth-contour and the coast. Initially the survey area did not include the area of the Minch and the north channel of the Irish Sea but gradually the spatial coverage has been altered until now it mimics the Quarter 1 survey. In 1998 the new research vessel *Scotia III* was used and the duration of the hauls was decreased from 60 minutes to 30 minutes.

Up until 1995 the target species for this survey was mackerel but the Mackerel Assessment Working Group detected a discrepancy between the survey index and the VPA derived recruitment index. This led to a withdrawal of the survey index from the assessment. Given this situation the whole survey was re-designed to follow more closely the demersal quarter 1 survey. The mackerel survey now ends in the region of the northern part of Donegal Bay and also extends into the northern Irish Sea.

The target species have now been extended to include cod, haddock, whiting, saithe and herring as well as the original target species mackerel.

The gear deployed on all the Scottish surveys is the 36/47 GOV trawl fitted with heavy ground gear 'C' and a 20 mm internal liner. The vessel undertaking this survey changed to *Scotia III* in

March 1998 from the previous *Scotia*. The gear includes a full suite of Scanmar sensors; headline height, wing and door spread and speed through the water.

### **Northern Ireland surveys**

In the Northern Ireland surveys, the sampling design is stratified with fixed-position stations. Stratification is by depth and seabed type. Haul duration is 3 nautical miles at 3 knots over the seabed. Stations in the St George's Channel are 1 nautical mile at 3 knots and have only been surveyed since October 2001. Number of stations is 45 in northern Irish Sea and 12 in St George's Channel. Tows are during day time only.

The surveys have been carried out in March and October since 1992. From March 1992 to March 2001, the survey extended from 54° 43' N to 53° 22' N. From October 2001 the survey is extended into the St George's Channel to 52° 18' N. Depth range is from 20 to 120m. The surveys are carried out on the R.V. *Lough Foyle*, a 43.5m stern trawler of 880 kw and GRT 547 tonnes. The fishing gear is a rock-hopper otter trawl with a 17m footrope fitted with 250 mm non-rotating rubber discs. The gear has a mean vertical opening of 3 m. The door spread varies from around 25m at 20 m depth to 40m at 80m depth. A 20mm (inside mesh) codend is fitted.

### **Irish surveys**

The Irish West Coast Groundfish Survey started in 1990 and for the first two years consisted of circa 25 stations concentrated around the Irish coast in ICES Areas VIa South and VIb. Adverse weather in 1992 limited station coverage to only 4 stations which effectively broke the time series. The survey was re-established in 1993 and has consisted of circa 70 stations, for Parts A and B combined, since then. Spatial coverage was extended west out to the 200m contour, but remains as VIa South and VIb.

Due to the restrictions of the current and previous research vessel commercial trawlers have been contracted to carry out the survey work. Wherever possible continuity of vessel and gear has been maintained and standard IBTS methodology applied. However, due to the reduced staffing possible on commercial vessels it has rarely been possible to completely sort the catch. Until 2000 all cod were sorted from the catch and then a sub-sample of two baskets was taken and completely sorted. From 2001 onwards, in response to an overall review of survey sampling undertaken during the IPROSTS Project (Anon 2001), the catch is completely sorted for all target species and a qualitative assessment made of the residual catch as a minimum.

The Irish surveys use an ICES rectangle based sampling strategy. The sampling design attempts to allocate at least two stations per rectangle (where the sea area is appropriate). Stations are selected randomly within each rectangle from known clear tow positions. On the WCGS and ISCGS circa 70 fishing stations are planned on each survey every year. This number of hauls is adjusted according to the ship time available at sea.

The Irish west coast groundfish survey is carried out in two parts: Part A covers ICES Division VIa (south) and VIIb (north); Part B covers ICES Division VIIb and VIIj. The survey is conducted from 15 to 300 m depths during the fourth quarter (October - November).

The Irish west coast groundfish survey is carried out on chartered commercial fishing vessels. Whilst the same vessel (MFV *Marliona*, 224 gross tonnage and 30 m LOA) has normally been used each year for Part A and Sionann for Part B, in 2001 Part B was conducted from the MFV *Regina Ponti* (34.5 m LOA). Both vessels use a Rockhopper net with 12 inch discs and 11 inch Thyboron doors. The nets are fitted with a 20-mm codend liner. Gear performance throughout the survey is monitored using Furuno Ch24 (Headline Monitor).

The Irish Sea Celtic Ground Survey commenced in 1997 and evolved from an earlier Irish Sea Juvenile Fish Survey. As a consequence early survey stations concentrated largely, though not exclusively, around a number of shallow spawning areas along the Irish east coast in VIIa. These

positions were expanded in combination with clear tows provided by the industry and CEFAS as well as some exploratory tows.

Spatial coverage therefore extended into the western Irish Sea from 2001 into the area around the Isle of Man, Liverpool and Cardigan Bays and the Welsh coast. The survey is carried out on the Irish research vessel the R.V. Celtic Voyager. The sampling procedure on board conforms to the IBTS standard protocols and as such all cod are sampled and aged, the entire catch is sorted and then sub-sampled as and where appropriate.

The Irish survey in the Irish Sea and Celtic Sea covers ICES Division VIIa and VIIg. The survey is conducted from 10 to 150 m depths during the fourth quarter (November-December) by the RV *Celtic Voyager*. This vessel is 32 m in length with gross tonnage of 340 t. The fishing gear used is a GOV 28.9/37.1 Trawl with Morgere Kite (0.85 by 0.85m). Mean vertical opening is 6 m and door spread 48 m. Morgere Polyvalent doors (Type AA4.5) are used and gear performance is monitored throughout the survey using the SCANMAR (RX400) net monitoring system (Headline height, Door spread).

Trials on the new 65m research vessel, the R.V. Celtic Explorer, are due to commence in late 2002. Therefore, from 2003 onwards all Irish Surveys will be conducted on this new vessel, starting around mid October through to late November.

### **English Survey**

During the late 1970's the Western mackerel stock fishery was expanding and concern for over-exploitation increasing. The Celtic Sea and Western Approaches Groundfish Survey was started in 1981, with the aim of investigating the distribution, biology and pre-recruit abundance of this mackerel stock. These objectives were almost immediately extended to all species that could be adequately sampled with a bottom trawl.

While mackerel was the primary target the survey covered all or part of the western continental shelf from the northern North Sea to the north coast of Spain. Later, as the objectives changed, the area shrunk in stages to its present boundaries: 47° 30' N to 52° 30' N and 3° W to 12° W. This has been the standard area since 1987. In the early years a March/April and December survey was carried out each year but since 1989 only the spring (quarter 1) survey has been conducted.

The survey fishes fixed station positions allocated by area (division lines at 48° 45'N and 50° 15'N) and depth strata (40-89, 90-114, 115-139, 140-179, 180-299 and 300-600m) and is carried out on the RV *Cirolana*, a stern trawler 74m in length with a gross tonnage of 1731 t. The fishing gear used is a modified Portuguese High-Headline trawl (PHHT) with 350mm rubber bobbins, a bunt tickler chain and a 20mm codend liner. Since 2004, this survey has been conducted from RV CEFAS Endeavour.

A fourth quarter survey is currently being established, as part of the internationally coordinated survey for southern and western waters. The first survey was conducted in 2002 on RV *Cirolana*, and used a Baca trawl. Due to major gear damage, the survey switched to a PHHT for the remainder of that year. During 2003, a modified GOV was used with rockhopper groundgear, deployed from RV CEFAS Endeavour.

### **French surveys**

The French demersal survey began in 1987. The survey area was first limited to the Bay of Biscay (ICES divisions VIIh, VIIa,b,c and d) and in 1990, the survey area was extended towards the north to cover the grounds of Celtic sea deeper than 100 meters (ICES divisions VIIe,f,g,h and j).

For the 1987 to 1996 period, the survey was conducted in the Bay of Biscay on an annual basis with the exception of the years 1993 and 1996. Most years it was conducted in the third or fourth

quarter (October-November) except in 1991 when it took place in May. In 1988 two surveys were conducted, one in May the other in October.

The Celtic Sea was surveyed from 1990 to 1994 but the sampling was restricted to a small geographical area. The duration is between 40 to 45 days depending on year and availability of ship. Since 1997, the survey covered all the Celtic Sea and Bay of Biscay during the 4th quarter.

Prior to 1997, the sampling designs were as follows: In the Bay of Biscay (ICES divisions VIIh, VIIIa,b,c and d) a stratified sampling scheme was originally used.

The area was divided according to latitude into 3 blocks and the hauls were distributed in seven depth zones (15-30,31-80, 81-120, 121-160, 161-200, 201-400, 401-600 m). 100 hauls were made at fixed locations and 35 at changeable stations from year to year. Since 1989, all the hauls (mean number by survey 135) are made at the same locations. In the Celtic Sea (ICES divisions VIIe,f,g,h and j) the sampling design was systematic, stations were located at the intersection points of a grid of lines 25 nautical miles apart both in latitude and in longitude. The mean number of sets was 56.

From 1997 onwards the whole area has been separated in 5 geographical strata or sectors: southern Bay of Biscay (GS) and northern Bay of Biscay (GN), southern Celtic Sea (CS), central Celtic sea (CC) and northern Celtic sea (CN). In each sector a depth-stratified sampling strategy has been adopted with 7 depth ranges: 0 - 30m, 31 - 80 m, 81-120 m, 121 - 160 m, 161 - 200 m, 201 - 400 m and 401 – 600m.

The sampling design is a stratified random allocation. The number of hauls per stratum is optimised by a Neyman allocation taking into account the most important commercial species in the area (hake, monkfish and megrim). Minimum of two stations per stratum is sampled and 140 fishing stations are planned every year. This number of hauls is adjusted according to the ship time available at sea.

Catch weight and catch numbers were recorded for all species, only selected finfish and shellfish species were measured until 1990. Since 1991, all finfish and a selection of shellfish (mainly nephrops and squids) are measured.

Since 1997, the French survey has been carried out on the R/V *Thalassa*, a stern trawler of 73.7 m length by 14.9 m wide, gross tonnage of 3022 t. The fishing gear used is a GOV 36/47 without exocet Kite which is replaced by 6 additional floats. In average, the gear has a horizontal opening of 20 m and a vertical opening of 4 m. The doors are plane-oval with 1350 Kg.

### **Spanish surveys**

The Spanish survey in the Porcupine bank began in 2001 and covers ICES Division VIIIb,k corresponding to the Porcupine Bank and adjacent area in western Irish waters from longitude 12° W to 15° W and from latitude 51° N to 54° N. The survey takes place in the third quarter (August- September) and covers depths between 180 and 800 m.

The survey area has been separated in two geographical strata and three depth strata (less than 200 m, 200-400, 400-800 m), resulting in 5 strata, given that there are no grounds shallower than 200 m in Outer geographical strata. The sampling design is random stratified with proportional allocation and a minimum of two stations per stratum with a total of 80 fishing stations.

The survey is carried on the R/V *Vizconde de Eza*. This vessel is a stern trawler of 53 m length and 13.5 m wide with gross tonnage of 1400 t. Fishing gear used is a Porcupine boca 60/72 with 59.46 m footrope and a 71.96 headline. Doors are oval with 800 kg and 4.5 m<sup>2</sup> surface. Diameter of warp used is 20 mm, of sweeps is 55 mm and the groundrope 98 mm with a double synthetic coat. Mean vertical opening is 3.5 m and door spread 120 m. Codend mesh size is 20 mm.

Since 1974 the IEO has performed bottom trawl surveys in the Atlantic continental shelf waters of the Iberian Peninsula (SESITS 1999). From 1980 the fishing resources of Divisions VIIIc and IXa of ICES were monitored through surveys, with the objective of following variations in the abundance of demersal and benthic species of commercial interest by means of indices

independent of fishing activity. At the same time estimations were obtained of the strength of recruitment of diverse species (principally hake) during the autumn. The evaluations were made according to a stratified sampling protocol, maintaining other factors constant, such as time of year, ship, fishing gear, speed, trawl time, etc. Tows were of one hour duration in all surveys before 1984, and were reduced to 30 minutes thereafter. Since 1990, gear geometry is monitored using Scanmar equipment.

Two series of surveys have been conducted, one at Spring (April-May), starting in 1984, and the other in the Autumn (September-October) starting in 1980. The spring series ended in 1988 and the autumn one has continued up to the present.

Two Spanish groundfish surveys are distinguished: The northern Spanish groundfish survey covers ICES Division VIIIc and the northern part of IXa corresponding to the Cantabrian Sea and off Galicia waters. This survey is conducted during the third and the fourth quarter (September – October) and covers a depth range of 35 to 700 m. Stratification is based on three depth strata (80-120, 121-200, 201-500 m) and 5 geographic sectors. Additional hauls both in deeper water (500 - 700 m) and shallower waters (30 – 80 m) may be conducted depending on the ship time available at sea. The coverage is approximately 5.4 hauls for every 1000 Km<sup>2</sup> (120 hauls per survey).

The southern Spanish groundfish survey is conducted during late Winter and Autumn in the southern part of ICES Division IXa, the Gulf of Cádiz. The covered area extends from 15 m to 700 m depth. In the southern surveys five depth strata have been used (15-30, 31-100, 101-200, 201-500 and 501-700 m).

All Spanish surveys in Spanish waters were carried out with R/V *Cornide de Saavedra* except in 1989 when another research vessel (N/V F. de P. Navarro) was used to conduct the survey. Several modifications were applied over time to this stern trawler: the engine power increased in 1983 (from 1700 Kw to 2651 Kw), in 1984 it was modified from its original 56 m (LL) and 990 GRT to 67 m and 1133 GRT at present, and a new bridge was used in 1990 (GPS, colour Echosounder, Plotter, Doppler log, etc.). The gear used is a Baka trawl 44/60 with a 43.6 m footrope and a 60.1 headline. The traditional trawl doors used are rectangular, weighting 650 Kg and 3.6 m<sup>2</sup> of surface (2.67\*1.34 m). The diameter of warp used is 22 mm (1.9 Kg/m). The mean vertical opening is 1.8 m and the horizontal opening is 21 m. Up to 1985, a codend cover of 20 mm mesh was used, and since then, a 20 mm mesh codend liner has been adopted.

### **Portuguese surveys**

The Portuguese groundfish surveys have been conducted twice a year (in Summer and Autumn), with R/V “Noruega”. Initially the main objectives of the surveys were to estimate the abundance and study the distribution of the most important commercial species in the Portuguese trawl fishery: hake, horse mackerel, blue whiting, seabream and Norway lobster. Recruitment indices of abundance and distribution for hake and horse mackerel were also evaluated in the autumn surveys. Additionally, trawl selectivity experiments for hake and horse mackerel with 40 mm mesh size, were also conducted during 1981 surveys using the covered cod-end method.

A stratified random sampling design was adopted during 1979-1989. The number of strata changed during this period: from 1979 to 1980 the surveyed area was divided into 15 strata and from 1981 onwards into 36 strata. Based on the statistical analysis of the previous surveys the design was revised in order to decrease the variance within stratum. The new strata are smaller than the previous ones and can be combined to get the previous ones. The aim of increasing the number of strata was to increase the probability of spreading the random sampled units in order to decrease the total variance of the species' mean abundance indices.

The stratification is based on depth and geographical areas. The depth ranges used during 1979-1988 were 20-100m, 101-200m and 201-500m. Each stratum was divided into units of approximately 25 nm<sup>2</sup>, sequentially numbered. During 1979-1980 the number of random hauls

per stratum was based on the previous information of the relative abundance of the target species in each geographical area and on the ship time available. During 1981-1989, when the number of strata was 36, two random units were sampled by stratum whenever possible, to achieve an estimate of the standard error of the stratified mean by stratum. The tow duration was 60 minutes during 1979-1985 at a trawling speed of 3.5 knots, changing to 30 minutes during 1986-1988, and changed back again to 60 minutes in 1989 as it was observed that the large adults of horse mackerel were not caught in 30 minutes tows at this trawling speed.

In the Portuguese surveys the sampling design is based on fixed stations. A total of 97 fixed stations are planned, spread over 12 sectors. Each sector is subdivided into 4 depth ranges: 20-100m, 101-200m, 201-500m and 501-750 m, with a total of 48 strata. The positions of the 97 fixed stations were selected based on common stations made during 1981-1989 surveys and taking into account that at least two stations per stratum should be sampled. A maximum of 30 supplementary stations are planned, fixed in each season, to be carried out if ship time is available or to replace positions that due to particular factors are not possible to sample.

The Portuguese surveys cover Division IXa in Portuguese waters. The area surveyed extends from latitude 41°20' N to 36°30' N, and from 20 to 750 meters depth.

The surveys are carried with the R/V Noruega, which is a stern trawler of 47.5 m length, 1500 horse power and 495 G.T.R. The fishing gear used is a bottom trawl (type Norwegian Campell Trawl 1800/96 NCT) with a 20 mm codend mesh size. The main characteristic of this gear is the groundrope with bobbins. The mean vertical opening is 4.6 m and the mean horizontal opening between wings and doors is 15.1 m and 45.7 m, respectively. The polyvalent trawl doors used are rectangular (2,7 m x 1,58 m) with an area of 3,75 m<sup>2</sup> and weighting 650 Kg.

## **2.2 BITS**

The Baltic cod stock has been monitored annually since 1982 through bottom trawl surveys carried out by most countries surrounding the Baltic. The national research vessels have each surveyed part of the area with some overlap in coverage and applied a depth stratified sampling design. However, different gears and design were applied and in 1985 ICES established a Study Group on Young Fish Surveys in the Baltic in order to standardize the surveys. After agreement a common standard trawl gear and standard sampling procedures were implemented in 2000 resulting in the coverage of the whole Baltic Sea. To calibrate the national surveys from before 2000 with the TV3 gear used from 2000, a set of conversion factors are produced by making comparative hauls. The work with standardising gear and creating conversion factors for old data was done under the EU project ISDBITS and gear specifications and conversion factors can be found in the report (ISDBITS 2001).

The type of trawl is called TV3 and come in two sizes for different sizes of research vessels, one 520 meshes in circumference and one 930 meshes. The small trawl is used for vessels up to around 800 HP and the larger trawl for vessels with higher engine power. The design and construction of the standard trawls are given in ICES (1997) and can also be found in the BITS manual (Anon. 2000). Denmark is still using the rock hopper gear on hard fishing ground as the small TV3 trawl can't fish in these areas. These 2 gears are assumed to have a similar catchability.

The BITS is conducted as a depth-stratified survey. The strata are based on Sub divisions and depth layers. Each year the necessary stations are randomly selected before the beginning of the international trawl surveys from a list of clear haul data. These stations are a selected sub-sample of the possible trawl tracks. If the number of possible tracks is not large enough for a random selection in some strata fixed stations can be used every year. The standard haul is a 30 minute haul with a towing speed of 3 knots. The speed should be measured as the speed over the ground.

Start time is defined as the moment when the vertical net opening is stable at the stated towing speed. Stop is defined as the start of hauling back the trawl. Trawling shall only take place during daylight, defined as the time between 15 minutes before sunrise until 15 minutes past sunset.

## 2.3 Beam trawl Surveys

During the 1980's, five countries bordering the North Sea and western waters of the UK had developed a range of beam trawl surveys. Some of these surveys were designed to sample pre-recruit (0- and 1- group) plaice and sole on nursery grounds with light gears, while others used beam trawls of commercial design to catch juveniles and adults. Collation and analysis of some of these data was initially focussed on the North Sea and eastern Channel through the Study group on beam trawl surveys, but at the ICES Statutory meeting in 1990, it was resolved to broaden the remit of this Study Group and to summarise and evaluate six of these surveys covering North Sea, Channel and Irish Sea. Initial efforts of the Study Group were focussed on standardising the methods of sampling and of data collection but despite these efforts a number of different gears and survey designs are used, depending on the different aims of the surveys, the vessels used and the type of substrate. Sampling methods and data collection and storage, however, have been standardised to a considerable extent.

The offshore beam trawl surveys for which a certain level of standardisation is achieved through the ICES beam trawl WG are shown in table 2.2 (see ICES 2003).

Table 2.2 National offshore beam trawl surveys and the quarters and areas they cover

Country	Quarter	Area
The Netherlands	3	North Sea
England	3	Channel and Irish Sea
Germany	3	German Bight
Belgium	3	southern North Sea

### The Netherlands

The Netherlands BTS was initiated in 1985 to estimate the abundance of the dominant age groups of plaice and sole including pre-recruits. Initially the survey was only carried out in the south-eastern North Sea (ICES area IV) using RV "Isis" equipped with a pair of 8 m beam trawls rigged with nets of 120 mm and 80 mm stretched mesh in the body and 40 mm stretched mesh cod-ends. A total of 8 tickler chains are used, 4 mounted between the shoes and 4 from the ground-rope. The survey was designed to take between one and three hauls per ICES rectangle (boxes of 0.5° latitude by 1° longitude). The stations are allocated over the fishable area of the rectangle on a "pseudo-random" basis to ensure that there is a reasonable spread within each rectangle. No attempt is made to return to the same tow positions each year. Towing speed is 4 knots for a tow duration of 30 minutes and fishing occurs during daylight only.

In 1995, the survey was expanded into the central and northern part of the North Sea using RV "Tridens". This vessel uses the same gear but is equipped with a flip-up rope as it covers rougher grounds. Sampling strategy is also similar but only one haul per rectangle is taken, preferably close to the centre of the rectangle.

### England

An English beam trawl survey has been carried out annually in July / August since 1989 using a commercial 4m beam trawl. The primary aim was to assess the relative abundance of pre-recruit

plaice and sole in ICES Division VIId. Consequently, most of the sampling was carried out in areas known to be nursery grounds for these species. In 1995 the survey was extended to include the southern North Sea in order to sample the whole population of plaice and sole. The standard survey had a total of 107 tows of 30 minutes duration, of which 79 were in the eastern Channel. Since 1999 the number of tows worked has been reduced to 91 of which 75 are in the Eastern Channel. The sampling gear consists of commercially rigged (1989 style) 4m beam trawl (measured between inside edges of shoes) fitted with a chain mat, flip-up ropes, and a 40mm cod-end liner. The gear is towed at 4 knots (ground speed) for 30 minutes on a warp length appropriate to the depth of water. Fishing is only carried out in daylight on this survey, as catch rates can be very different in the dark under certain conditions.

### **Germany**

The German beam trawl survey started in 1991 and covers much of the South-eastern North Sea. It takes about 60 hauls each year using a pair of 7m beam trawls rigged with tickler chains. The cod-end mesh size is 75 mm.

### **Belgium**

The Belgium survey started in 1985 and covers much of the South-western North Sea. It takes about 60 hauls each year using a single 8m beam trawl based on the design of the standard Netherlands survey but fitted with “flip-up” ropes.

## **3 Data**

Based on the characteristics of the surveys (see chapter 2) and how they are organised in terms of working groups (WGs) or the level of standardisation between surveys we distinguished four survey/area combinations. Each of these survey/area combinations are conducted by several countries and may even consist of different surveys. However, the survey/area combinations are chosen such that similarity within them allows the data from one survey/area combination to be used together for analyses. Guidelines are provided aimed at decreasing the erroneous use of the data available through DATRAS.

### **3.1 Data collection procedures**

The data collection procedures of the various surveys were compared by going through the existing manuals and by discussing these at the various survey working groups (BITSWG, IBTSWG and WGBEAM). As each of these surveys already has a long history of improving and standardizing their data collection procedures within the survey/area combination it is not within the remit of DATRAS to suggest any alterations to their data collection procedures. In several cases, however, it has resulted in procedures that were implicit to becoming more explicit and formulated in the manual. In some cases this has resulted in new and more extensive versions of the manual being drafted. The manuals of the various surveys are not included in this report but can be viewed through the DATRAS website (<http://www.ices.dk/datacentre/datras/datras.asp>).



## **3.2 Description of the data**

### **3.2.1 Existing databases**

#### **IBTS North Sea**

In the early 1980's a computerised database was set up at ICES headquarters in Copenhagen. All IBTS data collected from 1983 onwards in the North Sea, Skagerrak and Kattegat were stored in this database. Through the EU-financed study "Input of historic IBTS data" (CFP 95/069) all data collected before 1983 that were not yet available in the ICES database were computerised and added to the database. At present a database of surveys carried out in the North Sea, Skagerrak and Kattegat during the 1<sup>st</sup> quarter (1965-present) and other quarters (1991-present) is well established and maintained at ICES Headquarters.

#### **IBTS western and southern divisions**

No central database exists for the IBTS survey data conducted in the western and southern divisions. These data are still held in the respective national institutes that conducted the surveys.

#### **BITS**

Data from 1991 onwards of the 1st and 3rd quarter surveys in the Baltic have been stored in a database at ICES as part of the EU funded project "Establishing a Baltic International Trawl Survey (BITS) database" (No 96-072). The database and the exchange format were set up according to the principles of the IBTS database. The database does not contain all fish species caught as Poland will only deliver data on cod and flounder, the other countries deliver data on all fish species caught.

#### **BTS**

The data of the offshore beam trawl surveys that fall under WGBEAM (e.g. Belgium, Germany, Netherlands and England, UK) are held by the institute that chairs WGBEAM. For the past three years this was the Netherlands, the coming years it will be England, UK. The data cover much of the North Sea, Channel and Irish Sea for the period 1985- present.

### **3.2.2 Exchange format**

Before the DATRAS project IBTS and BITS had two similar exchange formats but with a few minor differences. For example, BITS reported the mean weight of fish in each length class while IBTS did not. Within the DATRAS project one common exchange format was created where all the fields needed to report the data collected during all the trawl surveys are included. Therefore the participants of specific surveys do not necessarily need to fill in all fields. Fields may be mandatory in one survey but not in another and the coding and value ranges of a certain field may differ between surveys. This type of information is in a separate database and the reporting format that was agreed on, can be found for each of the surveys at <http://www.ices.dk/datacentre/datsu/selrep.asp> (see appendix I for an example).

## **3.3 Availability of data within DATRAS**

The database is designed such that it can hold all necessary types of data. An overview of the type of bottom trawl surveys conducted in European waters and the countries that participate in the surveys and/or whose data are part of DATRAS (Table 3.1) shows that the database contains

at least some data of every gear/area combination thereby confirming the adequacy of the design. However, as not all national institutes were involved in the DATRAS project and it was decided that all data should be checked for errors and corrected by the institute responsible for these data, only the data of the institutes part of DATRAS could be loaded into the database.

Most of the remaining institutes have indicated that they will start submitting their data to the database soon. However, for a few institutes there is a more structural problem in that they do not wish to submit their data to a central database that can be accessed by other parties. Although the data access policy (chapter 5) covers this objection there is, as yet, no commitment of them to a central database such as DATRAS.

Table 3.1. Gear and area combinations against the countries that conduct the surveys. S indicates the country participates in the survey, D indicates the data are in the database.

Country	Bottom trawl			Beam trawl	
	Baltic	North Sea	Western Areas	Southern Areas	North Sea
Belgium					S
Denmark	S,D	S,D			
England		S,D	S,D		S
Estonia	S,D				
Finland	S,D				
France		S,D	S,D	S,D	
Germany	S,D	S,D			S
Ireland			S		
Latvia	S,D				
Lithuania	S,D				
Netherlands		S,D			S,D
Norway		S,D			
Poland	S,D				
Portugal				S	
Russia	S,D				
Scotland		S,D	S,D		
Spain			S	S	
Sweden	S,D	S,D			

### 3.4 Indices

All surveys described in this report and for which data are expected to become part of the database were initiated to deliver fishery-independent data (i.e. indices) to be used for the tuning of stock assessments of one or more commercial species. Thus far these indices were not calculated centrally but with all the data in one central database the possibility exists to make the calculation of these indices an integral part of DATRAS. To this end an inventory was made of the indices the surveys BTS, IBTS and BITS are supposed to deliver and the assessment working groups to which these indices must be delivered. Table 3.2 gives an overview of the different species, where they are caught and in which working groups the indices are used.

Table 3.2: Overview of the species, for which the surveys provide indices.

Survey	Area	Species	Assessment WG
<b>BTS</b>	North Sea, Celtic Sea and Irish Sea	plaice	northern shelf WG (VIIa), southern shelf (VIIe, VIIe, VIIg,f), and North Sea and Skagerrak WG
		sole	northern shelf WG (VIIa), southern shelf (VIIe, VIIe, VIIg,f), and North Sea and Skagerrak WG
<b>IBTS Q1</b>	North Sea and Skagerrak	cod	North Sea and Skagerrak WG
		whiting	North Sea and Skagerrak WG (Usually but not in 2001)
		saithe	
		haddock	North Sea and Skagerrak WG
		Norway pout	North Sea and Skagerrak WG
		herring	Herring Assessment south of 56 WG
		sprat	Herring Assessment south of 56 WG
<b>IBTS Q3</b>	North Sea and Skagerrak	mackerel	Mackerel and Horse Mackerel Assessment WG (Made available–not used but deemed important for distribution etc)
		Cod	North Sea and Skagerrak WG (Scottish & English separately)
		Whiting	North Sea and Skagerrak WG (Usually but not in 2001)
		Haddock	North Sea and Skagerrak WG (English data only)
<b>IBTS Q1</b>	VIa	Norway pout	North Sea and Skagerrak WG (Scottish & English separately)
		Cod	Northern Shelf WG
		Haddock	Northern Shelf WG
		Whiting	Northern Shelf WG
		Monkfish	Northern Shelf WG (Exploratory data analysis)
<b>IBTS Q4</b>	VIa	Mackerel	Mackerel and Horse Mackerel Assessment WG (Made available–not used but deemed important for distribution etc)
<b>BITS</b>	Baltic Sea	cod	Baltic WG

Note that the aim is not to develop any new indices. DATRAS just provides the opportunity to have indices calculated in a standardized manner from a central database of which the quality is assured, not just within the surveys but also between surveys. The development of new indices should still be done by the survey- or assessment working groups.

### 3.4.1 Beam trawl survey

The BTS survey consists of many surveys not totally integrated and in different areas. The target species are flatfish. Indices are therefore calculated separately for each survey/area combination. In Division VII, area stratified age–length keys are used to derive mean numbers at age by rectangle within the survey area. The index of abundance is derived from the un-weighted arithmetic mean across all rectangles (ICES C.M. 2001/G:06). In the North Sea, Netherlands BTS, the indices are calculated as above except that they are based on a reduced section of the full survey area known as the “Index Area”.

### 3.4.2 BITS Indices

In the BITS survey, indices are only calculated for cod. There are two cod indices: one for the eastern Baltic, sub-divisions 25+26+28 and one for the western Baltic, sub-divisions 22+24. In each of these sub-divisions the CPUE at length per haul from before 2000 is standardised to the standard TV3 trawl used by all vessels since 2000 by multiplying with a conversion factor (Anon. 2001). Then the mean age at length per depth stratum and sub-division are calculated and weighted with the surface area (m<sup>2</sup>) of the stratum. From these means, the mean catch at age per sub-division is calculated and then the mean catch per index area.

### 3.4.3 IBTS indices

In IBTS North Sea the indices are calculated per index area, which are specific for each species. The indices are calculated as mean at age per statistical rectangle and then as a mean of the statistical rectangles over the index area. The exact location of these index areas for the various stocks (notably cod and whiting in the Eastern Channel and North Sea and haddock and Norway pout in the North Sea and Skagerrak/Kattegat) are the subject of an ongoing debate and may still change. Some statistical rectangles are reduced in size due to land or very shallow water. For herring, sprat and saithe the mean CPUE at age are weighted with the percent covered with water for these statistical rectangles.

The index is calculated as the unadjusted arithmetic mean over the statistical rectangles. The survey is designed so that each rectangle is sampled at least twice (once by each of two countries) - some rectangles (those that have special importance for herring) may be sampled up to six times. The Index is constructed as follows :

Stage 1 -- standardise results from all hauls to a 60 minute tow

$$I_i = \frac{x \cdot 60}{t}$$

Stage 2 -- calculate mean catch per 60 minutes within each rectangle

$$I_{ii} = \frac{\text{Sum } [I_i(1) - I_i(k)]}{k}$$

Stage 3 -- Index = arithmetic mean of all rectangles within the species Standard Area

$$SI = \frac{\text{Sum } [I_{ii}(1) - I_{ii}(n)]}{n}$$

where x is the number of a particular species caught in a tow  
t is the time in minutes  
k is the number of tows in a rectangle  
n is the number of rectangles in a Standard Area

The arithmetic mean has the advantage of being simple and easily understood and should be proportional to the actual abundance. Disadvantages are :

- a few relatively large catches weigh heavily in the index
- the stated confidence limits for the Index may be inaccurate since the latter may not be normally distributed.

Both problems derive from the fact that the distributions of catches from trawl surveys are usually highly skewed.

The Index can be adjusted to take into account rectangles that have not been fished by creating data for the missed rectangles. The average contribution of a rectangle is calculated from the long term average catches in the period 1974 - 1979; this average catch is adjusted by an index of importance for each rectangle. The resulting catch rate can then be fed into the second equation above. In recent years the surveys have been very comprehensive and few, if any, rectangles have been missed and consequently the adjusted Standard Index is little used at the present time for the gadoid species. The herring Index is adjusted to make allowance for missing rectangles.

Other indices have been looked at e.g.

Stratification by density. Density strata were selected on the basis of the average density distribution over the period 1972 -1981 and then stratified means were calculated for the standard area on the assumption that within each stratum the hauls were randomly distributed. This index appears to be more resistant against outliers than the Standard Index but needs further investigation.

'Bootstrap' Index. This involves estimating an index of abundance from the distribution of the mean catch rates in a stratification by standard roundfish reporting area. Again this index requires further investigation.

Salinity. There is some evidence that the highest catches of young cod are encountered in waters with less than 34 promille salinity and by excluding tows made in waters below the 34 promille salinity band it is possible to obtain a better correlation between the Index and VPA. There are difficulties in constructing such an index because the boundary of the salinity band moves from year to year and therefore entails some very precise sampling. Currently the problem is being solved by using a new Standard Area for cod which excludes the 5 rectangles most affected by the 34 promille salinity band. For haddock the 35 promille salinity band may be significant; whiting shows no obvious correlation with salinity.

### **3.5 Expert advice**

The descriptions of the various surveys indicate that over time the sampling procedures, gear characteristics, timing of the survey or the area covered may have changed to a more or lesser extent. This probably affects the catches and hence may bias the analyses and cause erroneous conclusions to be drawn from these analyses. In order to prevent this some recommendations are provided in this chapter that can be used as guidelines using these survey data. As the survey/area combinations provide a certain level of consistency within them, these recommendations are given for each survey/area combination separately.

In providing the guidelines two aspects of the data were considered:

- Time-series
- Spatial coverage

As some surveys or countries within surveys initially only sampled a suite of commercial or "standard" species there is also information included on the year all participants of the survey started sampling all species. This is particularly relevant for studies on biodiversity or the entire fish community.

For extractions of data for analyses on anything other than the community there is a chance that the results are biased if only those hauls in which that species or suite of species is caught, are used. Hauls with 0 catches should be included in analyses unless there is a good reason not to.

## **IBTS North Sea**

The IBTS North Sea is structured by quarter and therefore the data can only be analysed on a quarterly basis. The first quarter survey represents the longest time-series but has also been subject to the biggest changes over time. All of these changes had taken place before the third quarter survey started (1997) and any type of analyses can be performed on the whole dataset. The 2<sup>nd</sup> and 4<sup>th</sup> quarter were only sampled as part of a series of quarterly surveys in the period 1991-1996. This period can therefore be used when exploring seasonal patterns. As the changes that may cause bias only occurred during the 1<sup>st</sup> quarter survey, expert advice is only provided for this quarter.

### Time-series

**Gear** – Prior to 1977 a 78' Dutch herring trawl was recommended as the standard trawl but in 1977 this recommendation was altered in favour of the GOV. The latter became the standard gear operated by all nations from 1983 onwards.

**Haul Duration** – Initially the hauls were of one hour duration but in the early 1970's this recommendation was changed to 30 minutes. In practice this did not become the standard until 1999. Catches are always reported as numbers per hour trawling.

**Age Determination** – From 1980 onwards age length keys (ALKs) were expanded to include ages 1 to 6. Prior to 1980 only the younger ages were reported.

**Sex/Maturity** – From 1981 onwards sex and maturity (using the 4 stage ICES scale) were reported by all nations.

### Spatial coverage

The whole North Sea, the Skagerrak and the Kattegat was covered from 1974 onwards. The stratification of the survey has always been based on the grid of ICES statistical rectangles. Each rectangle is usually fished by the ships of two different countries so that at least two hauls are normally made per rectangle. Within rectangles vessels are free to choose fishing positions but because most nations choose from a limited list of trawl tracks the procedure is defined as semi-random.

### Community

Records on all species exist for all countries from 1988 onwards.

## **NE Atlantic, Quarter 1, Scottish West Coast Surveys**

### Spatial coverage

The Scottish west coast, within the 250 metre contour line, from west of the Orkneys to Northern Ireland and the Clyde. In 1996 the survey area was extended to the northern Irish Sea. The stratification of the survey has always been based on the grid of ICES statistical rectangles. Each rectangle is usually fished at least once except for the Irish Sea where multiple hauls are made in each of the rectangles surveyed.

### Time-series

The whole area was sampled consistently from 1996 onwards

Age Determination – Ages are reported from 0 to the maximum obtained.  
Sex/Maturity – Sex and Maturity (ICES 4-stage scale) are reported.

#### Community

Records on all species exist for the complete time-series

### **NE Atlantic, EVHOE survey**

#### Time-series

From the start of the survey in 1997 there have been no major changes in the survey. In 1999, however, the survey was conducted a month later in the year

Ages are reported for: *Lophius budegass*, *Lophius piscatorius*, *Merluccius merluccius*, *Gadus morhua*, *Merlangius merlangius*, *Lepidorhombus whiffiagonis*

#### Spatial coverage

Celtic Sea and Bay of Biscay covered. In 1999 part of the Bay of Biscay could not be covered due to the oil spill from the ERIKA.

#### Community

From the beginning of the survey all fish species as well as three cephalopods : *Loligo vulgaris*, *Loligo forbesi* and *Illex coindetti* have been sampled for length composition.

### **Baltic Sea, BITS**

#### Time-series

The survey is a quarterly survey with surveys in the first and fourth quarter. However, as there is the risk of ice in the north eastern part of the Baltic Sea some part of the first quarter survey is carried out in April. These data should be considered part of the first quarter.

Standard gear is applied from 2001 onwards. Cod data before that are converted from the old gears into the new gear (TV3) using conversion factors. For all other species these gear differences may cause a bias.

#### Spatial coverage

The Baltic Sea has not been covered very well before 2001. After 2001 the entire Baltic Sea can be considered covered from a cod assessment point of view (ICES sub-division 22-28). This implies that areas where cod does not occur are not necessarily covered.

The use of different gears in the period before 2001 may also cause bias in the spatial distribution patterns of species.

#### Community

For all years all species are delivered from all countries except from Poland where only cod and flounder are provided.

### **North Sea, Beam trawl survey**

#### Time-series

The longest time-series that is consistently sampled from the beginning of the survey (1985) is that of RV “Isis” using a BT8 gear.

Age and sex and maturity are only consistently reported for Plaice (*Pleuronectes platessa*) and Sole (*Solea vulgaris*).

#### Spatial coverage

RV “Isis” using a BT8 gear has covered the south-eastern North Sea consistently ever since the start of the survey. In 1996 the area was expanded using RV “Tridens” and now also covers the south-western and central North Sea. As these vessels use different types of gear the two can not be combined without at least a conversion factor.

#### Community

Records on all species exist for the complete time-series

## 4 The DATRAS system

The bottom trawl survey data go roughly through three stages in the DATRAS system starting with the compilation of the files by the data submitter and ending with the data being available on the web (figure 4.1).

The first stage is the data submitting stage where data go through a screening procedure and the data submitter must rectify all errors before the data can pass to the data manager in ICES. The data manager in ICES also screens the data before data are allowed to be loaded into the database. Three databases were developed to assist the data screening process:

- RECO DB
- ITIS
- DATSU DB

These databases and the data screening process will be described in more detail in the following chapters.

The second stage is after the data are screened for errors and they are loaded into the DATRAS database. The DATRAS database also builds on the RECO and the ITIS databases. After loading they are checked by comparing the age/length data against age/length data from other countries already in the database. Outliers are checked and errors are corrected.

The third stage is after indices and standard aggregations of the data are calculated from the checked data and the results are transferred to the data warehouse database where they can be accessed from the web

In the following chapters both the structure of the database and the flow of the data processing is described in more detail including documentation of the databases and programs.



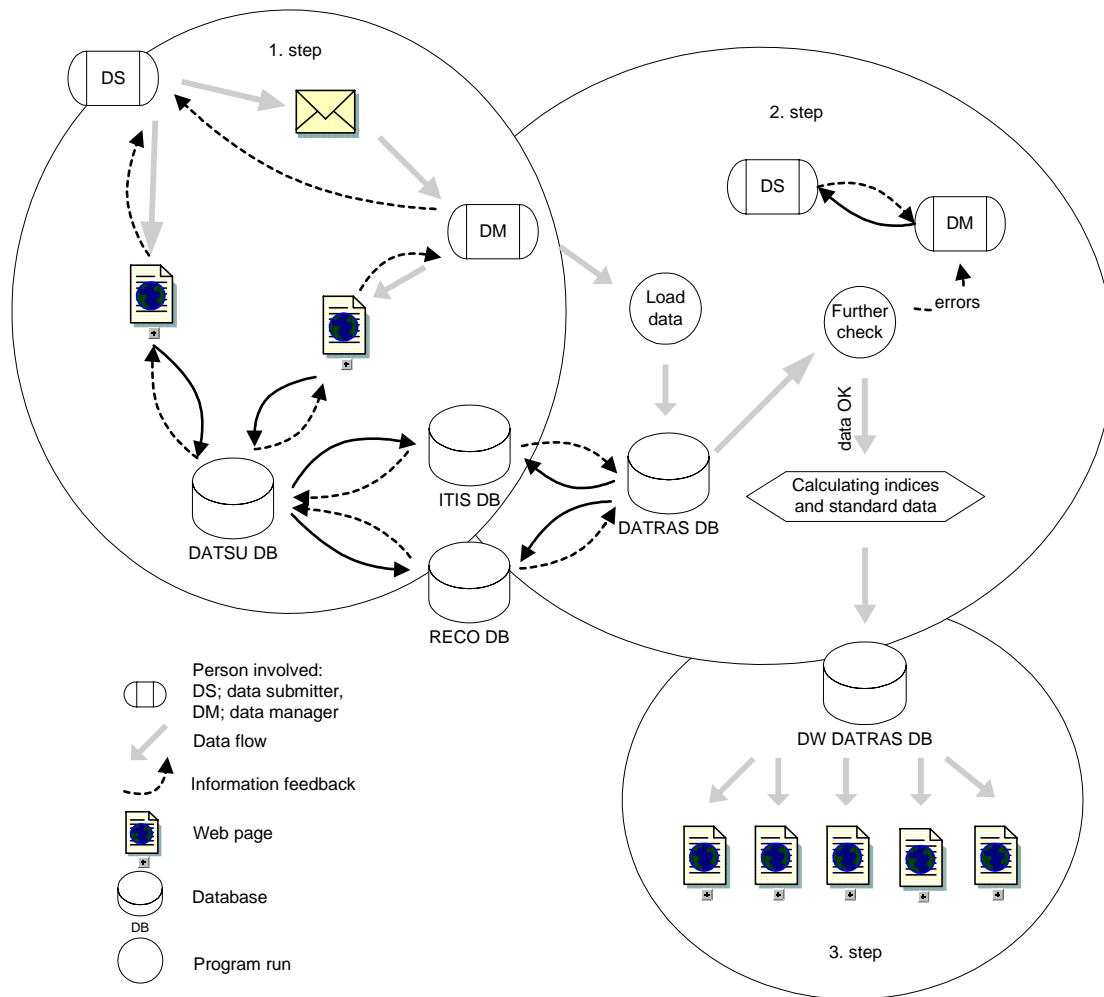


Figure 4.1. Data flow within the DATRAS system distinguishing five major database components part of three stages.

#### 4.1 Software

All databases are placed at an MS SQL server and all data handling on the server side is done in SQL server stored procedures except the import procedure which is done in Data Transformation Services (DTS packages). The client side is split in two: internal use and external use. Internal use is mainly developed for ICES data managers where ACCESS project is used as front-end to the SQL server and all coding is done in Visual Basic for Application (VBA) and SQL server stored procedures. The external front-end to the databases is always a web page coded in Active Server Pages (ASP). Through this web page data are selected directly from the database, either by calling procedures stored on a SQL server or calling a Visual Basic (VB) dll which calls several of these SQL server stored procedures.

#### 4.2 Database structure

Five major database components can be distinguished (Figure 4.1): two databases that contain the reference codes (RECO), and species codes (ITIS), one database holding the information on the structure of the various survey datasets and the checks required on them (DATSU), and

finally the DATRAS database system which consists of a part where the raw data are stored (DATRAS) and a part where all aggregated data are stored (DW DATRAS). All output is generated from the latter two databases.

#### 4.2.1 Reference code system (RECO)

As part of the DATRAS project the RECO database was developed to store all reference codes necessary for ICES' reporting formats (e.g. ship codes, country codes, gear codes, etc.). All other databases and systems receive codes from this database. This implies that codes are only updated once in the RECO database and all systems use references to that database. The reference codes and the related types of codes can be found at the web <http://www.ices.dk/datacentre/reco/>.

The RECO database diagram can be found in appendix II and an overview is given in table 4.1. Codes and relationships between codes can be retrieved from this simple structure.

E.g. to get all the ships used by Sweden one should select the code type 'TS\_Country' from tblCodeType, relate it to tblCode and select 'SWE' from tblCode, relate tblCode to tblCodeRel and relate tblCodeRel to a new tblCode. Then all ships belonging to Sweden are found in tblCode.

Table 4.1. Tables in the RECO database

<b>Tables</b>	<b>Description</b>
tblCode	Stores codes (ARG or SWE) and description of the code (Argos or Sweden).
tblcodeType	Stores the type of code (TS_ship) and the description of codes (trawl survey ship codes)
tblCodeTypeRel	Stores the type of join (One-To-Many) between two types of codes (TS_Ship and TS_Country) and the ids that make the join
tblCodeRel	Stores the relationship between two codes (SWE and ARG)
tblData	Data related to a code. This could be weighting factor of a statistical rectangle or the horse power of a ship
tblDataType	Describes the type of data.

The TblCode and tblCodeType are copied into the other databases which are part of the DATRAS project. Data are moved from the RECO tables to the other database tables when they are changed or updated with new codes. This task is done by specially developed triggers at the SQL server which ensure that the databases always reference the same codes as in RECO. No code is stored in other database tables than those occurring in tblCode. In all other tables the codes are replaced with the code ID from tblCode in the database in question.

#### 4.3 Species classification and coding (ITIS)

As the DATRAS database is expected to be an easily accessible database with reliable information on species names and their hierarchical classification it is necessary to incorporate a taxonomic coding system that ensures high quality with valid classifications, revisions, and additions of newly described species.

For the DATRAS project it was decided to use the TSN codes from the Integrated Taxonomic Information System (ITIS), <http://www.itis.usda.gov> as this is maintained and regularly updated. For each scientific name, ITIS includes the authority (author and date), taxonomic rank, associated synonyms and vernacular names where available, a unique taxonomic serial number, data source information (publications, experts, etc.) and data quality indicators.

Previously, most surveys and ICES used the NODC coding system. However, this is not maintained anymore and was therefore abandoned. As most national laboratories still use the NODC codes in their national systems it is still possible to report data with these codes but they are converted in ICES to the TSN codes. ICES has a complete copy of the ITIS database which is updated every month from the ITIS web page. To allow data submitters and users of the products to convert between the two code systems DATRAS provides a web page that can be used to search for and convert between TSN codes, NODC codes, scientific name and common names: <http://www.ices.dk/datacentre/reco/qryspec.asp>. In addition it is possible to download the complete species list of a survey with TSN, NODC and scientific name as CSV file <http://www.ices.dk/datacentre/datsu/selrep.asp>.

#### **4.4 Data screening utility (DATSU)**

In order to assure the quality of the data in the database all data need to pass through a data screening procedure before they can enter the database. Therefore a data screening program, DATSU, was developed with two requirements:

1. DATSU should be able to handle all existing survey formats in ICES that have a hierarchical structure with a main record type and related record types below
2. To prevent different versions of the data screening program existing in various institutes there should only exist one version of the program maintained by ICES

These requirements were met by creating one database at ICES, holding all information on the datasets and their structure as well as the checks that apply to the structure and record fields together with a web interface that allows access to this database. However, before data can be submitted to the database they have to pass through the data screening utility and only files that pass are added to the database. Storing all information in a database makes the system flexible and easy to maintain while the web interface allows access from anywhere with an internet connection and as all code runs at the ICES server the existence of only one checking program version is assured.

Figure 4.2 shows the flow of the data and the process involved in the data screening. The process is as follow:

- The file is uploaded to the ICES server: <http://www.ices.dk/datacentre/datsu/>
- The file is screened according to the information stored in the database
- The data submitter gets an e-mail with reference to a web page on which resulting errors are shown. This web page can be entered directly from the web page where the data screening program was run or from the address in the e-mail.
- The user updates the file and resubmits it to the ICES server
- When the data file has been screened and passes it is send to the data manager at ICES for a final check

- If the data manager finds any errors they will be reported back to the data submitter who can correct the errors otherwise the data are loaded into the database

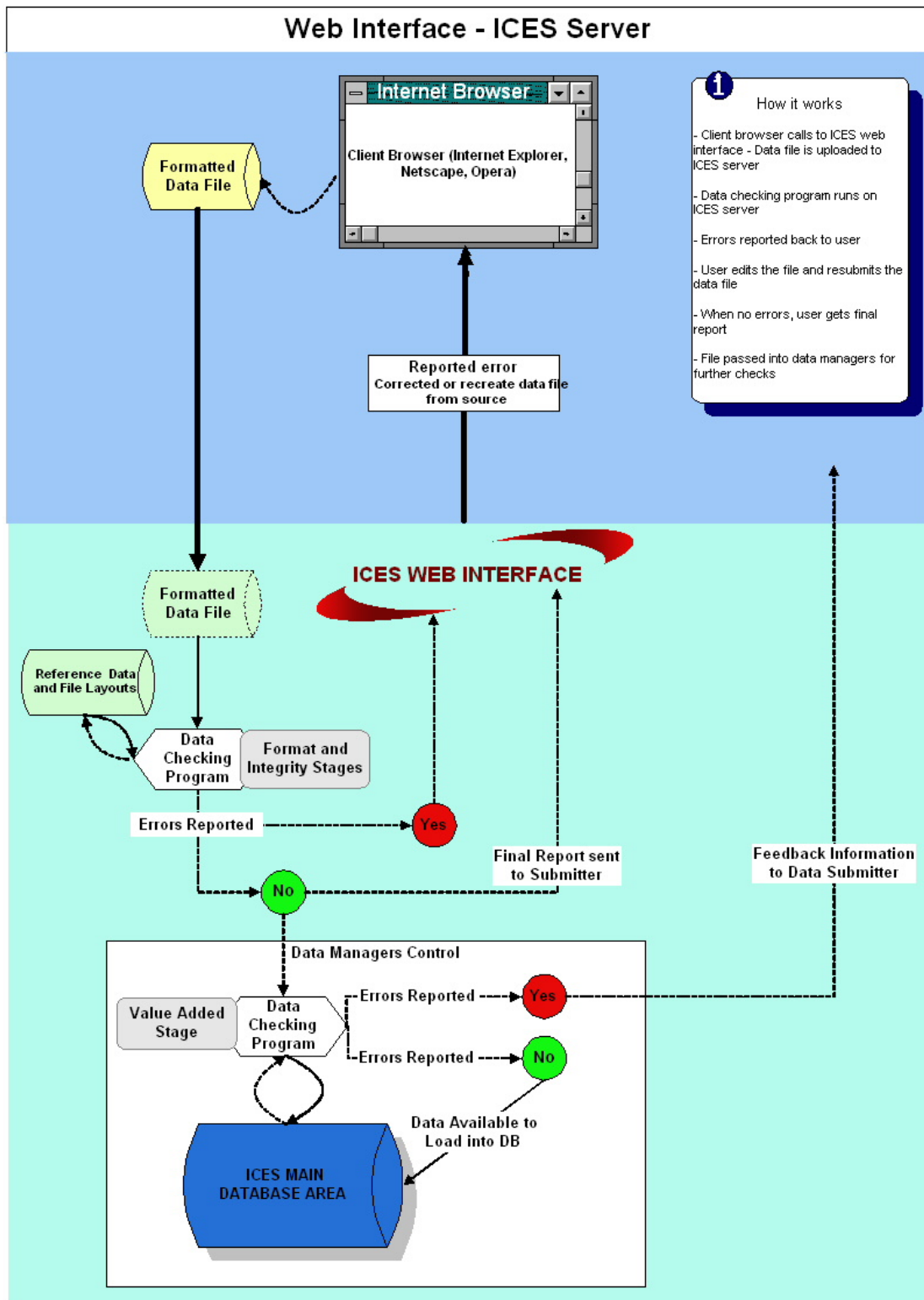


Figure 4.2. Flow of data in the data screening process

The data screening program builds of the following components:

- The DATSU Database
- A user front-end for the data manager where all the database tables can be maintained
- The program, made in Visual Basic (VB) and compiled to a dll and a number of stored procedures on the SQL server
- A number of web pages which calls the program and presents the results

These will be described in the following sections.

#### 4.4.1 Checks

The format of the datasets and the checks performed on each of them can be found on the web page <http://www.ices.dk/datacentre/datsu/selrep.asp>.

The following types of checks can be distinguished:

1. Format and integrity: checks the right numbers of fields in each record in the data file. It checks that there is a main key (e.g. year and quarter) running through the whole data file, that the sub keys are consistent in the file (e.g. ship, haul number etc.) and that there exist a unique constrain in each record.
2. Mandatory records and fields: checks that the mandatory records are there and that the fields are filled in.
3. Range: checks that if the value of a field is a number then the value is within the allowed range (e.g. latitude of the North Sea IBTS should be between 51 and 62 degrees)
4. Reference codes: if the value in a field is a code it will be checked against the reference code list. This can depend on dataset and/or other fields.
5. Other survey-specific checks: e.g. checks on positions, depth, day/night codes, length code against length classes and identical values in non key fields.

#### 4.4.2 DATSU Database

The DATSU database stores all information needed to build the format for a version of a dataset and run the checking program on that dataset (for the database diagram see appendix III). The program is designed to be able to handle more than one version of the same format. For example for the same survey (e.g. IBTS and BITS) we defined different versions:

- before 2004 which must allow all historic codes
- from 2004 onwards which is more restrictive with regard to the codes allowed.

The information required to build a dataset format can be found in the tables described in table 4.2:

Table 4.2. Tables in DATSU used to generate a format for a dataset

<b>Table name</b>	<b>Description of table</b>
tblSU_DatasetVersion	Dataset version ID (where the code for the dataset and its version is stored in tblcode) and information on release
tblSU_RecDataset	RecordID connected to the dataset and information on mandatory and record order

tblSU_Record	Lookup table where name and description of the dataset can be found.
tblSU_FieldsRec	Fields related to records (E.g. what fields belong to a record) and information on what fields are mandatory or legal (e.g. rigging only in BTS surveys), or part of a key. If so, there is a check that identical values are used within that key (e.g. total number, number measured, total weight and length code is the same for each species).
tblSU_FieldCodes	Lookup table with field codes and data type for the field
tblSU_FieldDefinition	Definition of fields with description and where different versions can have different width and headings
tblSU_KeyType	Lookup table with possible codes for key and a description of the key
tblSU_IdenticalValue	Lookup table with codes for identical value and a description of the identical value
tblDataType	Lookup table with data types

Checks can be made on each level in a dataset (dataset, record and field). All error messages are stored in tblSU\_Check which also identifies the checks done by the program. The tables necessary for the checking of data are described in table 4.3.

Table 4.3. Tables in DATSU storing information need when screening data

Table name	Description of table
tblSU_Check	Check ID and their messages
tblSU_Error	Describes if a check results in an error or a warning
tblSU_DatasetCheck	Relationship between dataset and checks
tblSU_RecCheck	Relationship between records and checks
tblSU_FieldCheck	Relationship between fields and checks
tblSU_DataRange	Information on data range for a field check
tblSU_FieldLookupCheck	Information on check for reference codes
tblSU_FieldLookupRel	Relationship between a field check and tblSU_FieldLookupCheck
tblSU_RelSpecies, tblSU_SpeciesVariables, tblSU_SpecVarType, tblSU_SpecLkpRel and tblSU_SpecLkp	These tables deal with checks on whether species are legal in a specific dataset and check on the relationship between species code and other related fields e.g. the relationship between length and species.

Information on each data screening session is stored in tblSU\_Session which includes information on user, data file and records. All reported errors are stored in tblSU\_ReportedErrors which links to checks (tblSU\_Checks) and the session (tblSU\_Session). This logging of information is provided to the user when the user gets the results from the data screening program and it is used for extraction of errors to the error report (figure 4.2).

### 4.4.3 Internal front-end

The database is managed internally in ICES through a front-end with the following functionality:

- Update lookup tables. The database have a number of lookup tables as described in table 1. These are updated through the front-end:
  - Dataset
  - Record type
  - Checks
  - Fields codes
  - Error types
  - tblCode: done by getting codes from RECO and IT IS
- Combine datasets with records and include fields to records
- Updates records and fields with mandatory and key information
- Include check on dataset, records and fields
- Create relation between dataset and species and species and other fields (like length or age)
- Create reports on the content of the database, like reports on the format of dataset and checks
- Copy a dataset and related checks into a new dataset. Thereby a new dataset with related checks can be created without having to rebuild all the relations between dataset, records, fields and checks.

Figure 4.3 shows the front-end. Each page on the front-end holds different parts of the functionality described above. In the page showed in figure 4.3 the user selects a dataset and a related record. The user then gets the list of field codes related to the record and can now add more fields and checks. Here the user has selected to include the check where codes in the Country field are checked against another field which can be selected in the right side of the form. When the other field is selected e.g. ship, the user clicks on “join check” and next time the data screening program is run for this dataset the check between Country and Ship is performed automatically.

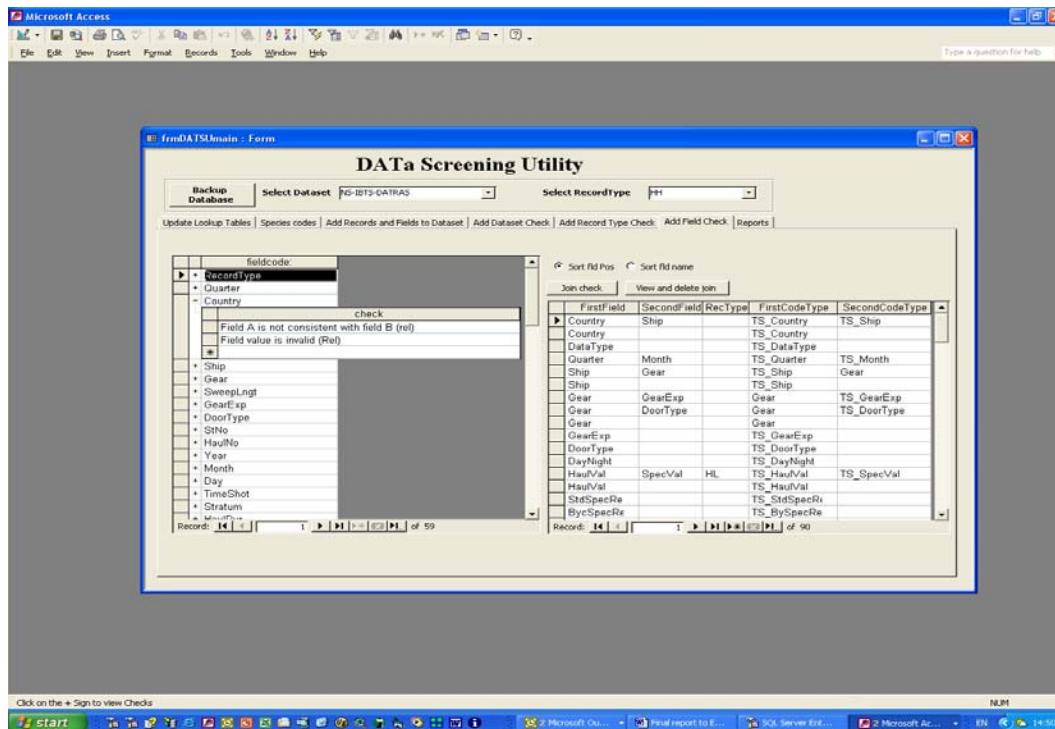


Figure 4.3. Data Screening Utility front-end

#### 4.4.4 Web Interface and running the program

The main entrance to the DATSU program for the external user is the web page <http://www.ices.dk/datacentre/datsu/>. Here, the user can upload the data file, select the type of dataset and set the number of errors that should be returned. After providing this information the data can be submitted.

The web page now calls on the program (VB dll) that splits the dataset into a table per record at the SQL server created from the information in the DATSU database. The program then calls program parts at the SQL server (stored procedures) that do most of the checks by comparing data in the record tables with information stored in the DATSU database (figure 4.4). A few checks are hard-coded in the program e.g. comparing day/night code with position and time shoot and check of day and month in relation to leap year checks.



# VB dll

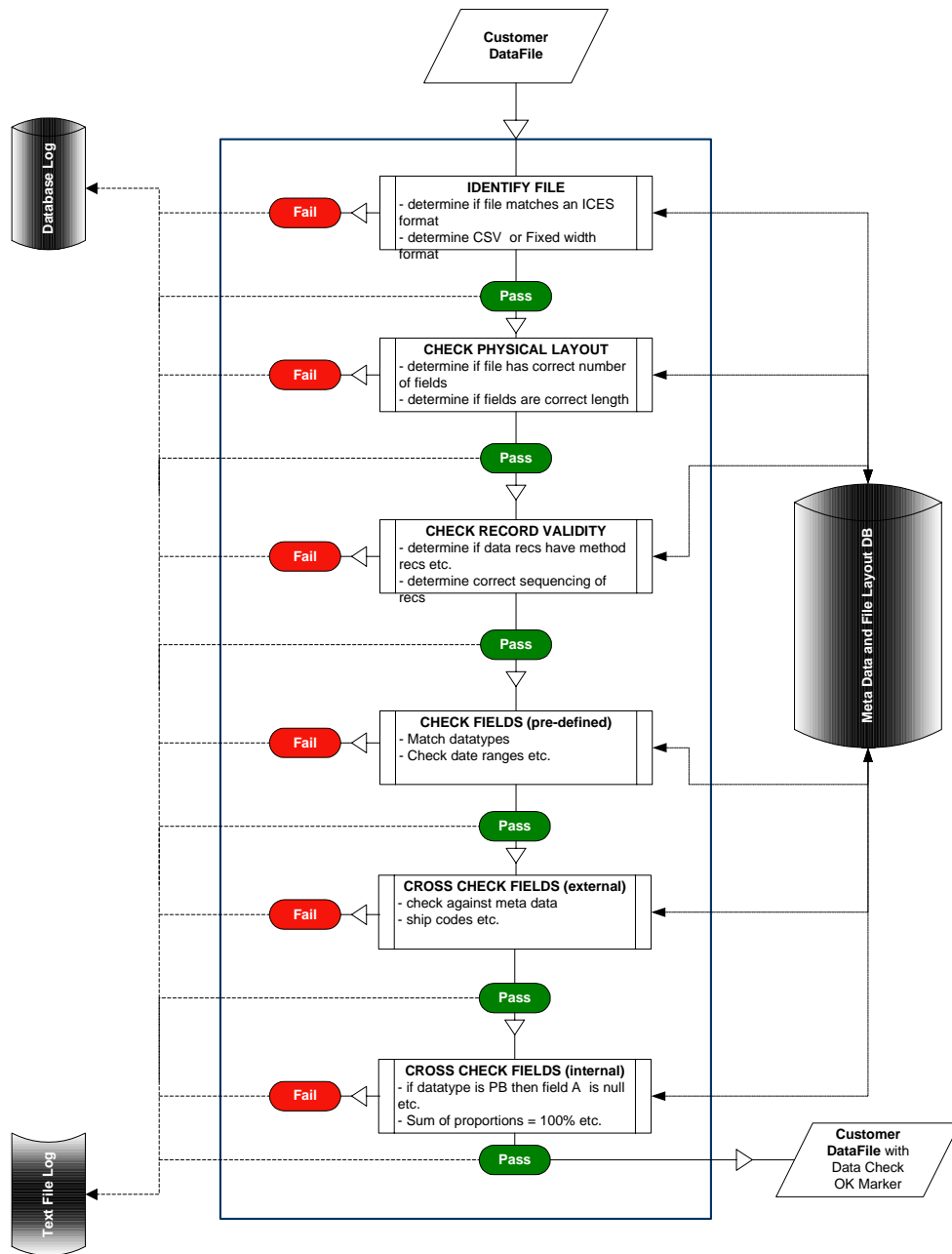


Figure 4.4. The basic idea of how the flow is in the VB dll each check is done by calling a stored procedures

When the program has finished, the user gets a link to the results page at the standby page and by e-mail (e.g. [http://www.ices.dk/datacentre/datsu/datsu\\_err\\_rep.asp?session=3653&Page=1](http://www.ices.dk/datacentre/datsu/datsu_err_rep.asp?session=3653&Page=1)).

The results page shows what is stored in the tblSession, tblReportedErrors and the description of errors form the tblSU\_Check in the DATSU database. The results can be sorted by line where

the errors occurred, type of error, description of error, value of the field and the field where the errors are.

## 4.5 DATRAS database

The DATRAS database system consists of two databases; one where the raw data are stored (DATRAS) and one where aggregated data are stored (DW\_DATRAS). All calculations are done in the DATRAS database after which the calculated or aggregated data are passed to the DW\_DATRAS database. All data presented on the web as well as all downloads, except the raw exchange files which come from the DATRAS database, are extracted from the DW\_DATRAS database.

In the following sections we will describe the database and the procedures that generate the aggregated or calculated output. We distinguish three types of data:

- Raw data: catch at length per species on a haul-by-haul basis
- Aggregated data: catch at length per species aggregated at a higher level, e.g. catches are reported as mean catch per ICES rectangle or depth stratum
- Indices: the ICES working groups that do the stock assessments use surveys to tune their Virtual Population Assessment (VPA). Usually an index is a time series of the mean catch-at-age for a particular survey in a specific area.

When calculating the indices the aggregated data are produced and aggregated data and indices will therefore be dealt with in the same session.

### 4.5.1 Structure DATRAS database

The DATRAS database is built as a relational database where the exchange format is mapped into normalized tables (Appendix IV). Data in all surveys are given as catch-at-length in numbers per haul or per hour or as catch-at-length in numbers per sub-sample. In addition, number at age, maturity and sex per length class are given for sub-samples of the target species. The sub-samples are given either on a haul-by-haul basis, per ICES statistical rectangle or per sampling area (e.g. roundfish or otoliths area) depending on the survey.

Table 4.4 describes the tables in the database. The main entrance into the database is the tblTrawlCruise with cruise information. The tblTrawlCruise is related to tblHaul with all mandatory haul information and to area collected information on Sex, Maturity and Age/Length Keys (SMALK) in tblALK, where the area can be statistical rectangles, round fish area or other sampling areas.

TblHaul is also related to optional information on gear (tblGearDetail) and environment (tblAncillary) and to tblSpecies with recorded species. The table tblSpecies holds all the species related information and as there can be several or no species in a haul the tblHaul is related to tblSpecies with a one to many relationship or no link. For each species there can be both SMALK or length distribution information and the tblSpecies is therefore linked with a one to many or zero to the tables tblALK containing haul related SMALK data and tblSampleInfo with sample information.

tblSampleInfo holds information on sub-sample levels like at what category and sub-sample level are species sampled and total measurements by category, sub-sample and sex. If data are delivered as number per hour and not by sub-sample both the category and the sub-sample level is one. tblSampleInfo is related to tblLength which contain the actual number measured per length or the number per length per hour depending on what ever data are reported as sub-sampled or by hour.

Table 4.4. Tables in the DATRAS database

<b>Table name</b>	<b>Unique constrain</b>	<b>Description of table</b>
TblTrawlCruise	Survey, Quarter, Year, Ship, Gear, Country	Main entrance to the database. Information on the cruise and date of import and which data should be included in further calculations.
tblHaul	tblTrawlCruiseID, ShootLat, ShootLon, DateTime, Depth	All mandatory haul information
tblAncillary	tblHaulID	Optional environmental information related to the haul
tblGearDetail	tblHaulID	Optional gear information related to the haul
tblSpecies	tblHaulID, SpecCode, SpecValCode	Recorded species and the validity code for that species
tblSampleInfo	tblSpeciesID, CatIdentifier, Sex, SubFactor	Species information per sub-sample level
tblLength	tblSampleInfoID, LngtClasMM, LngtCode, NoAtLngt	Length information per species and sub-sample
tblALK	tblSpeciesID, LngtclassMM, Sex, Maturity, Age	Information on haul-based species' Sex, Maturity and Age/Length Keys (SMALK)
tblWeightHaul	tblALKID, individualWeight	Optional information on weight on haul based SMALK
tblALKArea	tblTrawlCruiseID, AreaType, AreaCode, LngtClasMM, Sex, Maturity, Age	Information on area related SMALK
tblWeightArea	tblALKAreaID, individualWeight	Optional information on weight on area based SMALK
tblIndexSetting	Survey, SpecCode, Quarter, IndexArea	Information used when calculating indices and extracting SMALK information
tblSubKeyInfo	tblIndexSettingID	Stores if substitution of ALK is needed for a species in a particular year and area (see paragraph 4.5.2)
tblSubArea	tblSubKeyInfoID, SubArea	Stores the substitution keys

Data are imported into the database by two DTS packages made at the SQL server. A detailed description of how the data import works can be found in appendix V.

## 4.5.2 Aggregation of data and calculations of indices

When all data of a particular survey and year are loaded into the database the program can be executed that calculates standard extractions for that survey and year and updates the data warehouse. This is a two-stage process, first the age/length keys (ALKs) are checked and substitution keys are provided if needed (figure 4.5), then the indices are calculated (figure 4.4).

Before the data manager updates the data warehouse with standard extractions the ALK per species need to be checked for outliers. If there are outliers these are deleted and the countries providing them are notified about it. The ALK is also checked for missing values. If an area has less than 25 otoliths or there are missing age readings for too many age or length classes, ALKs from other areas have to be borrowed (see appendix VI for further description of substitution in North Sea IBTS). The substitution key for that area is also stored in the DATRAS database so that next time indices need to be calculated the substitution is done automatically.

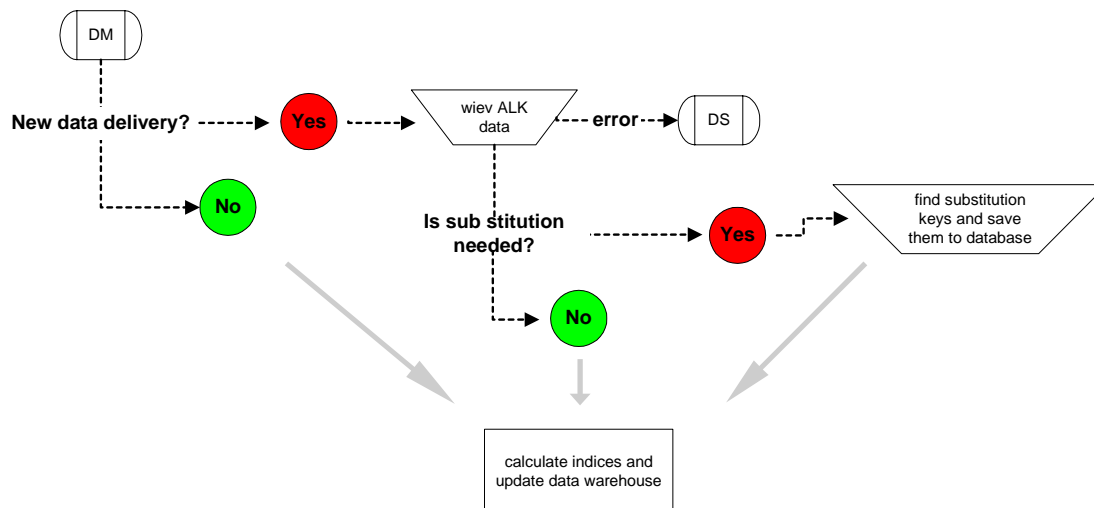


Figure 4.5. The flow describing how the age/length key is handled

When the ALK is checked and the necessary substitutions are done the data manager can decide that data are at a stage where they can be included in the data warehouse database. This is done by running a program that does all calculations and updates the data warehouse database. Although the final product is the indices, a number of intermediate results are also passed on to the data warehouse. The data warehouse therefore contains:

- number per length by haul
- number per age by haul
- mean number per age by statistical rectangle
- mean number per length by statistical rectangle
- indices

The program use tblIndexSettings shown in table 4.5. In this table the program can look up settings like aggregation area and min and max values for age and length.

Table 4.5. Settings for updating DW DATRAS (tblIndexSettings) for those stocks for which indices are calculated

Survey	Species	Quarter	Index area	Max age	Min length	Max length	ALK area	Weighted with area	Day haul only
BITS	Gadus morhua	1	BS_CodEast	10	140	900	TS_SubDiv	Yes	No
BITS	Gadus morhua	4	BS_CodEast	10	70	900	TS_SubDiv	Yes	No
BITS	Gadus morhua	1	BS_CodWest	10	140	900	TS_SubDiv	Yes	No
BITS	Gadus morhua	4	BS_CodWest	10	70	900	TS_SubDiv	Yes	No
NS-IBTS	Clupea harengus	1	NS_Her	5	150	320	TS_RFArea	Yes	Yes
NS-IBTS	Clupea harengus	2	NS_Her	5	60	340	TS_RFArea	Yes	Yes
NS-IBTS	Clupea harengus	3	NS_Her	5	60	340	TS_RFArea	Yes	Yes
NS-IBTS	Clupea harengus	4	NS_Her	5	60	340	TS_RFArea	Yes	Yes
NS-IBTS	Clupea harengus	1	NS_Her1to7	5	60	125	TS_RFArea	Yes	Yes
NS-IBTS	Clupea harengus	1	NS_Her1to9	5	60	125	TS_RFArea	Yes	Yes
NS-IBTS	Gadus morhua	1	NS_Cod	6	150	900	TS_RFArea	No	No
NS-IBTS	Gadus morhua	2	NS_Cod	6	70	1100	TS_RFArea	No	No
NS-IBTS	Gadus morhua	3	NS_Cod	6	70	1100	TS_RFArea	No	No
NS-IBTS	Gadus morhua	4	NS_Cod	6	70	1100	TS_RFArea	No	No
NS-IBTS	Gadus morhua	1	NS_CodCat	6	150	900	TS_RFArea	No	No
NS-IBTS	Gadus morhua	3	NS_CodCat	6	70	1100	TS_RFArea	No	No
NS-IBTS	Melanogrammus aeglefinus	1	NS_Had	6	150	600	TS_RFArea	No	No
NS-IBTS	Melanogrammus aeglefinus	3	NS_Had	6	100	700	TS_RFArea	No	No
NS-IBTS	Merlangius merlangus	1	NS_Whit	6	150	450	TS_RFArea	No	No
NS-IBTS	Merlangius merlangus	3	NS_Whit	6	80	500	TS_RFArea	No	No
NS-IBTS	Pollachius virens	1	NS_Saithe	6	250	900	TS_RFArea	Yes	No
NS-IBTS	Pollachius virens	3	NS_Saithe	6	70	1100	TS_RFArea	Yes	No
NS-IBTS	Scomber scombrus	1	NS_Mack	6	200	450	TS_RFArea	No	No
NS-IBTS	Scomber scombrus	3	NS_Mack	6	50	450	TS_RFArea	No	No
NS-IBTS	Sprattus sprattus	1	NS_SpratIIIa	5	70	160	TS_RFArea	Yes	No
NS-IBTS	Sprattus sprattus	3	NS_SpratIIIa	5	70	160	TS_RFArea	Yes	No
NS-IBTS	Sprattus sprattus	1	NS_SpratIV	5	70	160	TS_RFArea	Yes	No
NS-IBTS	Sprattus sprattus	3	NS_SpratIV	5	70	160	TS_RFArea	Yes	No
NS-IBTS	Trisopterus esmarkii	1	NS_NorPout	6	100	250	TS_RFArea	No	No
NS-IBTS	Trisopterus esmarkii	3	NS_NorPout	6	50	250	TS_RFArea	No	No

The program is run by survey, year, quarter and species. The flow is shown in figure 4.6 and below the tasks at different stages in the program is described.

1

- Extraction of raw length data
- Species recording code is taken into account
- Only valid data
- Check for day/night code
- Handles zero hauls
- Raise data from sub sampling to total number per haul
- If BITS data then CPUE are converted to standard trawl by multiplying with a conversion factor
- Data product is passed on to DW DATRAS. BITS data will be converted back to gear specific data before being included in DW DATRAS

2

- Extraction of ALK data, both haul and area based
- Included ALK from substitution areas
- Check if plus group (e.g. ages 1, 2, 3, 4+) is equal to or below the maximum age for that species (found in tblIndexSetting), if so then plus group is used in the further calculations
- Sum all ages above the plus group

3

- Combining ALK with catch at length
- If there is no ALK for a length for which there are catches recorded then there are three possibilities for obtaining age information for that length class:
  1. If length is less than min length in tblIndexSettings then the age is set to age 1 in first quarter and 0 in all other quarters.
  2. If length is between min length and max length then age is set to the nearest ALK either at a length class before or at a length class after the one which misses an ALK. If there is one below and one after the length class at equal distance in length a mean is taken
  3. If the length is larger than max length the age is set to the plus group.
- After ALK and catch are merged DW DATRAS is updated with number per age per haul

4

- Calculating a mean CPUE per age and length by stratum. The stratum chosen depends on the survey and is looked up in tblIndexSettings. Some surveys use the ICES rectangle as stratum (e.g. IBTS), others use a depth stratum (e.g. BITS)
- DW DATRAS is updated with mean CPUE per stratum

After point 4 the codes splits depending on survey and species as the indices are calculated differently.

5

- For North Sea IBTS herring, saithe and sprat data are weighted with a weighting factor for each statistical rectangle (based on its surface area) before the mean of the index area is calculated. As herring in area 8 and 9 are autumn spawners, ages above one are set to zero for North Sea IBTS herring in the first quarter in these areas.

6

- For all IBTS species the (weighted) mean of all statistical rectangles in an index area is the final produced index.
- DW DATRAS is updated with the indices

7

- For indices in the EVHOE and BITS survey mean CPUE per age and length per depth stratum is weighted with the area of a depth stratum in  $\text{km}^2$
- CPUE per age is summed over all depth stratum
- The index is calculated as the total CPUE in index area is divided by total area in index area

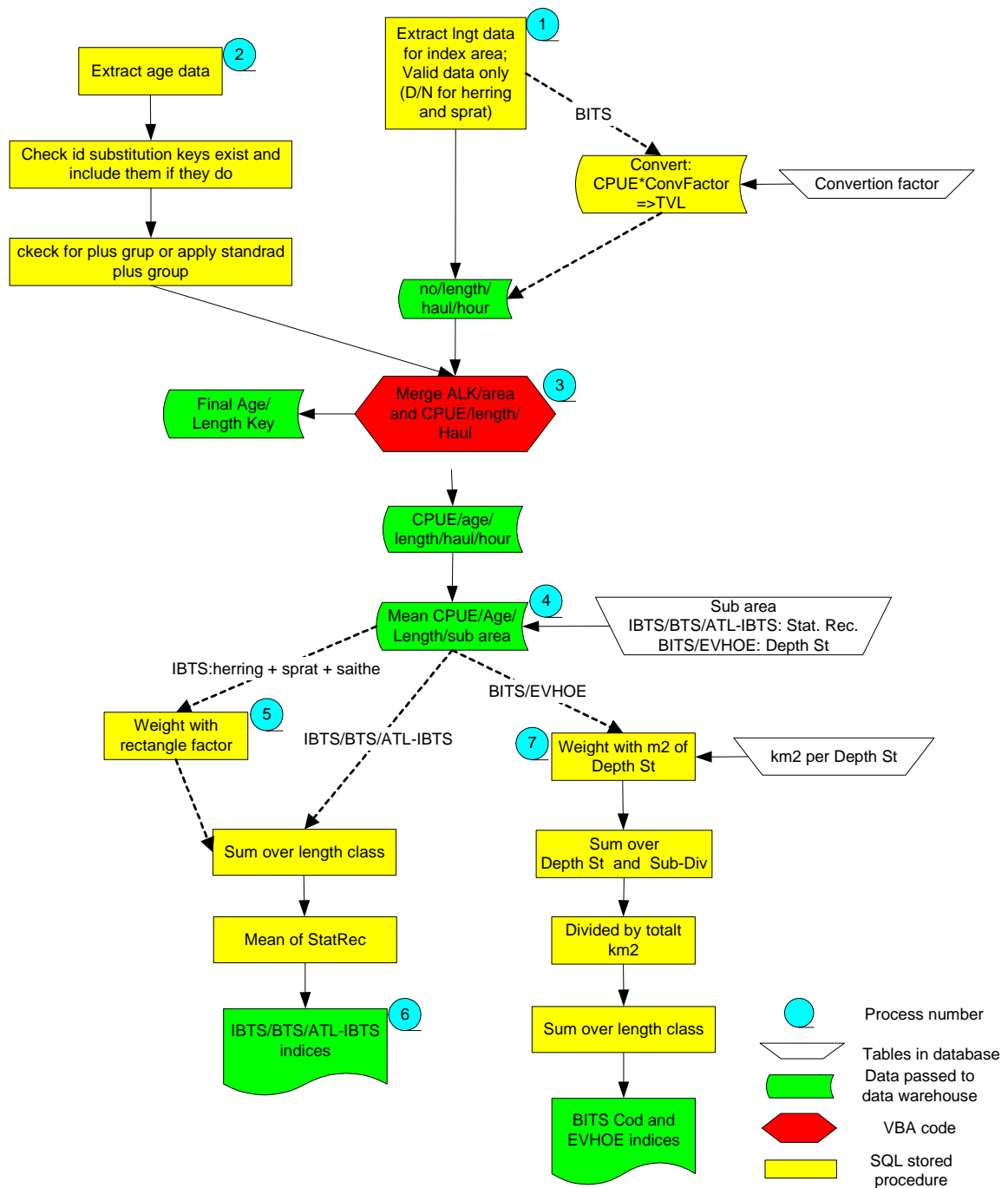


Figure 4.6. Flow of the program calculating indices and output to DW DATRAS

### 4.5.3 Data warehouse

As indices are calculated intermediate results are passed on to the DATRAS data warehouse (DW DATRAS) as described in the previous section. DW DATRAS consists of a relational database and five solitary tables (appendix VII).

Table 4.6 describes these solitary tables which are used for the downloading of CSV files. Two tables contain data on a haul-by-haul basis. Two other tables contain the mean CPUE at the sub area level (table 10) and one table contains the indices.



Table 4.6. Solitary tables in DW DATRAS

<b>Table name</b>	<b>Unique constrain</b>	<b>Description of table</b>
tblCHA	tblHaulID, tblSpeciesID	CPUE per age per hour
tblCHL	tblHaulID, tblSpeciesID, LngtClas	CPUE per length per haul per hour
tblDWMeanCPUEAge	Survey, Year, Quarter, SubArea, Stratum, SpecCode, LngtClas	Mean CPUE per age and length per stratum
tblDWMeanCPUELngt	Survey, Year, Quarter, SubArea, Stratum, SpecCode, LngtClas	Mean CPUE per length per stratum
tblDWIndices	Survey, Year, Quarter, SpecCode, IndexArea	Indices for “standard” species per survey

Pivot tables were considered by the survey WGs the most flexible, and therefore preferred, solution for the downloading of aggregated data. The database is set up such that in the future this should become possible. However, at this stage aggregated data can only be downloaded in some pre-defined formats.

In the following the data structure and the dimensions in the data warehouse cubes (overlying structures in the database needed for pivot tables) are described. The relational database will be able to provide data to two data warehouse cubes/pivot tables, one for extracting CPUE per length and age and one for CPUE per length by sub area.

The relational database contains 2 fact tables and 6 dimension tables (table 9 + appendix VII). The fact tables hold information on CPUE per length and age (tblDWAge\_fact) for the “standard” species (i.e. commercial species for which the survey is supposed to deliver indices for the tuning of the stock assessments) only and CPUE per length (tblDWLngt\_fact) for all species while the dimension tables hold information related to the fact table.

Table 9. Relational tables in DW DATRAS

<b>Table name</b>	<b>Unique constrain</b>	<b>Description of table</b>
tblDWSpecies	TSN	Information on the taxonomic classification of the reported specimen
tblDWAge	tblDWSpeciesID, Age, LngtClas	Age and length information for age related data
tblDWLngt	tblDWSpeciesID, LngtClas	Length information for length related data
tblDWDayNight	DayNight	Day/night information
tblDWArea	Survey, SubArea	Area and Stratum information. SubArea Stratum are survey-dependent (table 10).
tblDWTime	Year, Quarter	Temporal information
tblDWLngt_fact	tblDWLngtID, tblDWDayNightID, tblDWAreaID, tblDWTimeID	Length CPUE data related to species, Day/night code, stratum and time
tblDWAge_fact	tblDWAgeID, tblDWDayNightID, tblDWAreaID, tblDWTimeID	Age CPUE data related to species, Day/night code, stratum and time

Table 10. Survey dependent areas

Survey (1 <sup>st</sup> level)	Area (2 <sup>nd</sup> level)	Sub area (3 <sup>rd</sup> level)
NS-IBTS	NS-Roundfish area (RF)	ICES statistical rectangle
EVHOE	EVHOE areas	Depth stratified EVHOE areas
ATL-IBTS	Scottish sampling areas	ICES statistical rectangle
BITS	Sub-division	Depth stratified Sub-division
BTS	Otolith areas	ICES statistical rectangle

There are four dimensions for each cube/pivot table in the data warehouse (figure 4.7);

- Day/night (tblDWDayNight)
- hauling date (tblDWTime)
- area (tblDWArea)
- species (tblDWSpecies and tblDWAge or tblDWLngt)

The species dimension is split into two tables as the age information only is relevant for the tblDWAge\_fact table and some length classes might not exist for both age and length CPUE. Due to the nature of the taxonomic system where all species are not ordered into the same hierarchical levels the table tblDWSpecies will get columns with null values.

Each dimension has a number of levels at which data can be aggregated by pivot tables or by the SQL server analysis manager (figure 4.7).

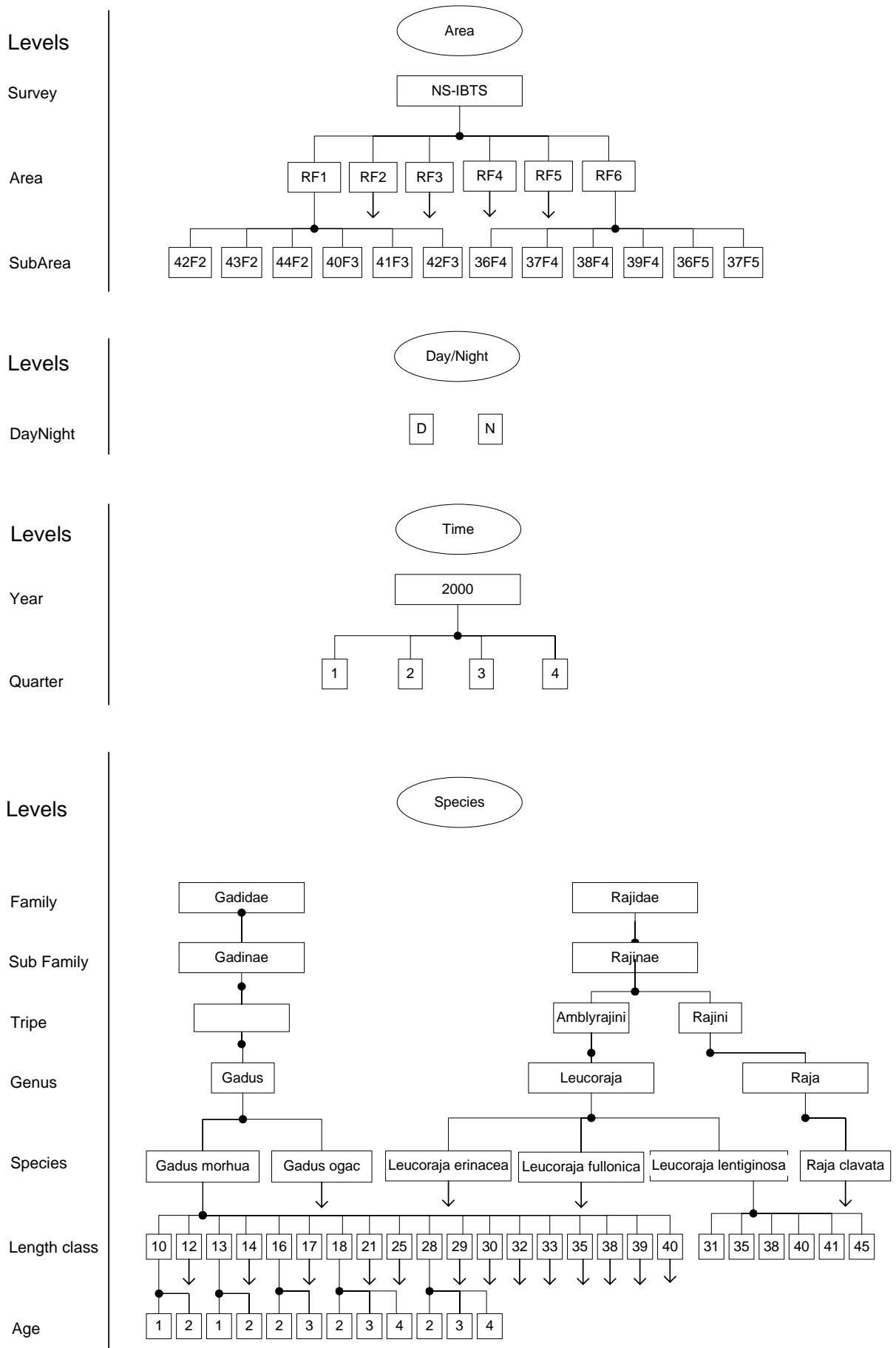


Figure 4.7. Levels at which data can be aggregated in DW DATRAS

## 5 Access policy

In chapter 2 four survey/area combinations were identified. Within each of these survey/area combinations there was agreement on the data access policy. This is exemplified in table 5.1, which distinguishes three user categories:

- Public and other parties that request data, typically for research purposes.
- ICES working groups
- Institutes that have supplied data to the database.

and three data types:

- Standard maps and graphs: Per survey/area combination for all relevant ages of species for which assessments are conducted. Maps will show plots indicating abundance per ICES rectangle or per haul. Time series of the indices and a graph showing the proportion of the age-groups will be generated.
- Aggregated data. A query of the database using pivot tables. Based on these tables, plots and graphs can be made on an interactive basis. The minimum level of aggregation differs between survey/area combinations.
  - ICES rectangle: IBTS in the North Sea, Skagerrak, Kattegat and the BTS in the North Sea, Channel and Irish Sea
  - Stratum: IBTS western and southern divisions
  - Sub-division and stratum: BITS Baltic Sea
- Raw (Un-aggregated) data. These are catch (numbers at length and/or numbers at age) data on a haul-by-haul basis and SMALK (Sex, Maturity, Age-Length-Keys) data per individual.

Data type	User categories		
	ICES WG <sup>1</sup>	Data supplier <sup>2</sup>	Public and other parties
Standard maps and graphs	F	F	F
Aggregated data	P	P	P/R <sup>3</sup>
Raw (Un-aggregated) data	R	P	R <sup>4</sup>

Table 5.1. Matrix showing how data access is determined by “User category” and “Data type” F is the abbreviation for “free access”, P for “password protected access” and R for “access to extracted data after granted request”. Notes:

<sup>1</sup> ICES WGs will have access to data from only those survey/area combinations that are relevant for their recommendations and as such should be specified in those recommendations.

<sup>2</sup> Data suppliers will only have access to data of those survey/area combinations to which the institute has provided data.

<sup>3</sup> Per survey/area combination the members can decide whether individuals will have free access to aggregated data or only after request. If a request is granted, an extraction of the data will be made available

<sup>4</sup> Access can be requested and if granted, an extraction of the data will be made available

All data (other than standard maps and graphs) are protected by passwords and controlled by NT security system. Each institute delivering data to the database can suggest a username and password to ICES which will give them access to the data in those survey/area combinations they are (for a sufficiently long period) part of. ICES WGs will have a password that allows them access to aggregated data, raw data will be issued to the chair of the WG after request. Other parties can request access to both aggregated or non-aggregated data through the ICES website. A standard form must be filled in to inform the institutes involved in the survey(s) on:

- Who is requesting data, including partners in the research project
- The purpose of the data request
- Which data (at what aggregation level) are requested
- A confirmation that the ICES rules for acknowledging the data source will be observed

Completing the form will result in a request to the relevant survey contact person of each institute involved with that survey/area combination and this person will be requested to reply to ICES within 14 days. For the IBTS the contact person has to reply within this time limit, if not it will be taken as acceptance of the request for data access. When after 14 days no relevant data supplier has objected, ICES will extract the requested data from the database and make them available. For BITS the contact person has to give a positive reply before the data can be provided. This is to prevent that the data are issued if the contact person has not read the e-mail.

## Output

As mentioned in previous chapters, DATRAS can deliver different types of output aimed at different types of users and/or user requests. These types of output are:

- Public data
- Aggregated data
- Raw data

Here we will present the output that DATRAS can generate. Depending on access rights (see chapter 5) this output can be retrieved through the web.

### 6.1 Public data

The public data (table 6.1) consist of a table with indices per year per species, a graph showing the time trend of the indices (figure 6.1) and for the North Sea IBTS distribution maps are provided for the eight “standard” species (a suite of species for which the survey is expected to deliver abundance estimates, often to be used for the tuning of stock assessments) with mean number per hour per statistical rectangle per age class 0, 1, 2 and 3+ per quarter over the last 4 years 2000 to 2004 (figure 6.1). The public data can be accessed at

<http://www.ices.dk/datacentre/datras/indices.asp>

Table 6.1. DATRAS output that can be accessed by the general public

Data	Description
Viewing and downloading of indices	Index per “standard” species. Index time series can be viewed and downloaded. The time series is selected by survey, start year, end year, quarter and species.
Distributions maps	Mean CPUE per age 0 to 3+ for all “standard” species by ICES statistical rectangle.

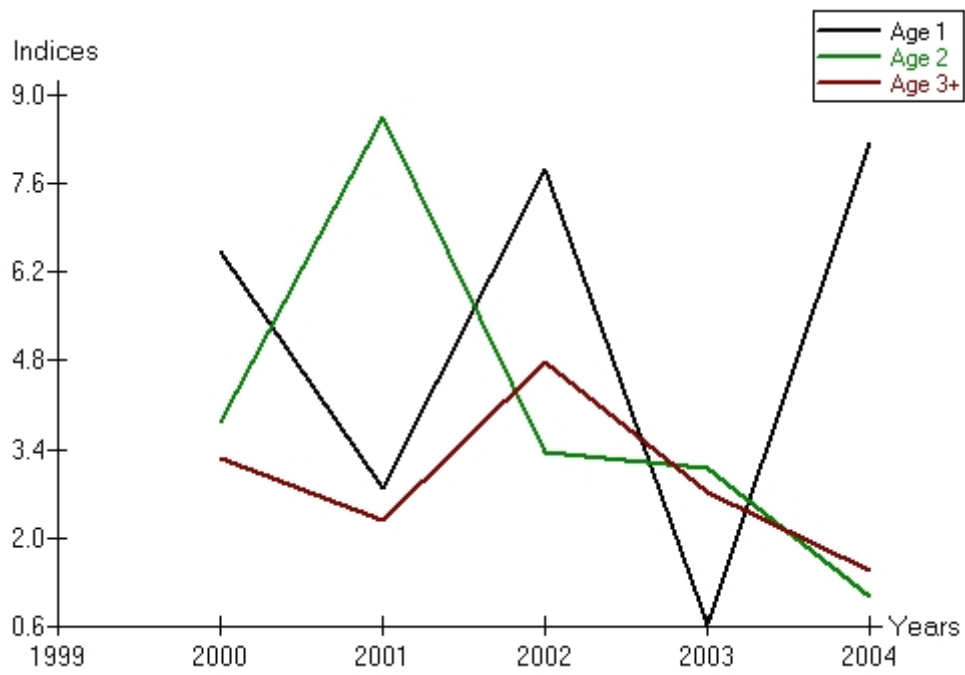


Figure 6.1. The time trend for cod indices found at the public DATRAS web page

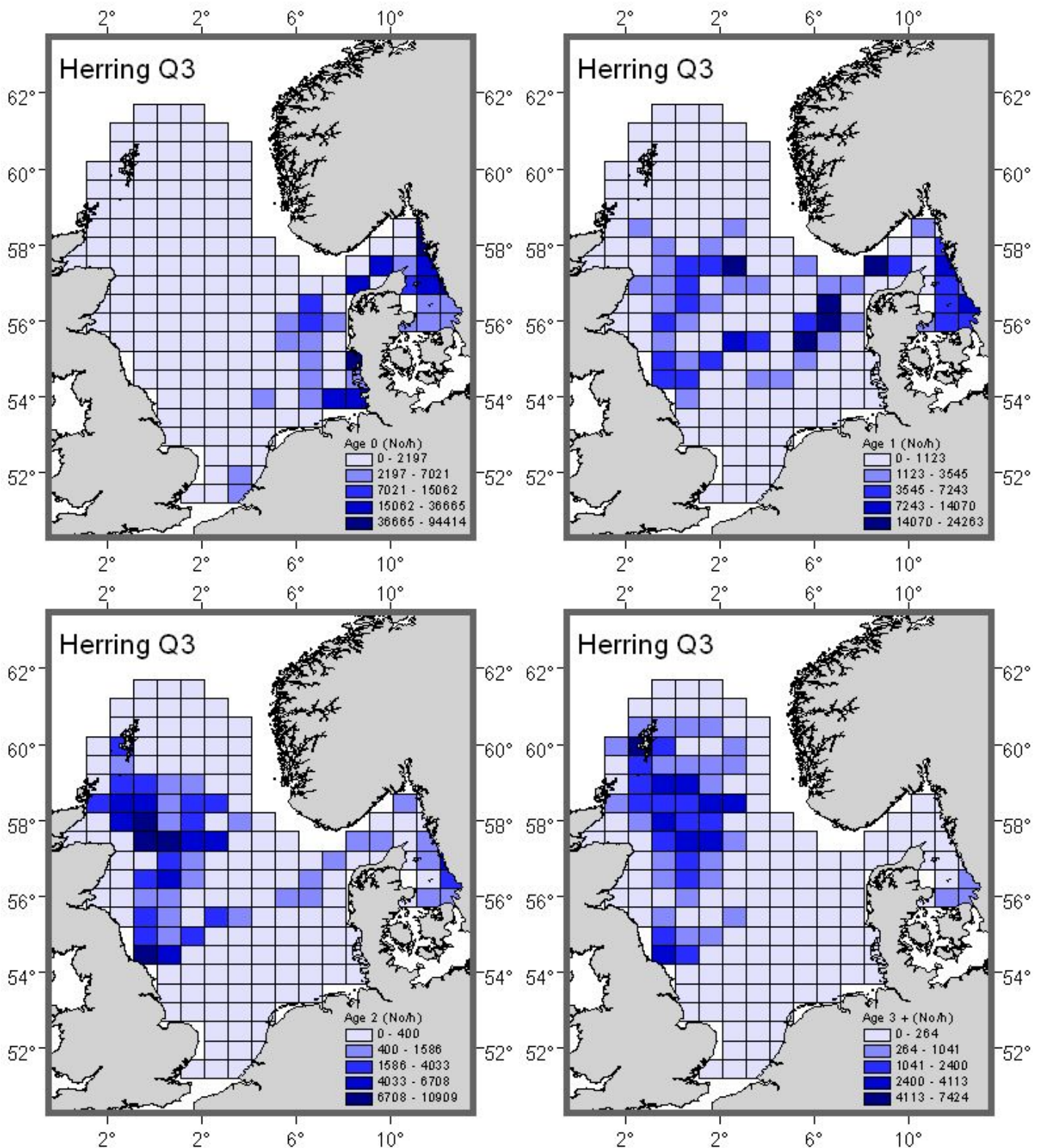


Figure 6.2. An example of the distribution maps. Quarter 3 catches of herring ages 0, 1, 2 and 3+ as a mean over the years 2000 to 2004

## 6.2 Aggregated data

The aggregated data can be downloaded from ICES through the DATRAS web page under restricted data (<http://www.ices.dk/datacentre/datras/restricted.asp>). The pivot tables are not yet

operational but ICES will work on a future solution. At present data can be downloaded as CSV files (table 6.2).

Table 6.2. Aggregated DATRAS output

Data	Description
Pivot tables: CPUE per length by sub area CPUE per age and length by sub area	Two pivot tables for downloading mean CPUE data by the aggregation levels described under the data warehouse section (Chapter 4.4.3 and figure 4.7).
CSV files: CPUE by age and length by sub area CPUE by length by sub area	Time series can be downloaded as CSV files. The time series is selected by survey, start year, end year, quarter, and species.
SMALK data: (CSV) - Raw ALK - Final used ALK - Raw SMALK	The “Final used ALK” is the “Raw ALK” where missing length classes are substituted. “Raw SMALK” also contains information on Sex and Maturity. Data can be obtained for one year by survey, year, quarter and species

### 6.3 Raw data

Two types of format are distinguished for the raw data: the original exchange format and catch-at-length or catch-at-age in numbers per hour per haul (Table 6.3).

Table 6.3. Raw data

Data	Description	Placed
Exchange format (CSV)	The exchange format can be downloaded as time series by indicating survey, quarter, start year and end year.	<a href="http://www.ices.dk/datacentre/datras/restricted.asp">http://www.ices.dk/datacentre/datras/restricted.asp</a>
Catch-at-length or catch-at-age in numbers per hour per haul	Catch files can be downloaded as time series by indicating survey, quarter, species, start year and end year	<a href="http://www.ices.dk/datacentre/datras/restricted.asp">http://www.ices.dk/datacentre/datras/restricted.asp</a>

### 6.4 ‘Health’ Warning

Whilst the data have been extensively checked by both the supplying institutes and ICES there are inherent flaws in gathering the information. For example:

- In spite of the fact that now all bottom trawl surveys are in the same database and often at least partially overlap in time period or area covered, there are differences between the surveys that do not allow the combination of data from different surveys in one analysis unless this point is specifically addressed.
- There is no guarantee that the gears deployed adequately sample all species (e.g. catchability of flatfish in beam trawls is markedly higher than that of GOV). It should be realized that each survey gear catches a gear-dependent subset of the fish community
- A common property of all surveys is that over time species determination skills improve. As a result new taxa will emerge in the dataset that were previously reported as part of another closely-related taxon.



- Over time the sampling procedures, gear characteristics, timing of the survey or the area covered may have changed thereby influencing the catches. In chapter 3.5 (Expert advice) recommendations are given in order to prevent bias in the analyses using DATRAS catch data, caused by such factors.

Thus users are urged to treat the data with caution. If the user has any queries on the validity of the data or the conclusions to be drawn they should contact either the ICES Secretariat or the Chairs of the relevant Working Group (IBTS, BITS or BEAM)

## 7 References

- Anon. 2000. Manual for the Baltic International Demersal Trawl Surveys, Version 3.0
- Anon. 2002. Manual for the International Bottom Trawl Surveys in the western and southern areas, Revision II
- Anon. 2004. Manual for the International Bottom Trawl Surveys, Revision VII
- ICES 1979. Report of the International Gadoid Survey Working group. ICES C.M. 1979/G:35
- ICES 1981. Report of the joint meeting of the International Young Herring Survey Working Group. ICES C.M. 1981/H:10.
- ICES 1983. Report of the International Gadoid Survey Working group. ICES C.M. 1983/G:62.
- ICES 1997. Report of the Workshop on Standard Trawls for Baltic International Fish Surveys, Gdynia, 1997. ICES C.M. 1997/J:6
- ICES 2002. Report of the Working Group on beam trawl surveys. ICES C.M. 2003/G:07
- IPOSTS. 2001. International Program of Standardized Trawl Surveys, Final Report. EU Project No. 98/057
- ISDBITS. 2001. Improvement of Stock Assessment and Data Collection by Continuation, Standardisation and Design Improvement of the Baltic International Bottom Trawl Survey for Fishery Resource Assessment, Final Report. EU Project No. 98/099
- SESITS 1999. Evaluation of Demersal Resources of South-Western Europe from standardised groundfish surveys, Final report. EU Project No. 96/029

## Appendix I: Exchange format

This table shows the exchange format for North Sea IBTS data as provided on the web <http://www.ices.dk/datacentre/datsu/selrep.asp>.

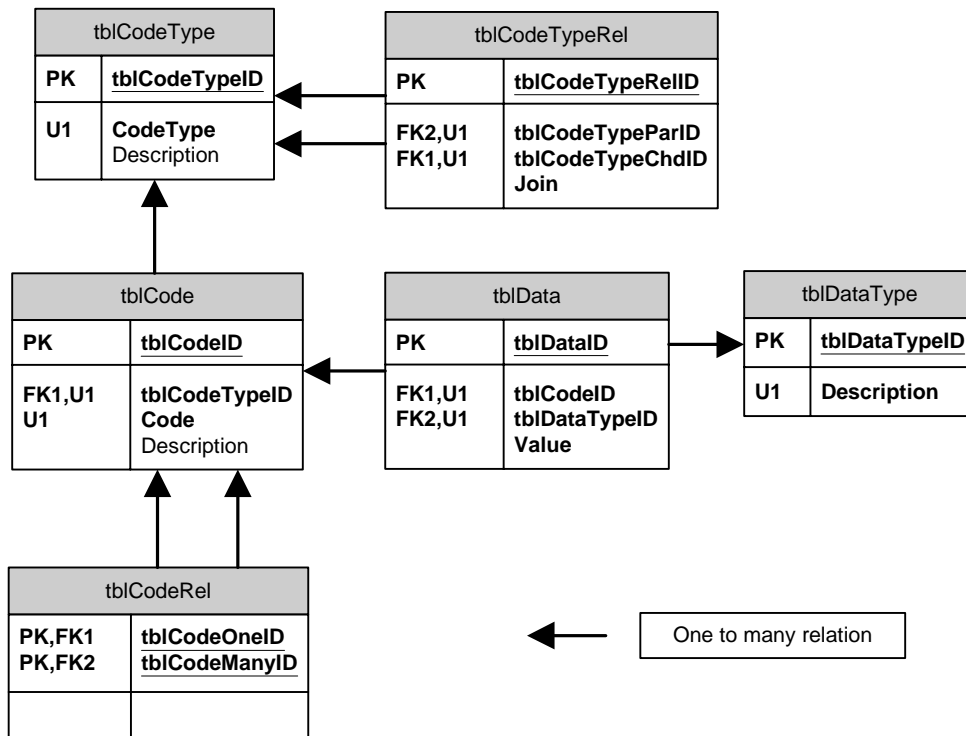
### North Sea International Bottom Trawl Survey - before 2004 (NS-IBTS)

Start/Order	Field Name	Width	Mandatory	Data Type	Code List
<b>HH Haul Information</b>					
1	RecordType	2	✓	char	
2	Quarter	1	✓	int	
3	Country	3	✓	char	<a href="#">TS_Country</a>
4	Ship	4	✓	char	<a href="#">TS_Ship</a>
5	Gear	6	✓	char	<a href="#">Gear</a>
6	SweepLngt	3		int	
7	GearExp	2		char	<a href="#">TS_GearExp</a>
8	DoorType	2		char	<a href="#">TS_DoorType</a>
9	StNo	6		char	
10	HaulNo	3	✓	int	
11	Year	4	✓	char	
12	Month	2	✓	int	
13	Day	2	✓	int	
14	TimeShot	4	✓	char	
15	Stratum	4		char	<i>Not used in this format</i>
16	HaulDur	3	✓	int	
17	DayNight	2	✓	char	<a href="#">TS_DayNight</a>
18	ShootLat	8	✓	decimal4	
19	ShootLong	9	✓	decimal4	
20	HaulLat	8		decimal4	
21	HaulLong	9		decimal4	
22	StatRec	4		char	
23	Depth	4	✓	int	
24	HaulVal	1	✓	char	<a href="#">TS_HaulVal</a>
25	HydroStNo	8	✓	char	
26	StdSpecRecCode	1	✓	char	<a href="#">TS_StdSpecRecCode</a>
27	BycSpecRecCode	1	✓	char	<a href="#">TS_BycSpecRecCode</a>
28	DataType	2	✓	char	<a href="#">TS_DataType</a>
29	Netopening	4		decimal1	
30	Rigging	2		char	<i>Not used in this format</i>
31	Tickler	2		int	<i>Not used in this format</i>
32	Distance	4		int	
33	Warplngt	4		int	
34	Warpdia	2		int	
35	WarpDen	2		int	
36	DoorSurface	4		decimal1	
37	DoorWgt	4		int	
38	DoorSpread	3		int	
39	WingSpread	2		int	
40	Buoyancy	4		int	
41	KiteDim	3		decimal1	
42	WgtGroundRope	4		int	
43	TowDir	3		int	

44	GroundSpeed	3		decimal1	
45	SpeedWater	3		decimal1	
46	SurCurDir	3		int	
47	SurCurSpeed	4		decimal1	
48	BotCurDir	3		int	
49	BotCurSpeed	3		decimal1	
50	WindDir	3		int	
51	WindSpeed	3		int	
52	SwellDir	3		int	
53	SwellHeight	4		decimal1	
54	SurTemp	4		decimal1	
55	BotTemp	4		decimal1	
56	SurSal	5		decimal2	
57	BotSal	5		decimal2	
58	ThermoCline	2		char	<a href="#">TS_ThermoCline</a>
59	ThClineDepth	4		int	
<b>HL Length frequency distribution</b>					
1	RecordType	2	✓	char	
2	Quarter	1	✓	int	
3	Country	3	✓	char	<a href="#">TS_Country</a>
4	Ship	4	✓	char	<a href="#">TS_Ship</a>
5	Gear	6	✓	char	<a href="#">Gear</a>
6	SweepLngt	3		int	
7	GearExp	2		char	<a href="#">TS_GearExp</a>
8	DoorType	2		char	<a href="#">TS_DoorType</a>
9	StNo	6		char	
10	HaulNo	3	✓	int	
11	Year	4	✓	char	
12	SpecCodeType	1	✓	char	<a href="#">TS_SpecCodeType</a>
13	SpecCode	10	✓	char	
14	SpecVal	2	✓	char	<a href="#">TS_SpecVal</a>
15	Sex	2		char	<a href="#">TS_Sex</a>
16	TotalNo	7		decimal2	
17	CatIdentifier	2	✓	int	<a href="#">TS_CatIdentifier</a>
18	NoMeas	3	✓	int	
19	SubFactor	9	✓	decimal4	
20	SubWgt	5		int	
21	CatCatchWgt	8	✓	int	
22	LngtCode	2	✓	char	<a href="#">TS_LngtCode</a>
23	LngtClass	3	✓	int	
24	HLNoAtLngt	6	✓	int	
<b>CA SMALK</b>					
1	RecordType	2	✓	char	
2	Quarter	1	✓	int	
3	Country	3	✓	char	<a href="#">TS_Country</a>
4	Ship	4	✓	char	<a href="#">TS_Ship</a>
5	Gear	6	✓	char	<a href="#">Gear</a>
6	SweepLngt	3		int	
7	GearExp	2		char	<a href="#">TS_GearExp</a>
8	DoorType	2		char	<a href="#">TS_DoorType</a>
9	StNo	6		char	
10	HaulNo	3		int	

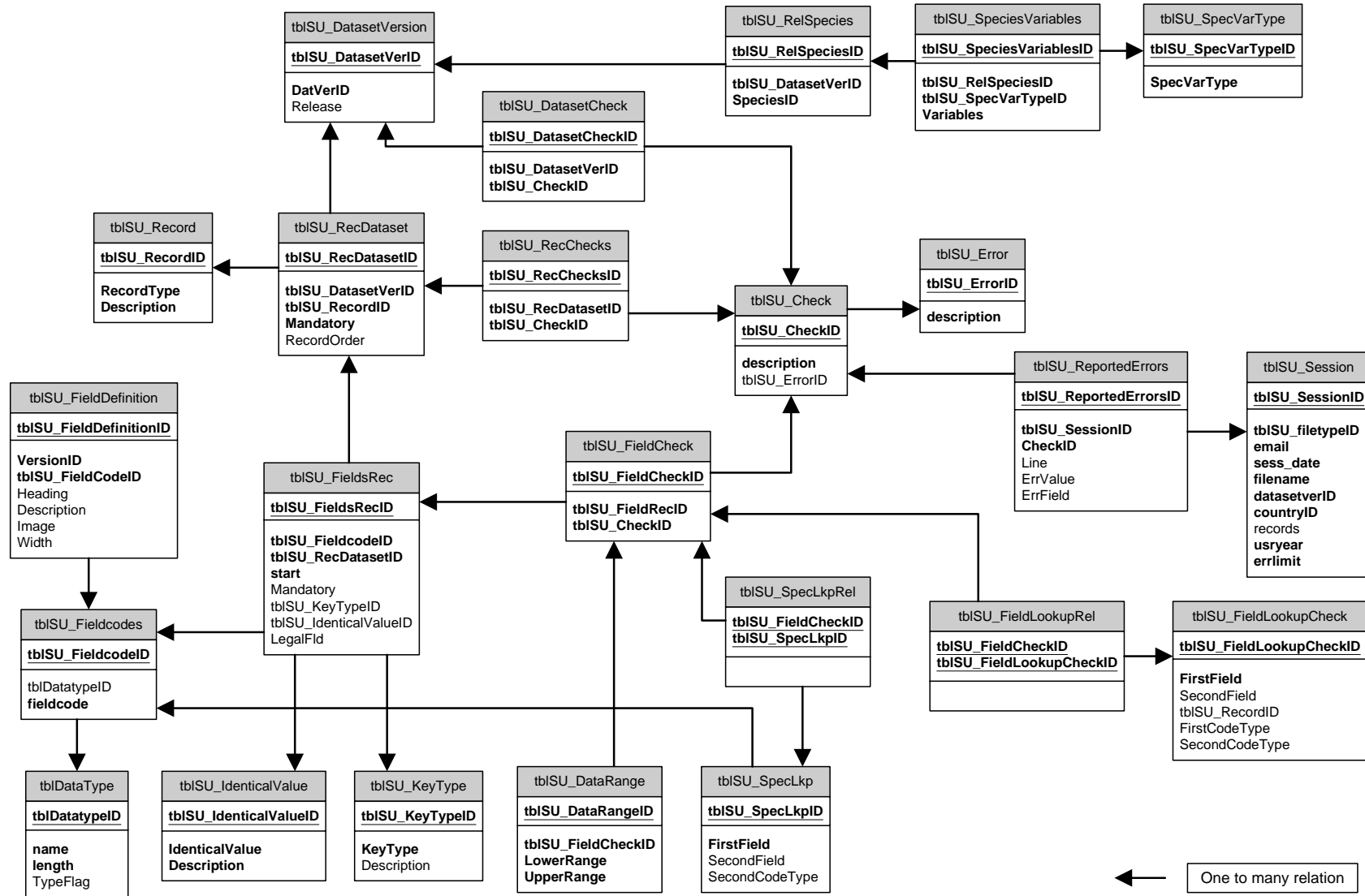
11	Year	4	✓	char	
12	SpecCodeType	1	✓	char	<a href="#">TS_SpecCodeType</a>
13	SpecCode	10	✓	char	
14	AreaType	2	✓	char	<a href="#">TS_AreaType</a>
15	AreaCode	4	✓	char	
16	LngtCode	2	✓	char	<a href="#">TS_LngtCode</a>
17	LngtClass	3	✓	int	
18	Sex	2	✓	char	<a href="#">TS_Sex</a>
19	Maturity	2	✓	char	<a href="#">TS_Maturity</a>
20	PlusGr	2		char	<a href="#">TS_PlusGr</a>
21	age	2	✓	int	
22	CANoAtLngt	3	✓	int	
23	IndWgt	5		int	

### Appendix II: RECO database diagram

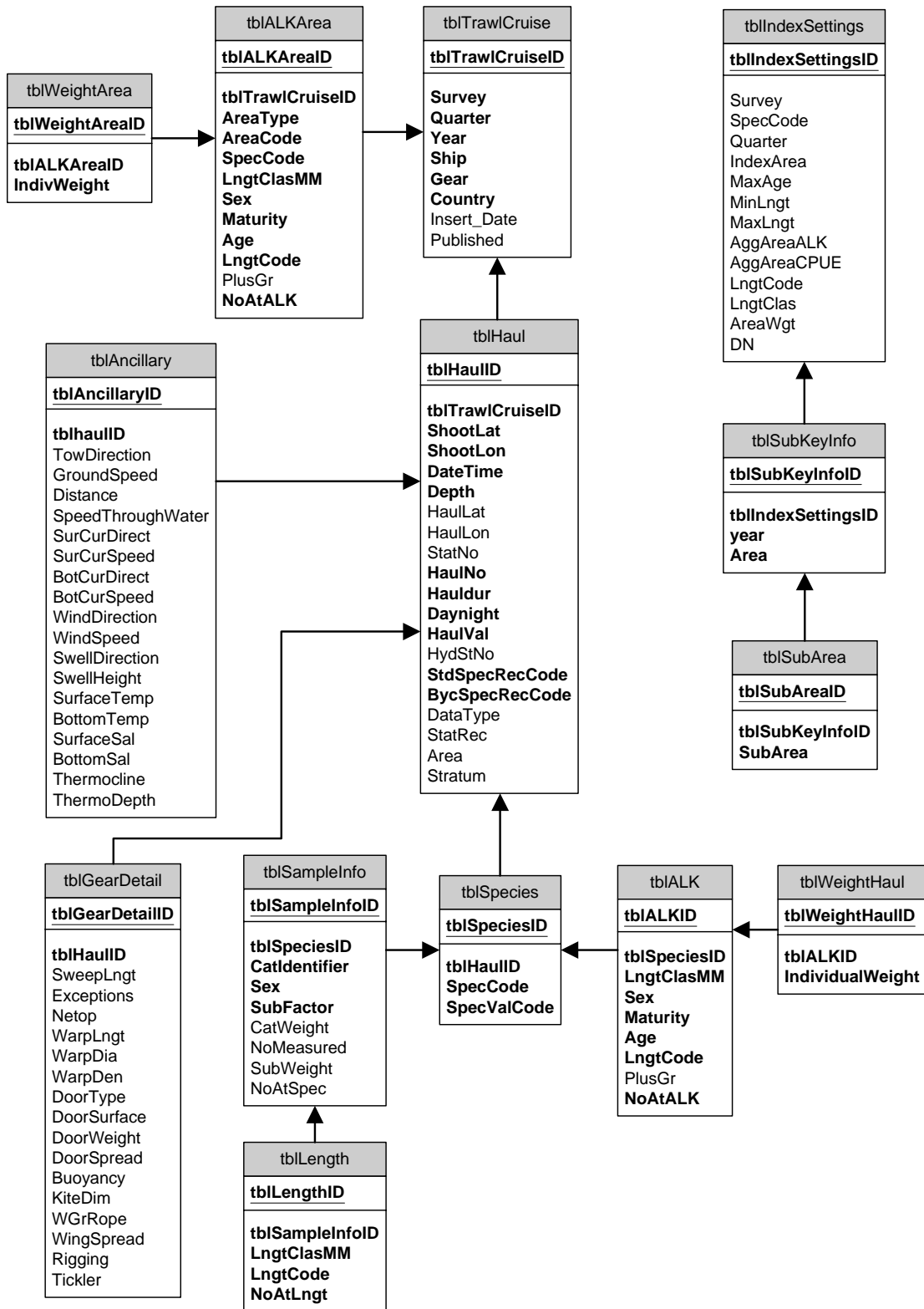




### Appendix III: DATSU database diagram



# Appendiks IV: DATRAS database diagram

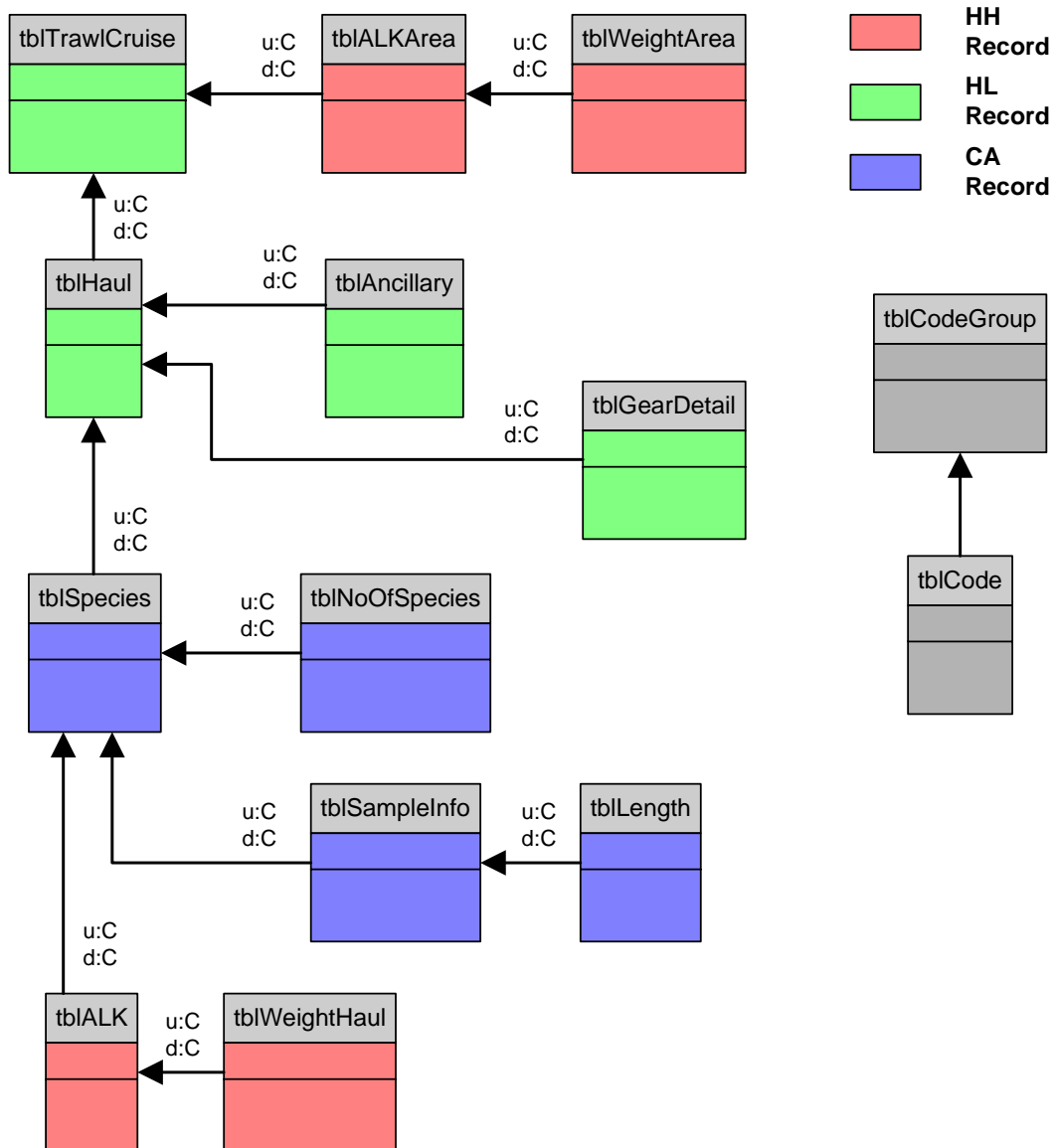


## Appendix V: Data import

The data import module in DATRAS accept data only in the exchange format as shown in appendix I (or see web page <http://www.ices.dk/datacentre/datsu/selrep.asp>) and passes the data to the tables in the DATRAS database.

### Diagram of Records

Below is a schematic of the tables in the DATRAS database, which includes their relationships.




Each of these tables and the data transformations needed from the data files will be listed in sequence.

### TblTrawlCruise

Database Field	Database type	Format Field	Location (Rec/Posn)	Notes
----------------	---------------	--------------	---------------------	-------



1	tblTrawlCruiseID	int	N/A	N/A	Created by DataBase
2	Survey	int	N/A	N/A	User defined at time of import <a href="#">RECO Lkup</a> <sup>1</sup>
3	Quarter	tinyint	Quarter	HH2	
4	Year	smallint	Year	HH11	
5	Ship	int	Ship	HH4	 <a href="#">CO Lkup</a>
6	Gear	int	Gear	HH5	<a href="#">RECO Lkup</a>
7	Country	int	Country	HH3	<a href="#">RECO Lkup</a>

### TblALKArea

	DataBase Field	Database type	Format Field	Location (Rec/Posn)	Notes
1	tblALKAreaID	int	N/A	N/A	Created by DataBase
2	tblTrawlCruiseID	int	N/A	N/A	Created by DataBase
3	AreaType	int	AreaType	CA14	<a href="#">RECO Lkup</a>
4	AreaCode	int	AreaCode	CA15	<a href="#">RECO Lkup</a> , used in Conjunction with CA45 to determine AreaType
5	SpecCode	int	SpecCode	CA13	<a href="#">RECO Lkup</a> , used in conjunction with CA12 to determine <a href="#">NODC/TSN</a> <sup>2</sup> type code
6	LngtClas	tinyint	MinLengthClass	CA17	
7	Sex	int	Sex	CA18	<a href="#">RECO Lkup</a>
8	Maturity	int	Maturity	CA19	<a href="#">RECO Lkup</a>
9	Age	smallint	Age	CA21	

<sup>1</sup> This list will be a lookup in the tblCode in the DATRAS database. If codes are not found in tblCode the lookup will be in tblCode in RECO. If the code is not found data are not imported and a message is given in the log.

<sup>2</sup> The NODC reported in data is actually erroneous in that it contains trailing zeroes, which are an ICES/WG invention. The actual codes have no trailing zeroes. To ensure we can map to ITIS NODC codes we will remove the trailing zeroes at import and convert to TSN serial numbers.

10	LngtCode	int	LngtCode	CA16	<a href="#">RECO Lkup</a>
11	PlusGr	int	PlusGr	CA20	<a href="#">RECO Lkup</a>
12	NoAtALK	tinyint	CANumberAtLength	CA22	Summed for each unique key in record

### TblWeightArea

	DataBase Field	Database type	Format Field	Location (Rec/Posn)	Notes
1	tblWeightAreaID	int	N/A	N/A	Created by DataBase
2	tblALKAreaID	int	N/A	N/A	Created by DataBase
3	IndivWeight	int	IndividualWeight	CA23	Each one recorded

### TblHaul

	DataBase Field	Database type	Format Field	Location (Rec/Posn)	Notes
1	tblHaulID	int	N/A	N/A	Created by DataBase
2	tblTrawlCruiseID	int	N/A	N/A	Created by DataBase
3	ShootLat	dec (9,4)	ShootingLatitude	HH18	
4	ShootLon	dec (9,4)	ShootingLongitude	HH21	
5	DateTime	smalldatetime	Year + Month + Day + TimeShot	HH11 HH12 HH13 HH14	+ + + Concatenate and Cast.
6	Depth	smallint	Depth	HH23	
7	HaulLat	dec (9,4)	HaulingLatitude	HH20	
8	HaulLon	dec (9,4)	HaulingLongitude	HH21	
9	StatNo	char	StdStationNumber	HH9	
10	HaulNo	smallint	HaulNo	HH10	
11	Hauldur	smallint	HaulDuration	HH11	
12	Daynight	int	DayNight	HH17	<a href="#">RECO Lkup</a>
13	HaulVal	int	HaulValidity	HH24	<a href="#">RECO Lkup</a>
14	HydStNo	char	HydrographicStationNu	HH25	
15	StdSpecRecCode	int	StandardSpecRecCode	HH26	<a href="#">RECO Lkup</a>

16	BycSpecRecCode	int	BycSpecRecCode	HH27	<a href="#">RECO Lkup</a>
17	DataType	int	DataTypes	HH28	<a href="#">RECO Lkup</a>
18	StatRec	char OR calculated column	N/A	N/A	fn_Statrec(Lon,Lat) returns StatRec Code
19	Area	char	N/A	N/A	Calculated by stored proc. at import
20	Stratum	int	Stratum	HH15	<a href="#">RECO Lkup</a> or iscalculated by stored proc. at import, depending on dataset

### TblGearDetail

	DataBase Field	Database type	Format Field	Location (Rec/Posn)	Notes
1	tblGearDetailID	int	N/A	N/A	Created by DataBase
2	tblHaulID	int	N/A	N/A	Created by DataBase
3	SweepLngt	smallint	SweepLength	HH6	
4	Exceptions	int	GearExp	HH7	<a href="#">RECO Lkup</a>
5	Netop	dec (9,2)	NetOpening	HH29	
6	WarpLngt	smallint	WarpLength	HH33	
7	WarpDia	tinyint	WarpDiameter	HH34	
8	WarpDen	tinyint	WarpDensity	HH35	
9	DoorType	int	DoorType	HH8	<a href="#">RECO Lkup</a>
10	DoorSurface	dec (9,2)	Door Surface	HH36	
11	DoorWeight	smallint	DoorWeight	HH37	
12	DoorSpread	smallint	DoorSpread	HH38	
13	Buoyancy	smallint	Buoyancy	HH40	
14	KiteDim	dec (9,2)	KiteDimensions	HH41	
15	WGrRope	smallint	WeightGroundRope	HH42	
16	WingSpread	smallint	WingSpread	HH39	
17	Rigging	int	Rigging	HH30	<a href="#">RECO Lkup</a>

18	Tickler	tinyint	Tickler	HH31	
----	---------	---------	---------	------	--

### TblAncillary

	DataBase Field	Database type	Format Field	Location (Rec/Posn)	Notes
1	tblAncillaryID	int	N/A	N/A	Created by DataBase
2	tblhaulID	int	N/A	N/A	Created by DataBase
3	TowDirection	smallint	TowingDirection	HH43	
4	GroundSpeed	dec (9,2)	GroundSpeed	HH44	
5	Distance	smallint	Distance	HH32	
6	SpeedThroughWater	dec (9,2)	SpeedThroughWater	HH45	
7	SurCurDirect	smallint	SurfaceCurrentDirect	HH46	(0-360 Deg)
8	SurCurSpeed	dec (9,2)	SurfaceCurrentSpeed	HH47	
9	BotCurDirect	smallint	BottomCurrentDirect	HH48	(0-360 Deg)
10	BotCurSpeed	dec (9,2)	BottomCurrentSpeed	HH49	
11	WindDirection	smallint	WindDirection	HH50	
12	WindSpeed	smallint	WindSpeed	HH51	
13	SwellDirection	smallint	SwellDirection	HH52	
14	SwellHeight	dec (9,2)	SwellHeight	HH53	
15	SurfaceTemp	smallint	SurfaceTemp	HH54	
16	BottomTemp	smallint	BottomTemp	HH55	
17	SurfaceSal	smallint	SurfaceSalinity	HH56	
18	BottomSal	smallint	BottomSalinity	HH57	
19	Thermocline	char(2)	Thermocline	HH58	
20	ThermoDepth	smallint	ThClineDepth	HH59	

### TblSpecies

	DataBase Field	Database type	Format Field	Location (Rec/Posn)	Notes
1	tblSpeciesID	int	N/A	N/A	Created by DataBase
2	tblHaulID	int	N/A	N/A	Created by DataBase
3	SpecCode	int	SpecCode	HL13	<a href="#">RECO Lkup</a> , used in conjunction with HL12 to determine <a href="#">NODC/TSN</a> type code
4	SpecValCode	int	SpeciesValidity	HL14	<a href="#">RECO Lkup</a>

### TblNoOfSpecies

	DataBase Field	Database type	Format Field	Location (Rec/Posn)	Notes
1	tblNoOfSpeciesID	int	N/A	N/A	Created by DataBase

2	NoAtSpec	int	TotalNo	HL49	
3	tblSpeciesID	int	N/A	N/A	Created by DataBase

### TblALK

	DataBase Field	Database type	Format Field	Location (Rec/Posn)	Notes
1	tblALKID	int	N/A	N/A	Created by DataBase
2	tblSpeciesID	int	N/A	N/A	Created by DataBase
3	LngtClas	tinyint	MinLengthClass	CA17	
4	Sex	int	Sex	CA18	<a href="#">RECO Lkup</a>
5	Maturity	int	Maturity	CA19	<a href="#">RECO Lkup</a>
6	Age	smallint	Age	CA21	
7	LngtCode	int	LngtCode	CA16	<a href="#">RECO Lkup</a>
8	PlusGr	int	PlusGr	CA20	<a href="#">RECO Lkup</a>
9	NoAtALK	tinyint	CANumberAtLength	CA22	Summed for each unique key in record

### TblWeightHaul

	DataBase Field	Database type	Format Field	Location (Rec/Posn)	Notes
1	tblWeightHaulID	int	N/A	N/A	Created by DataBase
2	tblALKID	int	N/A	N/A	Created by DataBase
3	IndividualWeight	int	IndividualWeight	CA23	

### TblSampleInfo

	DataBase Field	Database type	Format Field	Location (Rec/Posn)	Notes
1	tblSampleInfoID	int	N/A	N/A	Created by DataBase
2	tblSpeciesID	int	N/A	N/A	Created by DataBase
3	CatIdentifier	smallint	CatIdentifier	HL17	<a href="#">RECO Lkup</a>
4	Sex	int	Sex	HL15	<a href="#">RECO Lkup</a>
5	SubFactor	dec (9,2)	SubSamplingFactor	HL19	
6	CatWeight	int	CatCatchWeight	HL21	
7	NoMeasured	smallint	NoMeasured	HL18	
8	SubWeight	int	SubSamplingWeight	HL20	

### TblLength

	DataBase Field	Database type	Format Field	Location (Rec/Posn)	Notes
1	tblLengthID	int	N/A	N/A	Created by DataBase
2	tblSampleInfoID	int	N/A	N/A	Created by DataBase
3	LngtClas	tinyint	MinLengthClass	HL23	
4	LngtCode	int	LngtCode	HL22	<a href="#">RECO Lkup</a>
5	NoAtLngt	int	HLNoAtLength	HL24	

**General:**

- If Exists(HAULNO in CA record ) Then Haul Based ALK (tblALK) Else Area Based (tblALKArea)
- It is assumed that there is only one tblTrawlCruise record per file submission.

In addition to the Core Database tables there will also be a simple Log Table that relates to all of the tables in DATRAS. This would be structured as below:

**TblLog**

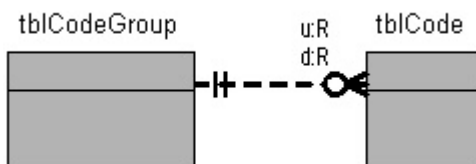
	DataBase Field	Database type	Notes
1	tblLogID	int	Created by DataBase
2	ObjName	varchar(50)	Object Name ie. tblHaul, test.csv
3	ObjAction	varchar(50)	Insert, Delete, Update, (Export, Import)
4	ObjUser	varchar(10)	sysusers.name
5	LogTime	datetime	Default Value: getdate()

**Lookup Codes**

All of the codes that are used in the database will be referenced from RECO or ITIS (Species TSN's). However, it will be necessary to include additional tables to deal with the management of the codes within the DATRAS database as it is inadvisable to link each code column directly to the outside database where the actual codes reside.

The SKY database has a similar 'problem' and so we will use this solution as then we can benefit from the existing triggers and procedures.

This will look like the diagram below:



**TblCode**

	DataBase Field	Database type	Notes
1	tblCodeID	int	Created by DataBase

2	Code	varchar(10)	Code as Imported from Data File
3	ForeignID	int	ID from External Database
4	tblCodeGroupID	int	ID from CodeGroup Table

### **TblCodeGroup**

	<b>DataBase Field</b>	<b>Database type</b>	<b>Notes</b>
1	tblCodeGroupID	int	Created by DataBase
2	CodeGroup	varchar(20)	Name of Group/Type
3	DB	varchar(20)	Name of Referenced Database ie. RECO

### **Triggers/Procedures for Code**

To maintain the code table, a number of triggers are used to ensure integrity between databases and within the code table.

#### **Datras..TR\_LocktblCode**

This is an Insert/Update trigger that prevents direct data entry into tblCode

#### **Datras..TR\_LocktblCodeGroup**

This is an Insert/Update trigger that prevents direct data entry into tblCodeGroup

#### **Datras..UP\_GetCodeID**

This is the stored procedure that manages the codes in tblCode and tblCodeGroup. It works by checking for the existence of a newly imported code in datras and then if it can't find it there, it looks to external databases to get the information (RECO and IT IS). Once it finds the code, it momentarily disables the locks on tblCode and tblCodeGroup and inserts the new information.

#### **RECO..TR\_Check\_DelCode<sup>1</sup>**

This is triggered when an attempt is made to delete a code in RECO..tblCode. A check is made on external databases (datras, sky etc.) to see if that code is in use, if it is then the delete is cancelled and an error is raised.

#### **RECO..TR\_Check\_UpdCode**

This is triggered when an attempt is made to update a code in RECO.tblCode. A check is made on external databases to see if that code exists there, if it is then the code value in these databases is updated.

#### **RECO..TR\_Check\_DelCodeType**

This is triggered when an attempt is made to delete a code in RECO..tblCodeType. A check is made on external databases (datras, sky etc.) to see if that code type is in use, if it is then the delete is cancelled and an error is raised.

#### **RECO..TR\_Check\_UpdateType**

This is triggered when an attempt is made to update a code in RECO.tblCodeType. A check is made on external databases to see if that code exists there, if it is then the code type value in these databases is updated.

---

<sup>1</sup> Because these triggers affect more than one database, to modify we use the file copy (P:\DPI\The4OfUs\reco\sql\\*.\*) and run this to update the triggers.

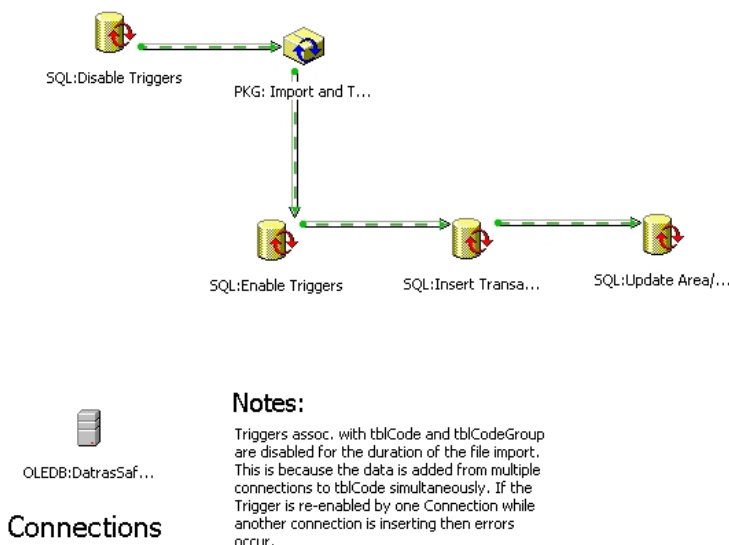
### DTS: ITIS\_Create<sup>3</sup>

When the ICES copy of the ITIS database is rebuilt (each month) from the new data downloads, a task is executed called **SQL: Cross Check**. The task queries the new tsn list against those present in other ICES databases and sends an e-mail of results to DB Managers. The results indicate if TSN codes used elsewhere in the ICES databases (ie. RECO) have been deleted from ITIS. The data administrators can then decide what action to take<sup>4</sup>.

### Data Transformation Packages:Import

#### Datras Part I (Diagram)

The green arrows on the diagram indicate “on Success” task execution order.



#### Datras Part I (Explanation)

DTS packages are made up by a series of interconnecting tasks that have specific properties and abilities depending on the type of object they are derived from. The tasks can be linked together with Success/Failure/Completion threads to make up a package.

#### Global Variables:

A package can have any number of global variables defined (values that are available to any object/task in the package). In this package, all of the global variables are set by the calling program i.e. from VBA in the DATRAS project front end.

- **User** (username performing operation)
- **Filename** (Name of the File been Imported)
- **FileLocation** (File Directory)
- **Survey** (Name of the Survey the data is being imported for)

#### OLEDB:Datras

<sup>3</sup> ITIS is not maintained here at ICES therefore, triggers are not necessary.

<sup>4</sup> In theory, TSN Codes are never deleted in ITIS as they are referred to other new TSN codes, if they become obsolete.



This is the connection object, which references the Datras database through an OLE DB type connection.

**SQL:Disable Triggers**

This is an execute SQL task, which disables the triggers on tblCode and tblCodeGroup. (see note in Diagram)

**PKG:Import and Transform**

This is an execute package task that executes *Datras Part II* (the child package).

**SQL:Enable Triggers**

This is an execute SQL task, which enables the triggers on tblCode and tblCodeGroup after running the previous tasks.

**SQL:Insert Transactions**

This is an execute SQL task that uses nested transactions to insert all of the data into the 10 tables which make up the DATRAS database.

**SQL:Update Area/StatRec**

This SQL task does two things; First it looks for Null Columns in the tblHaul.StatRec field and finds the missing Code from RECO and inserts it into DATRAS..tblCode. The Field is not 'changed' as it runs on a function. Secondly, tblHaul.Area is filled with the appropriate Area ID (RF Area for IBTS etc.) from RECO.

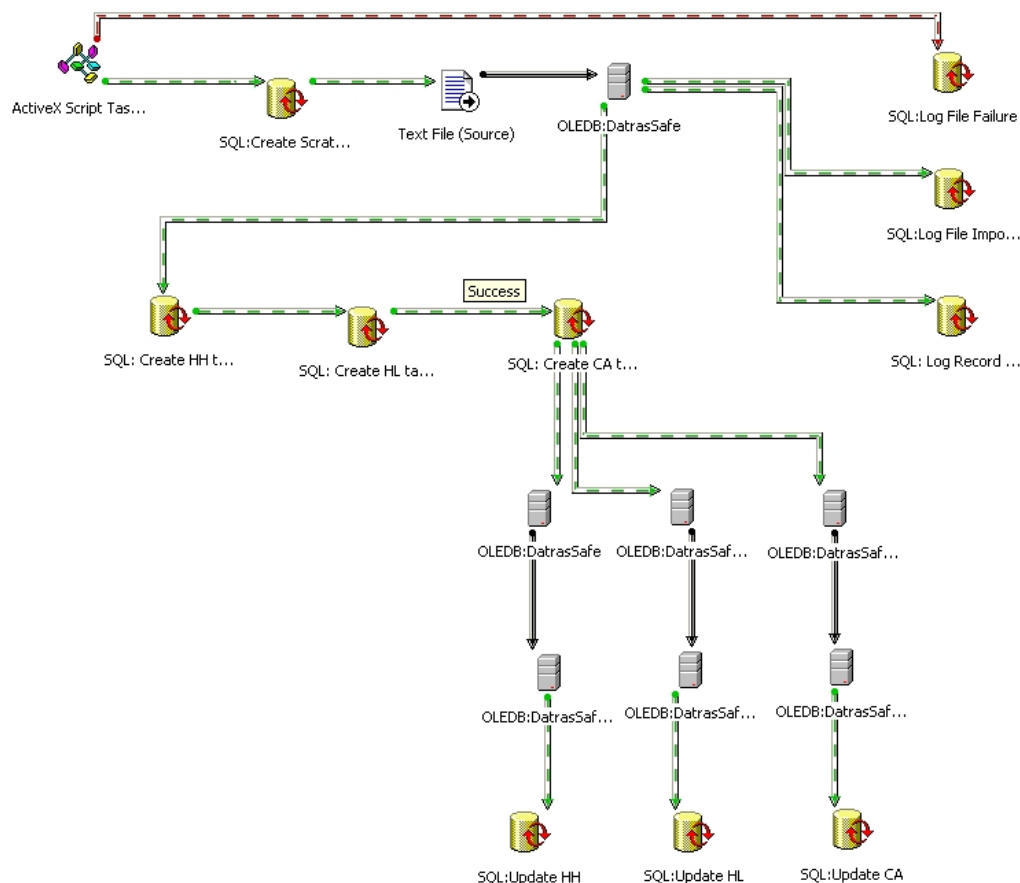
**Datras Part II (Diagram)**

The green arrows on the diagram indicate "on Success" task execution order.

The red arrows indicate "on Failure" task execution.

The dark grey arrows are data transformations and will be explained in the text.

The layout of the diagram may differ from how it is presented here, but the program flow is exactly the same.



Starting from the top left hand side, the tasks/objects will be explained, remember this package only runs from inside the Datras Package I.

### ActiveX Script task:CopyFile

This is written in VBScript and takes the information in the global variables (*filename*, *filelocation*) and uses that to copy to a new file called *datras\_file.csv*

### SQL:Log File Failure

On failure of the CopyFile script an error is written to the *tblLog* in the Datras database.

### SQL:Create Scratch table

This SQL task creates a generic table (*xtblScratch*) with 55 columns for holding the data from the file.

### Text File (Source)

This is a text file connection object, which specifies *datras\_file.csv* as the source file.

### TFM:CSV File (Grey Arrow)

This is a transformation task and using the Text source and destination object (*xtblScratch*), it maps from the CSV to the SQL table.

### OLEDB:DatrasSafe

This is the connection object for the destination node of the transformation (above).

### SQL:Log file Import

This SQL task writes a record to the *tblLog* reporting the filename etc.

### **SQL:Log Record Counts**

This SQL task writes a record to the tblLog reporting the number of records for each record type in the data file.

### **SQL:Create HH Table + HL Table + CA Table**

This is 3 linked SQL tasks which create the tables (xtblCA, xtblHH, xtblHL) for holding the 3 different record types from xtblScratch.

### **OLEDB:DatrasSafe (3 separate Connections)**

These are 3 separate connection objects (all pointing to Datras database) and mean we can process 3 transformations simultaneously.

### **TFM: HH, HL and CA (Grey Arrows)**

These transformations map the source connection (xtblScratch) to the destination connection (xtblCA, xtblHH, xtblHL)

### **SQL: Update HH, HL and CA**

These 3 separate tasks occur simultaneously and use a number of SQL statements to add the Id's from tblCode to replace the original code from the data file. This is achieved using the DTS\_GetCodeID stored procedure, which is identical to UP\_GetCodeID except that there is no disable/enable trigger clause. This is to prevent the 3 simultaneous tasks interfering with each other.

## Appendix VI: The IBTS ALK substitution procedure

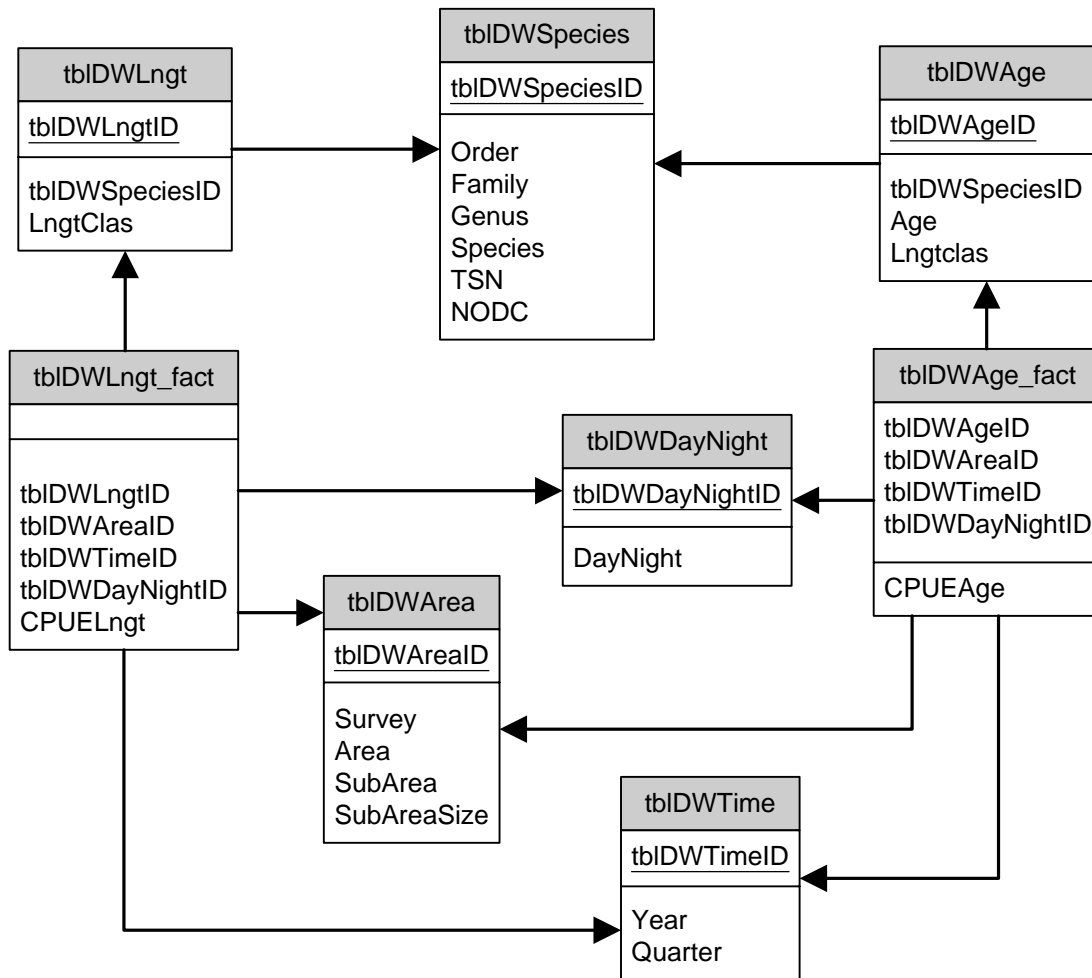
The procedure is as given below:

1. A table with no. of otoliths by species and Roundfish (RF) area is inspected to see which cells (i.e. species \* RF) are missing or has below 25 otoliths. If zero otoliths → there will always be a substitution with data from other RFs, following the procedure given below.
2. The age data table mentioned above is compared with a table with no./hr of fish caught, by species and RF. If the number of otolith for a given cell is low, but the number caught is high special care has to taken to check whether the given otolith data are sufficiently good, i.e. spread out over length and ages and extensive enough. If not a substitution is made.
3. All ALKs are inspected manually and if there are only a few age groups represented or the length range is limited a substitution is made.
4. A substitution is actually a supplementation. Age data from neighbouring RF's are used. Often only one or a few with the largest set of age data are used.
5. For saithe all age data are merged and applied to all RFs.
6. For mackerel often the situation is the same as for saithe.
7. The manual inspection of data is also a quality control procedure where peculiarities are spotted. Often individual outliers or bulks of data not consistent with the rest of data are spotted. These data are then looked into and send back to the country of origin for checking. Manual inspection of the ALK data is often a way of identifying missing data submission from a country, appearing as some RFs in the otolith table without age data.

The substitution can probably be simplified by deciding that a substitution for a given RF is made by age data from all neighbouring RFs. This will mean that:

RF 1	will be supplemented by data from RF 2, 3
RF2	“ RF 1, 3, 4, 6, 7
RF3	“ RF 1, 2, 4
RF4	“ RF 2, 3, 5, 6
RF5	“ RF 4, 6
RF6	“ RF 2, 4, 5, 7
RF7	“ RF 2, 6
RF8	“ RF 7, 9
RF9	“ RF 8.

## Appendix VII: Data warehouse database diagram



### Single tables in DW DATRAS

tbICHA
tblHaulID tblSpeciesID Survey SurveyDes Year Quarter Ship ShipDes Gear GearDes HaulNo ShootLat ShootLon DateTime Depth Area SubArea DayNight SpecCode Species Age_0 Age_1 Age_2 Age_3 Age_4 Age_5 Age_6 Age_7 Age_8 Age_9 Age_10

tbICHL
tblHaulID tblSpeciesID Survey SurveyDes Year Quarter Ship ShipDes Gear GearDes HaulNo ShootLat ShootLon DateTime Depth Area Subarea DayNight SpecCode Species LngtClas CPUE

tbIDWMeanCPUEAge
Survey DayNight Year Quarter Area SubArea SpecCode LngtCode LngtClas Age_0 Age_1 Age_2 Age_3 Age_4 Age_5 Age_6

tbIDWIndices
Survey Year Quarter SpecCode IndexArea
Genus Family PlusGr Age_0 Age_1 Age_2 Age_3 Age_4 Age_5 Age_6 Age_7 Age_8 Age_9 Age_10

tbIDWMeanCPUELngt
Survey DayNight Year Quarter Area SubArea SpecCode LngtCode LngtClas CPUE