ICES WKTQD Report 2007

ICES Resource Management Committee
ICES CM 2007/RMC:10
REF. LRC

Report of the Workshop on Taxonomic Quality Issues in the DATRAS Database (WKTQD)

23–25 January 2007

ICES, Copenhagen

ICES
International Council for the Exploration of the Sea
CIEM
Conseil International pour l'Exploration de la Mer

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H. C. Andersens Boulevard 44–46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

Recommended format for purposes of citation:

ICES. 2007. Report of the Workshop on Taxonomic Quality Issues in the DATRAS Database (WKTQD), 23-25 January 2007, ICES, Copenhagen. ICES CM 2007/RMC:10. 45 pp.
For permission to reproduce material from this publication, please apply to the General Secretary.

The document is a report of an Expert Group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the views of the Council.

© 2007 International Council for the Exploration of the Sea

Contents

Executive summary

Quality control of data collections is an issue of primary importance in conducting science and this is no different for surveys intended to monitor changes in fish abundance. Essentially, quality control is the responsibility of the national institutes conducting the surveys. However, if the national data are combined in a common international data base, as is the case with DATRAS, the consistency of the data submitted engages an international dimension, because the reliability of any comprehensive analysis of changes in the fish community at large that could serve as the basis for ICES advice depends on the reliability of the species identifications in all national subsets.

Many of the surveys that are routinely carried out jointly under the auspices of ICES have a long history during which the primary objectives have changed. Thus, the IBTS started as a Young Herring Survey, was then transformed in a Young Fish Survey to obtain recruitment estimates of commercial species, and became only a general monitoring survey of the entire fish community at a later stage. Despite these changes in general objectives, the emphasis in data use is still largely focused on the commercial species that are relatively easy to identify. Although measurement errors and punching errors for this group may have entered the data base, as evident from unrealistically small or large individuals reported in some cases, it seems generally safe to conclude that, in view of the large amount of detailed data collected for these species, these could lead only to minor and negligible distortions in the analyses. However, when it comes to the less common species, studies of the IBTS component in the past have proven major inconsistencies in species identification in the data set that has been entered in DATRAS. This problem is not restricted to a single or a few countries, but affects all countries, although the species involved may differ. This suggests that it is a direct consequence of the large number of people involved in data collection on board and of a generic lack of good taxonomic knowledge among the scientific staff at large that inhibits the maintenance of enough quality control.

The problems identified in the past have been the direct reason for holding this one-off workshop to discuss the various aspects of identifying inconsistencies and correcting species identifications in historic data sets and of ensuring correct species identification in future data collections. Taxonomic quality control is a complex issue, because the problems vary by region depending on the species that may be encountered and therefore may require specific approaches regionally. Moreover, the ultimate responsibility for introducing specific protocols for quality control rests with the survey working groups responsible for data collection rather than that the appropriate procedures can be prescribed by others. Therefore, the aim of the workshop has been to provide generic guidelines for development of suitable protocols by the survey working groups rather than to come up with a final answer.

It must be emphasized that so far progress in identifying inconsistencies in reporting of various taxa has been restricted to the IBTS component of DATRAS, which covers the North Sea, Skagerak and Kattegat. For all other surveys, similar analyses have not yet been conducted, but there is no reason to assume that the situation would be any different.

Following the Terms of Reference, the report is split in four sections that deal with each of these respectively.

Section 3 deals with ToR a): “Identify and correct taxonomic mis-identifications and input errors in DATRAS”. Obviously, this ultimate goal was beyond reach during a three-day workshop, and a lot more work needs to be done. This section lists dubious species, inconsistent information provided regarding taxonomic level reported, maximum attainable size and area of distribution, and examples of inconsistent information reported for some problematic taxa. The information given is restricted to the IBTS component of DATRAS, but should not be interpreted as a comprehensive analysis of all inconsistencies that may be

present in this data set. Rather, it highlights methodical aspects as to how inconsistencies may be elucidated.

Section 4 deals with ToR b): ”Development of protocols for ensuring the appropriate treatment of data reported at higher taxonomic levels”. Historically, different countries have reported variously species at different taxonomic levels (genus or families). Also, uncertain species records may have to be adjusted by using a higher taxonomic level. As a consequence, subsequent community analyses may require that these higher taxa are split into its constituting species based on identifications considered reliable. Because there are various ways to do such computations depending on the assumptions made, different analyses could give different answers. From an ICES perspective, some consistency in the approaches used by different working groups would seem appropriate. This section provides the essentials of an appropriate algorithm based on using length frequencies, area of distribution and year of catch, that could serve as a first guideline for comprehensive community analyses as well as for trend analysis of individual species.

Section 5 deals with ToR c): “Develop improved protocols to ensure that species identification in trawl surveys is appropriate for fish community studies, including the development of photo-ID keys for nations participating in surveys”. Various initiatives have been taken by individual countries to develop appropriate tools for species identification, including training courses. This section lists a number of ways by which future data collections may be improved or by which species identification can be ascertained at a later stage in the process.

Finally, section 6 deals with ToR d). “Develop protocols for (i) improving quality control during the submission of data to DATRAS and (ii) the future checking and quality assurance of DATRAS data. This ToR relates to the important aspect of the responsibility of ICES for ensuring that all taxonomic data in DATRAS are correct or, if they are dubious but cannot be corrected, that the information provided to external users is properly identified as being of a dubious nature. The section describes a warning system that should be developed at the submission stage of new data so as to inform each country that the data submitted contain information that is inconsistent with what is known about the biology of the species and therefore should be carefully checked. It is also stressed that historic data that cannot be corrected but remain dubious are properly flagged internally and potentially adjusted when made available to external users.

Proper taxonomic quality control will remain an issue of all monitoring programmes that needs continuous attention and adjustments. We can only hope that this comprehensive description of its many aspects helps the survey working groups to make real progress in achieving an urgently required revision of the historic information provided as well as in improving future data submission..

1 Opening of the meeting

As any database, DATRAS is not free of errors and inconsistencies in the historic records. Although a problem has been known to exist for years (Daan, 2001), a new and more detailed analysis presented at the IBTSWG in 2006 proved that the problem is even more extensive and involves more species than previously thought (ICES, 2006). This analysis compared length distributions reported of specific species among countries and species compositions reported by individual countries among consecutive years, as well a comprehensive listing of species records in excess of their known maximum reported length. The apparent inconsistencies suggest that taxonomic identification errors, whether caused by practical lack of knowledge, data entry errors or subsequent processing errors (recoding), persist though recent years rather than having been resolved. In 2006, the IBTSWG put forward a recommendation to address this quality control issue of DATRAS during a workshop devoted specifically to the problem. In the same year, the WGFE came up with a similar recommendation because it experienced great difficulties in carrying out fish community analyses for the North Sea, because the multitude of erroneous species records can easily distort the conclusions derived. However, for one unfortunate reason or another, these recommendations dropped out of the agenda of the responsible committees during the following Annual Science Meeting. Although this unintended mistake was discovered only after the final set of working group meetings had been decided upon, in late November, ICES decided to add this meeting of the “Workshop on Taxonomic Quality Control issues in DATRAS data” (WKTQD) to the list halfway through December.

WKTQD met at ICES Headquarters from 23 to 25 January 2007. In view of the short period between announcement of the meeting and its actual venue, it was a pleasant surprise to see 13 participants turn up, keen to resolve this important issue.

2 Adoption of the agenda

The terms of reference were to:

Identify and correct taxonomic mis-identifications and input errors in DATRAS;

Development of protocols for ensuring the appropriate treatment of data reported at higher taxonomic levels;

Develop improved protocols to ensure that species identification in trawl surveys is appropriate for fish community studies, including the development of photo-ID keys for nations participating in surveys;

Develop protocols for (i) improving quality control during the submission of data to DATRAS and (ii) the future checking and quality assurance of DATRAS data.

The meeting was opened on January 23rd at 10.00. After three presentations during the first morning and illustrating the extent of taxonomic misidentifications among elasmobranchs and other taxa in the North Sea (IBTS) component of DATRAS, the four ToR were discussed in general terms during consecutive sessions. Although the issues are strongly connected, the solutions in terms of quality control differ quite considerably and it was seen useful to keep these issues separate.

Ad (a): relates to historic data that have already been entered into DATRAS and are available for use. The short duration of the meeting did not allow any specific analyses to find potential errors in addition to those identified in Daan (2001) and ICES (2006), nor is a workshop an appropriate place to make corrections to data bases. Rather the workshop decided to lay out the procedures required to start this process and to set a reasonable deadline for finishing the process.

Ad (b): relates to the use of existing data in DATRAS. Although the ToR refers to treatment of higher taxonomic levels, we take a somewhat wider perspective on how to deal with information that is altogether trustworthy even though it might not be possible to correct them (health warnings, flagging).

Ad (c): relates to the pre-submission phase of data (quality control on board during sampling). Although data collection is largely a national responsibility that may be governed by other issues than prescriptions by an international group of scientists, general guidelines may help to increase the awareness of the problem and point to practical solutions that may improve species ID.

Ad (d): relates to the checking carried out at the time of submission of data to DATRAS and to periodical checks of historic data in DATRAS for new inconsistencies. Although the emphasis during the workshop was put on North Sea data based on the experience of most participants, it was emphasized to consider the broader DATRAS context in this respect.

The last day was largely spent on defining and describing the various protocols and report writing. The meeting was closed on January 25th at 17.00.

3 ToR a): Identify and correct taxonomic mis-identifications and input errors in DATRAS

In scrutinizing the IBTS part of DATRAS, Daan (2001) and ter Hofstede & Daan (2006a) have reported many obvious errors and inconsistencies that could have a major influence on the interpretation of the results of various community analyses as well as on the trends in abundance and distribution maps of specific species. However, although errors may be inferred from consistency checks, the appropriate correction is problematic, because their origin remains unknown.

The usual procedure on board is that the catch is sorted by species and then the different samples are processed by two members of the scientific crew, one measuring the fish and one tallying. In the old days, all the information was recorded on paper and these records are assumed to have been archived after punching the data for further electronic processing. Obviously, any errors entering the first reporting phase because of a wrong assignment or entry in a wrong size class can never be corrected properly, unless they are identified almost immediately on board. In contrast, punching or subsequent processing errors through the use of wrong codes might be detected by comparison with the original information sheets. This type of data checking and correcting involves an elaborate process that can only be carried out at the national institutes, where the original data are stored on paper and therefore is not a suitable task for a workshop.

In recent years, data are more and more entered directly into electronic devices (electronic measuring boards and pc software for data entry) and the handwritten sheets are disappearing. Thus, any punching errors cannot be distinguished from wrong measurements or misidentifications and the potential for data checking has become less, although one might hope that the number of errors has decreased by this process, because there is one step less in the procedure.

As a consequence of the procedures on board, it will never be possible to find all errors made in data collection, but only the flagrant ones may be detected. In this context, it is important to keep in mind those incidental mistakes rather than consistent errors may only marginally affect the outcome of any analyses and this is not the main concern. Only when they can lead to major bias in the results, ICES must be concerned about making the database available to the wider public.

National laboratories are responsible for submitting reliable information that is coded consistently. ICES does check for consistent coding of all information entering the data base and these checks may be extended to cover the likelihood of catch of particular species or cases where species records largely exceed reported maximum lengths. However, ICES can not correct individual records. However, in case the reliability of specific records is mistrusted on the basis of common taxonomic and biological information, these records should be flagged in the database to alert potential users. When data are made available to external users (as defined in the open-access policy), it would be appropriate to assign a higher taxonomic level to unreliable identifications, but keeping the flag. This can only help to prevent misinterpretation of the data.

Table 3.1 provides an overview of the cumulative number of specimens caught per 1 hour fishing and the number of hauls in which they have been reported for the IBTS within DATRAS (1970-2005). The tools that have been tested so far on this set are discussed in more detail below, but the comments provided suggests

The analysis has by no means been exhaustive and more comprehensive methods might be used, for instance by integrating consistency checks in terms of length, time, and space distributions by country. However, the analysis available should serve as a good starting point for repairing the database and action should not be delayed. Also, initiatives should be taken by working groups responsible for storing survey data from other regions or survey types to carry out comprehensive checks on such sets as well, because it seems only likely that the problems identified are not restricted to the IBTS and indeed widespread.

Table 3.1. List of taxa reported in the IBTS dataset within DATRAS with numbers reported (#c) and number of positive hauls (#h) and comments on likely or potential errors.

TSN	TAXON	#C	#H	COMMENTS
159700	Lampetra	8	4	genus with 1 species
159719	Lampetra fluviatilis	96	44	
159721	Petromyzon	10	5	genus with 1 species
159722	Petromyzon marinus	95	47	
159772	Myxine glutinosa	26934	3656	
159911	Lamna nasus	4	2	
159985	Scyliorhinidae	2	1	??
160034	Galeus melastomus	192	73	
160053	Scyliorhinus	4	3	?
160065	Scyliorhinus caniculus	12399	4996	
160067	Scyliorhinus stellaris	12	5	
160181	Galeorhinus galeus	793	294	
160226	Mustelus	248	48	
160240	Mustelus asterias	1470	518	uncertain species identification
160242	Mustelus mustelus	142	57	uncertain species identification
160611	Somniosus microcephalus	4	2	
160617	Squalus acanthias	27087	5669	
160670	Etmopterus spinax	92	44	
160838	Torpedo marmorata	4	2	input error: Raja clavata - NET
160845	Rajidae	128	54	??
160846	Raja	226	106	interpretation error
564140	Leucoraja lentiginosa	3	2	not in CLOFNAM
160876	Raja radiata	91	36	invalid taxon-> Amblyraja radiata
564149	Amblyraja radiata	81433	30203	inconsistencies (> Lmax) - SCO
160880	Raja brachyura	109	63	
160883	Raja montagui	3081	1294	
564126	Dipturus batis	170	94	inconsistencies in distribution- DEN
564148	Dipturus oxyrhinchus	4	2	
564134	Lecoraja fullonica	61	36	
564128	Leucoraja circularis	59	26	
564143	Leucoraja naevus	6109	3063	inconsistencies (>Lmax) - SCO
564141	Dipturus lintea	2	1	unlikely species - FRA
160900	Leucoraja undulata	3	2	identification ? (small specimens)
160901	Raja clavata	13315	2693	inconsistencies (> Lmax)
160959	Dasyatis pastinaca	3	2	
161022	Chimaera monstrosa	840	316	
161125	Anguillidae	150	57	??
161128	Anguilla anguilla	791	327	
161341	Conger conger	14	9	
161701	Alosa	7	4	?
161708	Alosa alosa	386	65	inconsistencies among countries
161716	Alosa fallax	19282	550	
161722	Clupea harengus	82508628	248587	
161789	Sprattus sprattus	75651841	107202	
161813	Sardina pilchardus	138545	1469	
161831	Engraulis encrasicolus	190447	2374	
161994	Salmo	2	1	?
161996	Salmo salar	10	6	

TSN	TAXON	#C	#H	COMMENTS
161997	Salmo trutta	41	20	
162039	Osmerus eperlaunus	5023	360	one spatial outlier - NOR
162057	Argentinidae	37113	2784	family with 1 genus
162061	Argentina	5490	367	?
162064	Argentina silus	28699	2982	
162071	Argentina sphyraena	234670	13717	
162187	Maurolicus muelleri	146296	1899	
162368	Myctophoidei	2496	5	??
162471	Notolepis rissoi	2	1	identification error (>Lmax) - NOR
164475	Lepadogaster	2	1	identification error (>Lmax) - FRA
164482	Diplecogaster bimaculata	3	3	
164497	Lophiidae	258	237	??
164501	Lophius piscatorius	11481	5830	
164502	Lophius budegassa	10	7	
164712	Gadus morhua	1250998	173606	
164727	Pollachius virens	291685	28150	
164728	Pollachius pollachius	7247	1757	
164740	Brosme brosme	1501	688	
164744	Melanogrammus aeglefinus	20943752	260723	
164748	Rhinonemus cimbrius	66450	13221	
164751	Physcis blennoides	88	45	
164754	Trisopterus minutus	460971	25747	
164755	Trisopterus luscus	76079	5602	
164756	Trisopterus esmarki	68069324	73928	
164758	Merlengius merlangus	33397666	316509	
164760	Molva molva	10233	4717	
164761	Molva dypterygia	23	14	unlikely ?
164764	Gaidropsarus	81	35	
164765	Gaidropsaus vulgaris	1361	336	overreported ? - NET
164766	Gaidropsaus mediterraneus	20	9	identification error (>Lmax) - GFR
164768	Gaidropsarus argentatus	6	1	identification error ?
164771	Gadiculus	47	1	genus with 1 species
164772	Gadiculus argenteus	156708	3377	
164774	Micromesistius poutassou	1418738	6447	
164777	Raniceps raninus	56	30	
164779	Ciliata mustela	1505	452	
164780	Ciliata septemtrionalis	61	20	never reported by DEN, GFR, NOR, SWE
550592	Gaidropsarus macrophthalmus	2	1	identification? (Syn: Antonogadus m.)
164789	Merlucciidae	48	19	family with 1 species
164795	Merluccius merluccius	33641	9760	
165116	Echiodon drummondi	248	73	
165215	Zoarcidae	646	46	??
165243	Lycenchelys sarsi	1048	263	
165255	Lycodes	6	3	genus with 1 species
165284	Lycodes vahli	7363	4387	
165324	Zoarces viviparus	2728	496	
165350	Coryphaenoides rupestris	42	25	
165419	Trachyrhynchus murrayi	3	3	identification error: (Triglops murrayi??) – SCO
165594	Belone belone	205	88	
165612	Scomberesox saurus	2	1	
166025	Atherina presbyter	20	3	
166271	Zeiformes	8	4	??
166283	Zenopsis ocellata	12	6	identification error - DEN
166287	Zeus faber	562	253	

TSN	Taxon	#C	#H	Comments
166309	Caproidae	8	4	family with 1 species
166320	Capros aper	87	45	
615903	Lamprididae	44	8	identification error (Lampridae?) - FRA
166363	Gasterosteidae	36	11	??
166365	Gasterosteus aculeatus	8745	486	
166401	Spinachia spinachia	1009	65	inconsistencies (>Lmax) - DEN
166438	Syngnathoidei	2	1	??
166443	Syngnathidae	1965	343	??
166444	Syngnathus	8	4	?
166463	Syngnathus rostellatus	1026	211	inconsistencies among countries
166464	Syngnathus acus	1101	157	inconsistencies among countries
166467	Syngnathus typhle	30	13	inconsistencies among countries
166591	Entelurus aequoreus	1251	451	
166595	Nerophis ophidion	37	17	identification error (>Lmax) – FRA, NOR
166613	Acentronura	2	1	not in CLOFNAM
166704	Scorpaenidae	2	1	??
166745	Sebastes marinus	473	108	
166756	Sebastes mentella	2	1	
166779	Sebastes viviparus	10833	2129	
166787	Helicolenus dactylopterus	7431	1136	
166839	Scorpaena scrofa	6	4	
166972	Triglidae	5472	526	??
167039	Trigla lucerna	6021	1651	inconsistencies among countries ?
167044	Eutrigla gurnardus	2423043	115975	
167046	Trigloporus lastoviza	12	2	
167049	Aspitrigla cuculus	6067	1502	inconsistencies among countries ?
167196	Cottidae	183	110	??
167209	Artediellus atlanticus	2	1	identification?
167311	Myoxocephalus	48	1	genus with 1 species - Net 1991
167316	Triglopsis quadricornis	70	12	identification error - FRA
167317	Myoxocephalus scorpioides	416	61	identification error - FRA
167318	Myoxocephalus scorpius	22962	4826	
167375	Triglops murrayi	120	64	
167390	Taurulus bubalis	3857	758	inconsistencies (>Lmax) – DEN, ENG, FRA, SCO
167391	Taurulus lilljeborgi	20	10	identification ?
167454	Agonus cataphractus	42005	5913	
167478	Leptagonus decagonus	25	3	identification ?
167483	Cyclopteridae	8	4	??
167550	Liparis	308	46	?
167578	Liparis liparis	3489	702	inconsistencies (>Lmax) - NED
167581	Liparis montagui	169	50	inconsistencies (>Lmax)
167612	Cyclopterus lumpus	6670	3097	
168588	Trachurus trachurus	5578547	27064	
169180	Sparidae	6	4	??
169215	Pagellus erythrinus	3	2	
169229	Spondyliosoma cantharus	389	87	
169418	Mullus surmuletus	14958	2001	
169419	Mullus barbatus	4	2	identification?
170316	Dicentrarchus	12	5	genus with 1 spspecies
170317	Dicentrarchus labrax	185	70	

TSN	TAXON	#C	#H	COMMENTS
170333	Mugilidae	20	9	?
170335	Mugil cephalus	4	2	
170371	Chelon labrosus	140	48	
170376	Liza ramada	4	2	
170377	Liza aurata	13	7	
614239	Symphodus melops	14	3	
170733	Ctenolabrus rupestris	34	13	
170737	Labrus bergylta	15	8	
170739	Labrus bimaculatus	1	1	
170991	Echiichthys vipera	287451	6544	
170992	Trachinus draco	30690	1256	inconsistencies in distribution and LFD
171124	Blenniidae	2	1	??
171125	Blennius	4	1	?
171335	Anarhichadidae	61	24	Family with 1 genus
171336	Anarhichas	4	2	?
171338	Anarhichas denticulatus	6	3	
171341	Anarhichas lupus	3912	2049	
171342	Anarhichas minor	17	8	
171554	Stichaeidae	50	9	??
171588	Lumpenus lampretaeformis	43069	4424	
171603	Leptoclinus maculatus	59	27	id. errors >Lmax - Lumpenus? Den, Eng, Nor, Sco
171645	Pholis gunellus	2361	234	
171670	Ammodytidae	870559	2909	??
171671	Ammodytes	721951	1114	
171676	Ammodytes tobianus	51341	573	inconsistencies among countries (and >Lmax)
171677	Ammodytes marinus	576955	2718	inconsistencies among countries (and >Lmax)
171680	Gymnammodytes semisquamatus	37389	70	inconsistencies among countries (and >Lmax)
171681	Hyperoplus	8687	20	
171682	Hyperoplus lanceolatus	402364	6849	inconsistencies among countries
171683	Hyperoplus immaculatus	58507	328	inconsistencies among countries
171691	Callionymidae	2810	469	family with 1 genus
171692	Callionymus	636	149	?
171698	Callionymus lyra	82212	16901	
171699	Callionymus maculatus	61394	6879	inconsistencies (>Lmax)
171712	Callionymus reticulatus	838	219	inconsistencies (>Lmax)
171746	Gobiidae	16465	1086	??
171833	Gobius	22622	68	input error (Pomatoschistus or Gobiidae)
171841	Gobius cobitis	2	1	identification?
171850	Gobius niger	119	37	
171971	Crystallogobius linearis	98	19	
171977	Pomatoschistus	34624	572	
171978	Pomatoschistus minutus	32290	679	identification ?
171982	Pomatoschistus microps	46	7	identification ?
172033	Aphia minuta	150	27	inconsistencies (>Lmax) - SWE
172034	Leseurigobius	34	11	genus with 1 species
172036	Leseurigobius friesii	321	68	
172414	Scomber scombrus	1844874	30075	
172421	Thunnus thunnus	2	1	
172714	Bothidae	61	17	??
616195	Psetta maxima	3161	1573	

TSN	TAXON	#C	#H	COMMENTS
172749	Scophthalmus rhombus	2480	1147	
172803	Arnoglossus	25	11	?
172805	Arnoglossus laterna	13392	2794	
172806	Arnoglossus imperialis	36	17	identification?
172809	Arnoglossus thori	1	1	identification?
172828	Zeugopterus	4	2	?
172829	Zeugopterus punctatus	454	119	identification? inconsistencies among countries
616613	Zeugopterus norvegicus	522	198	
616605	Zeugopterus regius	15	11	identification? -SCO
172834	Lepidorhombus boscii	6	2	
172835	Lepidorhombus whiffiagonus	15160	5098	
172873	Glyptocephalus cynoglossus	34671	10419	
172877	Hippoglossoides platessoides	3073054	131798	
172881	Limanda limanda	12538952	215196	
172888	Microstomus kitt	268833	54547	
172894	Platichthys flesus	100200	15219	
172902	Pleuronectes platessa	727806	95382	
172933	Hippoglossus hippoglossus	614	323	
172980	Soleidae	615	80	??
173000	Solea	1	1	genus with 1 species
173001	Solea vulgaris	19538	5376	
173020	Buglossidium	78	24	genus with 1 species
173021	Buglossidium luteum	74921	5927	
173022	Microchirus	10	5	genus with 1 species
173026	Microchirus variegatus	425	179	
173051	Pegusa lascaris	15	7	

3.1 Consistent taxonomy

The idea of taxonomy (and its coding in one system or the other) is to provide a unique interpretation of the taxon and code used. Therefore, taxa (and codes) that provide the same interpretation must be avoided, because they suggest a non-existent difference. An example is a genus represented by a single species in a particular area, such as *Lampetra*/*Lampetra fluviatilis*. In this case, the presence of *Lampetra* is redundant in DATRAS, because it can only be interpreted in terms of *L. fluviatilis,* and therefore should not be acceptable. However, it may well be that at the national level the use of this code had a slightly different meaning, for instance that it was not rigidly identified, but that it was likely to be *L. fluviatilis*. Such a differentiation cannot be maintained within an international data base, because different countries may use different interpretations and therefore the uniqueness criterion can no longer be maintained. It is up to the nation submitting the data how it will change the taxon, but it should either be Petromyzonidae, indicating that it could have been *Petromyzon marinus*, or *L. fluviatilis*, if it is virtually certain to have belonged to the species. The choice is not be made by DATRAS, but by the countries concerned, but *Lampetra* should be considered an invalid taxon for the North Sea, where there is only one species that belongs to that genus and might reasonably be caught. A similar problem exists for families represented by one genus (e.g., *Callionymidae* with the genus *Callionymus*). In this case, the family name is redundant and should be considered invalid.

Although this problem can be easily resolved by incorporation a suitable check upon entry in DATRAS, care must be taken that these checks are made area-specific. Also, changes in the taxonomy can easily distort the information existing in the database. For instance, originally unspecified *Raja* was a valid genus, indicating that it could have been one of the many ray species occurring in the North Sea. At a particular point in time, ICES has adopted the new nomenclature, where the genus *Raja* has been reserved for a small subset, whereas the other rays have been brought under several other genus. In this case, the unique interpretation of the genus *Raja* changed at a particular point in time and therefore a suitable correction is required: all *Raja* sp reported before that time have to be changed to Rajidae! Therefore, changes to the nomenclature used in DATRAS must be supervised by a small group of taxonomists before implementation to ensure consistency in interpretation over the entire period.

The redundant taxa present in the IBTS (North Sea) dataset within DATRAS and their appropriate interpretation if submitted to DATRAS are given in table 3.1.1. This list can also be used to amend other North Sea data, such as from the beam-trawl survey, but the appropriate set of invalid taxa for other areas still has to be determined. We recommend that a comprehensive list of species likely to be reported from various regions within the ICES area with their associated minimum and maximum length to be used as a check upon data entry.

Table 3.1.1. List of invalid taxa currently reported and their appropriate interpretation.

TSN	INVALID TAXON	TSN	VALID TAXON
159700	*Lampetra*	159719	*Lampetra fluviatilis*
159721	*Petromyzon*	159722	*Petromyzon fluviatilis*
160846	*Raja*	160845	Rajidae
162057	Argentinidae	162061	*Argentina*
164771	*Gadiculus*	164772	*Gadiculus thori*
164789	Merlucciidae	164795	*Merluccius merluccius*
165255	*Lycodes*	165284	Lycodes vahlii
166309	Caproidae	166320	*Capros aper*
166438	Syngnathoidei	166443	Syngnathidae
170316	*Dicentrarchus*	170317	*Dicentrarchus labrax*
171335	Anarhichadidae	171336	*Anarhichas*
171691	Callionymidae	171692	*Callionymus*
172034	Leseurigobius	172036	Leseurigobius friesii
173000	*Solea*	173001	*Solea vulgaris*
173020	*Buglossidium*	173021	*Buglossidium luteum*
173022	*Microchirus*	173026	*Microchirus variegatus*

3.2 Inappropriate taxa for the North Sea

In the marine world, it is never totally impossible that a species is recorded in a particular area that has never been observed there before. However, in such cases it would have been appropriate to document such catches and inform appropriate museums of natural history. As long as such documentation is lacking, ICES should not trust these records and they should be flagged accordingly.

Other species may have been reported from the North Sea as occasional vagrants, but if suddenly large numbers are reported, one should be hesitant to adopt these as being valid. Sometimes reported sizes provide additional clues as to whether these identifications have been correct.

Irregularities may also occur from taxonomic confusion. A good example is the genus *Mustelus*, for which two species are generally accepted to occur in the North Sea, *M. mustelus* and *M. asterias*, the distinction being generally made on the basis of the absence or presence of white spots, respectively. However, there is growing evidence that this criterion is insufficient to separate the two species because of very gradual differences in both number and size of these white spots, which are often barely visible. For all practical purposes, we suggest to bring all the historically collected information on the two species under the taxon *Mustelus*, without trying to distinguish species. New information by species should only be accepted if supported by good taxonomic evidence that the two species have been properly identified.

Table 3.2.1 provides a list of taxa for which identification is obviously in error or for which it seems unlikely that identification has been correct. Although these species have been reported for the North Sea and the reported sizes provide no clue, we suggest carrying out thorough checks using the original records.

Table 3.2.1. Listing of (a) identification errors and (b) unlikely but not impossible identifications in DATRAS.

a. Identification errors

TSN	TAXON	#C	#H	COMMENTS
160240	*Mustelus asterias*	1470	518	uncertain identification criteria
160242	*Mustelus mustelus*	142	57	uncertain identification criteria
160838	*Torpedo marmorata*	4	2	input error: Raja clavata – NET
564140	*Leucoraja lentiginosa*	3	2	not in CLOFNAM
160876	*Raja radiata*	91	36	invalid taxon -> *Amblyraja radiata*
564141	*Dipturus lintea*	2	1	outside usual area - FRA
162471	*Notolepis rissoi*	2	1	> >Lmax - NOR
164475	*Lepadogaster*	2	1	> >Lmax - FRA
164761	*Molva dypterygia*	23	14	outside usual area
164766	*Gaidropsaus mediterraneus*	20	9	>>Lmax - GFR
164768	*Gaidropsarus argentatus*	6	1	not in NS fauna
165419	*Trachyrhynchus murrayi*	3	3	(*Triglops m.*?) - SCO
166283	*Zenopsis ocellata*	12	6	not in NS fauna - DEN
615903	Lamprididae	44	8	(Lampridae?) - FRA
166595	*Nerophis ophidion*	37	17	>>Lmax – FRA, NOR
166613	*Acentronura*	2	1	not in CLOFNAM
167316	*Triglopsis quadricornis*	70	12	outside usual area - FRA
167317	*Myoxocephalus scorpioides*	416	61	not in NS fauna – FRA
171603	*Leptoclinus maculatus*	59	27	>>Lmax (*Lumpenus l.*?) - DEN, ENG, NOR, SCO

b. Unlikely identifications (checks required)

TSN	TAXON	#C	#H	COMMENTS
160900	*Leucoraja undulata*	3	2	only very small specimens reported
550592	*Gaidropsarus macrophthalmus*	2	1	Synonym*: Antonogadus m.*
167209	*Artediellus atlanticus*	2	1	rare
167391	*Taurulus lilljeborgi*	20	10	rare
167478	*Leptagonus decagonus*	25	3	rare
169419	*Mullus barbatus*	4	2	rare
171833	*Gobius*	22622	68	too many (*Pomatoschistus*/Gobiidae?)
171841	*Gobius cobitis*	2	1	rare
172806	*Arnoglossus imperialis*	36	17	rare
172809	*Arnoglossus thori*	1	1	rare
172829	*Zeugopterus punctatus*	454	119	rare
616605	*Zeugopterus regius*	15	11	rare - SCO

3.3 Length frequency

The length-frequency distributions of all species reported can be easily checked against available information from the literature. Because most bony fishes start their life as very small larvae, in principle there is no lower size limit (Lmin) for these species in the gear. However, given the 2cm liner used, it is unlikely that these small larvae are retained by the net or identified in the catch and therefore a general Lmin of 2cm might help to identify input errors. For elasmobranchs, the situation is different, because information on size at birth is often available from the literature and this Lmin can be used to spot errors. For instance, Denmark has reported 96 unrealistically small *Mustelus mustelus* of 3-13 cm, whereas the reported minimum size at birth is about 35cm!

Similarly, the maximum sizes reported in the literature (Lmax) do help to find clear identification errors. In this context, it must be emphasized that such Lmax figures do not represent absolute maxima. Their origin is generally obscure, they may apply to a specific area and ‘exceptionally large’ specimens may occur once in a while anyway. Given the enormous amount of data entering the data base, it is therefore only to be expected that Lmax figures will have to be adjusted upwards continuously. Nevertheless, they are extremely helpful in spotting errors in specific subsets, for instance from a specific country in a specific year. For instance, of the 214 *Amblyraja radiata* above the reported Lmax of 60 cm in DATRAS, 150 refer to Scottish catches in a single year (1989)! Similarly, France reported 74 sprat >19cm in 1999, whereas no sprat in the catches from other years or from other countries exceeded the Lmax. Denmark reported 108 *Engraulis engraulis* of 198-201 cm (Lmax =20cm!), suggesting that the original measurements were in mm rather than cm. More generally, whether such inconsistencies are caused by identification errors or measurement or coding errors remains unclear, but they may be resolved by checking the original data.

A suitable protocol to deal with unrealistic values for Lmin and Lmax would involve the following steps:

- i) An examination of size frequency (all years combined) to identify the minimum and maximum lengths of all species.
- ii) These data should be reported to the relevant survey Working Groups (e.g. WGBEAM, IBTSWG, WGBIFS), collated and a standardised length range established/agreed for all species. DATRAS should update their length flags accordingly.
- iii) Data for specimens outside this agreed length range should then be checked by national labs and corrected if possible.
- iv) If exceptional lengths can be validated, then data can be included without a quality flag. If there is still doubt as to whether entries are valid (i.e. the length and/or species may be incorrect), then these data should be flagged within the national database (if possible) and DATRAS
- v) There may be latitudinal differences in Lmax for some species, and survey working groups or WGFE could usefully examine this
- vi) Given that some nations have reported several skate species to be above Lmax, national labs should also check as to whether length data have been converted from original wing width data and, if so, whether the correct conversion factors were used for the various species.

Table 3.3.1. Species recorded in DATRAS that are larger than their reported Lmax (excessively large specimens in red).

TSN	SPECIES	COUNTRY	YEAR	Q	SHIP	HAULNO	NUMBER	SIZE CLASS	LMAX
564149	Amblyraja radiata	SCO	1989	1	SCO2	all	508	>	60
161716	Alosa fallax	FRA	1999	1	THA2	61	2	110	70
171722	Clupea harengus	NOR	1995	4	GOS	14	2	44	40
161789	Sprattus sprattus	FRA	1999	1	THA2	17	10	20	19
161789	Sprattus sprattus	FRA	1999	1	THA2	17	10	20	19
161789	Sprattus sprattus	FRA	1999	1	THA2	17	8	21	19
161789	Sprattus sprattus	FRA	1999	1	THA2	17	18	21	19
161789	Sprattus sprattus	FRA	1999	1	THA2	17	12	22	19
161789	Sprattus sprattus	FRA	1999	1	THA2	17	4	22	19
161789	Sprattus sprattus	FRA	1999	1	THA2	17	8	23	19
161789	Sprattus sprattus	FRA	1999	1	THA2	17	2	23	19
161789	Sprattus sprattus	FRA	1999	1	THA2	17	2	24	19
161831	Engraulis encrasicolus	DEN	2000	3	DAN2	31	44	198	22
161831	Engraulis encrasicolus	DEN	2000	3	DAN2	31	56	199	22
161831	Engraulis encrasicolus	DEN	2000	3	DAN2	31	6	200	22
161831	Engraulis encrasicolus	DEN	2000	3	DAN2	31	2	201	22
161831	Engraulis encrasicolus	FRA	2003	1	THA2	27	2	120	22
161831	Engraulis encrasicolus	FRA	2003	1	THA2	27	2	135	22
161831	Engraulis encrasicolus	FRA	2004	1	THA2	1	10	23	22
161831	Engraulis encrasicolus	FRA	2004	1	THA2	1	6	24	22
161831	Engraulis encrasicolus	FRA	2004	1	THA2	1	4	27	22
161831	Engraulis encrasicolus	FRA	2004	1	THA2	1	2	28	22
161831	Engraulis encrasicolus	FRA	2005	1	THA2	5	2	80	22
162964	Argentina silus	NOR	2003	3	HAV	553	11	69	60
162187	Maurolicus muelleri	SCO	1985	1	SCO2	11	1	12	8
162187	Maurolicus muelleri	SCO	2003	1	SCO3	33	4	9	8
162187	Maurolicus muelleri	SWE	1972	1	SKA	31	1	9	8
162471	Notolepis rissoi	NOR	2004	3	HAV	312	2	33	29
164475	Lepadogaster	FRA	1992	1	THA	29	2	20	8
164748	Rhinonemus cimbrius	DEN	1991	1	DAN2	21	8	45	41
164754	Trisopterus minutus	DEN	1971	1	DAN	8	2	28	27
164754	Trisopterus minutus	DEN	1986	1	DAN2	27	2	29	27
164754	Trisopterus minutus	DEN	1991	1	DAN2	14	2	29	27
164754	Trisopterus minutus	DEN	1992	1	DAN2	35	2	30	27
164754	Trisopterus minutus	DEN	1992	4	DAN2	29	4	28	27
164754	Trisopterus minutus	DEN	1992	4	DAN2	29	4	30	27
164754	Trisopterus minutus	DEN	1993	1	DAN2	15	10	30	27
164754	Trisopterus minutus	DEN	1999	1	DAN2	23	2	30	27
164754	Trisopterus minutus	FRA	1999	1	THA2	2	4	29	27
164754	Trisopterus minutus	GFR	1982	1	AND2	19	2	33	27
164754	Trisopterus minutus	GFR	1983	1	AND2	61	12	29	27
164754	Trisopterus minutus	GFR	1992	1	SOL	28	2	28	27
164754	Trisopterus minutus	GFR	1992	1	SOL	28	2	29	27
164754	Trisopterus minutus	NED	1995	4	ISI	3	16	30	27
164754	Trisopterus minutus	NOR	1982	1	MIC	65	2	29	27

TSN	SPECIES	COUNTRY	YEAR	Q	SHIP	HAULNO	NUMBER	SIZE CLASS	LMAX
164754	Trisopterus minutus	SWE	1972	1	THE	4	1	28	27
164754	Trisopterus minutus	SWE	1980	1	ARG	3	2	31	27
164754	Trisopterus minutus	SWE	1980	1	ARG	32	2	35	27
164756	Trisopterus esmarki	DEN	2003	1	DAN2	37	34	28	25
164756	Trisopterus esmarki	DEN	2003	1	DAN2	37	52	29	25
164756	Trisopterus esmarki	GFR	1977	1	PO	20	44	26	25
164756	Trisopterus esmarki	GFR	1977	1	PO	20	40	27	25
164756	Trisopterus esmarki	GFR	1977	1	PO	20	7	28	25
164756	Trisopterus esmarki	GFR	1977	1	PO	20	4	29	25
164756	Trisopterus esmarki	GFR	1977	1	PO	65	50	26	25
164756	Trisopterus esmarki	GFR	1977	1	PO	65	18	27	25
164756	Trisopterus esmarki	GFR	1977	1	PO	65	24	28	25
164756	Trisopterus esmarki	GFR	1977	1	PO	65	6	30	25
164756	Trisopterus esmarki	GFR	1979	1	PO	9	10	26	25
164756	Trisopterus esmarki	GFR	1979	1	PO	9	4	29	25
164756	Trisopterus esmarki	GFR	1979	1	PO	9	2	32	25
164756	Trisopterus esmarki	GFR	1992	1	WAH2	23	4	26	25
164756	Trisopterus esmarki	GFR	1992	1	WAH2	23	8	27	25
164756	Trisopterus esmarki	GFR	1992	1	WAH2	23	10	28	25
164756	Trisopterus esmarki	GFR	1992	1	WAH2	23	10	29	25
164756	Trisopterus esmarki	GFR	1992	1	WAH2	23	2	30	25
164756	Trisopterus esmarki	GFR	1992	3	WAH2	70	8	26	25
164756	Trisopterus esmarki	GFR	1992	3	WAH2	70	8	27	25
164756	Trisopterus esmarki	GFR	2002	1	WAH3	29	81	27	25
164756	Trisopterus esmarki	NOR	1991	2	JHJ	10	25	26	25
164756	Trisopterus esmarki	NOR	1995	4	GOS	42	191	26	25
164756	Trisopterus esmarki	NOR	1995	4	GOS	42	38	27	25
164756	Trisopterus esmarki	NOR	1995	4	GOS	59	7	26	25
164756	Trisopterus esmarki	NOR	1995	4	GOS	59	14	27	25
164756	Trisopterus esmarki	NOR	1999	3	MIC	581	50	31	25
164756	Trisopterus esmarki	SCO	1980	1	EXP	26	1	26	25
164756	Trisopterus esmarki	SCO	1983	1	EXP	40	10	26	25
164766	Gaidropsarus mediterraneus	GFR	1983	1	AND2	43	2	27	25
164766	Gaidropsarus mediterraneus	GFR	1983	1	AND2	75	2	28	25
164766	Gaidropsarus mediterraneus	GFR	1988	1	WAH2	12	2	29	25
164772	Gadiculus argenteus	ENG	2003	3	END	51	2	18	15
164772	Gadiculus argenteus	SWE	1995	3	ARG	195	12	19	15
164774	Micromesistius poutassou	SWE	1991	3	ARG	193	2	51	47
164779	Ciliata mustela	NED	1976	1	TRI	54	2	53	30
166365	Gasterosteus aculeatus	SCO	1987	1	SCO2	32	3	16	11
166401	Spinachia spinachia	DEN	1992	4	DAN2	8	4	33	22
166401	Spinachia spinachia	DEN	1992	4	DAN2	8	4	43	22
166401	Spinachia spinachia	DEN	2004	1	DAN2	26	2	25	22
166401	Spinachia spinachia	DEN	2004	3	DAN2	38	6	33	22
166401	Spinachia spinachia	DEN	2004	3	DAN2	40	4	39	22
166401	Spinachia spinachia	DEN	2004	3	DAN2	43	4	38	22
166401	Spinachia spinachia	DEN	2005	1	DAN2	31	2	28	22
166401	Spinachia spinachia	DEN	2005	1	DAN2	34	2	40	22

TSN	SPECIES	COUNTRY	YEAR	Q	SHIP	HAULNO	NUMBER	SIZE CLASS	LMAX
166401	Spinachia spinachia	DEN	2005	1	DAN2	40	2	26	22
166401	Spinachia spinachia	DEN	2005	1	DAN2	42	2	34	22
166401	Spinachia spinachia	DEN	2005	1	DAN2	43	2	26	22
166401	Spinachia spinachia	DEN	2005	1	DAN2	43	2	35	22
166401	Spinachia spinachia	DEN	2005	1	DAN2	44	6	28	22
166401	Spinachia spinachia	DEN	2005	1	DAN2	44	2	31	22
166463	Syngnathus rostellatus	GFR	1993	1	WAH2	59	2	26	17
166595	Nerophis ophidion	FRA	2005	1	THA2	42	2	38	30
166595	Nerophis ophidion	FRA	2005	1	THA2	44	2	42	30
166595	Nerophis ophidion	FRA	2005	1	THA2	47	2	36	30
166595	Nerophis ophidion	FRA	2005	1	THA2	47	2	39	30
166595	Nerophis ophidion	FRA	2005	1	THA2	55	2	31	30
166595	Nerophis ophidion	FRA	2005	1	THA2	55	2	42	30
166595	Nerophis ophidion	FRA	2005	1	THA2	58	2	37	30
166595	Nerophis ophidion	FRA	2005	1	THA2	59	2	37	30
166595	Nerophis ophidion	FRA	2005	1	THA2	67	2	37	30
166595	Nerophis ophidion	FRA	2005	1	THA2	70	2	41	30
166595	Nerophis ophidion	NOR	2004	1	HAV	26	2	33	30
166595	Nerophis ophidion	NOR	2004	1	HAV	26	2	40	30
166595	Nerophis ophidion	NOR	2004	1	HAV	31	3	33	30
166779	Sebastes viviparus	DEN	1985	1	DAN2	12	2	41	39
166779	Sebastes viviparus	DEN	1993	4	DAN2	45	2	41	39
167578	Liparis liparis	NED	1994	2	TRI2	13	2	35	18
167578	Liparis liparis	NED	1997	1	TRI2	30	2	32	18
167581	Liparis montagui	FRA	1999	1	THA2	3	10	13	10
167581	Liparis montagui	FRA	1999	1	THA2	3	8	14	10
167581	Liparis montagui	FRA	1999	1	THA2	3	8	15	10
167581	Liparis montagui	FRA	2000	1	THA2	43	2	14	10
167581	Liparis montagui	FRA	2000	1	THA2	43	2	15	10
167581	Liparis montagui	FRA	2002	1	THA2	68	2	12	10
167581	Liparis montagui	FRA	2002	1	THA2	68	2	14	10
167581	Liparis montagui	FRA	2002	1	THA2	68	2	15	10
167581	Liparis montagui	FRA	2002	1	THA2	69	2	12	10
167581	Liparis montagui	FRA	2002	1	THA2	69	10	13	10
167581	Liparis montagui	FRA	2002	1	THA2	69	10	14	10
167581	Liparis montagui	FRA	2002	1	THA2	69	16	15	10
170377	Liza aurata	ENG	1993	3	CIR	2	2	52	45
170377	Liza aurata	ENG	1996	3	CIR	3	2	48	45
170377	Liza aurata	NED	1994	3	TRI2	1	2	56	45
170991	Echiichthys vipera	FRA	2001	1	THA2	52	146	19	17
170991	Echiichthys vipera	FRA	2001	1	THA2	52	163	20	17
170991	Echiichthys vipera	GFR	1995	2	WAH3	7	6	19	17
170991	Echiichthys vipera	GFR	1995	2	WAH3	7	18	20	17
171603	Leptoclinus maculatus	DEN	1989	1	DAN2	35	2	21	20
171603	Leptoclinus maculatus	DEN	1991	4	DAN2	7	2	21	20
171603	Leptoclinus maculatus	DEN	1991	4	DAN2	7	2	27	20
171603	Leptoclinus maculatus	ENG	1997	3	CIR	34	2	23	20
171603	Leptoclinus maculatus	ENG	1997	3	CIR	40	2	31	20
171603	Leptoclinus maculatus	NOR	1999	3	MIC	556	2	21	20

TSN	SPECIES	COUNTRY	YEAR	Q	SHIP	HAULNO	NUMBER	SIZE CLASS	LMAX
171603	Leptoclinus maculatus	NOR	1999	3	MIC	556	2	24	20
171603	Leptoclinus maculatus	NOR	1999	3	MIC	556	2	26	20
171603	Leptoclinus maculatus	NOR	1999	3	MIC	556	2	27	20
171603	Leptoclinus maculatus	NOR	1999	3	MIC	556	2	28	20
171603	Leptoclinus maculatus	NOR	1999	3	MIC	556	2	29	20
171603	Leptoclinus maculatus	SCO	1991	3	SCO2	33	1	27	20
171603	Leptoclinus maculatus	SCO	1991	3	SCO2	33	1	34	20
171645	Pholis gunnellus	FRA	2005	1	THA2	67	4	26	25
171645	Pholis gunnellus	FRA	2005	1	THA2	67	2	27	25
171645	Pholis gunnellus	FRA	2005	1	THA2	67	2	29	25
171645	Pholis gunnellus	FRA	2005	1	THA2	67	2	31	25
171645	Pholis gunnellus	FRA	2005	1	THA2	67	2	33	25
171682	Hyperoplus lanceolatus	GFR	1994	2	WAH3	73	2	55	40
171698	Callionymus lyra	GFR	1985	1	AND2	75	2	35	30
171698	Callionymus lyra	GFR	1985	1	AND2	75	4	36	30
171698	Callionymus lyra	GFR	1985	1	AND2	87	2	38	30
171698	Callionymus lyra	SCO	1993	1	SCO2	6	1	38	30
171978	Pomatoschistus minutus	GFR	1988	1	WAH2	46	2	11	9
171978	Pomatoschistus minutus	GFR	1992	3	WAH2	83	6	19	9
171978	Pomatoschistus minutus	GFR	2000	1	WAH3	59	2	55	9
172033	Aphia minuta	SWE	1972	1	SKA	41	2	38	6
172033	Aphia minuta	SWE	1972	1	SKA	41	2	42	6
172033	Aphia minuta	SWE	1972	1	SKA	41	2	45	6
172033	Aphia minuta	SWE	1972	1	SKA	41	2	49	6
172033	Aphia minuta	SWE	1972	1	SKA	41	2	51	6
172033	Aphia minuta	SWE	1972	1	SKA	41	4	52	6
172033	Aphia minuta	SWE	1972	1	SKA	41	2	53	6
172033	Aphia minuta	SWE	1972	1	SKA	41	2	54	6
172033	Aphia minuta	SWE	1972	1	SKA	41	2	56	6
172033	Aphia minuta	SWE	1972	1	SKA	41	4	57	6
172033	Aphia minuta	SWE	1972	1	SKA	41	4	58	6
172036	Leseurigobius friesii	SWE	1998	1	ARG	78	2	11	10
172036	Leseurigobius friesii	SWE	2001	3	ARG	20	2	11	10
172036	Leseurigobius friesii	SWE	2005	3	ARG	17	2	11	10
172805	Arnoglossus laterna	SCO	2003	3	SCO3	88	2	22	20
172829	Zeugopterus punctatus	SWE	1981	1	ARG	13	2	29	25
172835	Lepidorhombus whiffiagonis	GFR	1987	1	WAH2	101	2	65	60
172835	Lepidorhombus whiffiagonis	SCO	1980	1	EXP	51	1	62	60
172877	Hippoglossoides platessoides	GFR	1994	1	WAH3	76	2	71	50
172877	Hippoglossoides platessoides	SCO	1997	2	SCO2	35	4	51	50
172877	Hippoglossoides platessoides	SCO	1997	2	SCO2	35	1	52	50
172877	Hippoglossoides platessoides	SCO	1997	2	SCO2	35	3	53	50
172877	Hippoglossoides platessoides	SCO	1997	2	SCO2	35	5	54	50

TSN	SPECIES	COUNTRY	YEAR	Q	SHIP	HAULNO	NUMBER	SIZE CLASS	LMAX
172877	Hippoglossoides platessoides	SCO	1997	2	SCO2	35	5	55	50
172877	Hippoglossoides platessoides	SCO	1997	2	SCO2	35	2	56	50
172877	Hippoglossoides platessoides	SCO	1997	2	SCO2	35	1	57	50
172877	Hippoglossoides platessoides	SCO	1997	2	SCO2	35	1	59	50
172881	Limanda limanda	NED	1970	1	WIL	11	8	47	42
172881	Limanda limanda	NED	1970	1	WIL	11	8	48	42
172881	Limanda limanda	NED	1970	1	WIL	11	16	49	42
172881	Limanda limanda	NED	1970	1	TRI	13	4	100	42
172881	Limanda limanda	NED	1984	1	TRI	30	2	43	42
172881	Limanda limanda	NED	1984	1	TRI	30	2	45	42
173021	Buglossidium luteum	FRA	2000	1	THA2	68	4	20	18
173021	Buglossidium luteum	FRA	2000	1	THA2	68	4	21	18
173021	Buglossidium luteum	GFR	2001	1	WAH3	35	2	19	18
173021	Buglossidium luteum	GFR	2005	1	WAH3	42	2	19	18
173021	Buglossidium luteum	SCO	1980	1	EXP	53	1	19	18
173021	Buglossidium luteum	SCO	1982	1	EXP	37	1	21	18
173021	Buglossidium luteum	SCO	1983	1	EXP	14	2	20	18
173021	Buglossidium luteum	SCO	1983	1	EXP	14	1	21	18
173021	Buglossidium luteum	SCO	1983	1	EXP	15	1	19	18
173021	Buglossidium luteum	SCO	1983	1	EXP	15	1	28	18
173021	Buglossidium luteum	SCO	1984	1	EXP	17	1	21	18
173021	Buglossidium luteum	SCO	1984	1	EXP	17	1	22	18
173021	Buglossidium luteum	SCO	1991	1	SCO2	47	2	19	18
173021	Buglossidium luteum	SCO	1991	3	SCO2	53	1	19	18
173021	Buglossidium luteum	SCO	1991	3	SCO2	54	1	19	18
173021	Buglossidium luteum	SCO	1992	1	SCO2	43	3	21	18
173021	Buglossidium luteum	SCO	1992	2	SCO2	51	1	19	18
173021	Buglossidium luteum	SCO	1992	2	SCO2	52	1	19	18
173021	Buglossidium luteum	SCO	1992	3	SCO2	66	2	20	18
173021	Buglossidium luteum	SCO	1993	3	SCO2	72	2	19	18

3.4 Geographical distribution

Another important check may be found in plotting the geographical distribution of the observations reported, because these may indicate clear outliers or doubtful observations (Figure 3.4.1). However, this kind of check will only work for species that are not too rare and that are characterised by a preference for specific habitats. Figure 3.4.1 provides examples of presence/absence plots for six species over all years. The smelt (*Osmerus eperlanus*) is a coastal species of brackish and even freshwater habitats. Therefore, it would seem unlikely that a specimen would have been caught on the slope of the Norwegian Trench. In contrast, the lesser argentine (*Argentina sphyraena*) is a typical slope species that would not be expected in the southern North Sea. The eelpout is also a coastal species typically living in musselbeds and offshore observations would be unlikely. In contrast, Vahl's eelpout (*Lycodes vahlii*) is typical of deep sea areas and it is regularly caught along the border of the Norwegian Trench, while observations in the southern North Sea raise suspicion. A similar reasoning applies to the tusk (*Brosme brosme*) and the megrim (*Lepidorhombus whiffiagonis*). Especially the gap in the distribution of the latter across the central North Sea and the presence of observations south of the Doggerbank where these are not normally found suggest that misidentifications play a role.

It should be stressed that such maps can never prove that misidentification plays a role, because outliers represent a normal feature of biological distributions and specimens may inadvertently drift into areas, where they do not belong. These data can only be used to indicate potential problems, but more work is required to actually trace their origin, for instance by investigating whether outliers can be traced to particular sets (a single country or a single year). For future checks, maps only based on trustworthy observations would really help to identify potential errors before the data are actually entered in DATRAS, as well as to ensure that new outliers are properly documented so that adjusted maps may be drawn.

We suggest the following guidelines for checking for geographical outliers:

i) Preparation of examination of distribution maps (all years combined) for all species to identify any obvious outliers or unlikely distributions.

ii) Tracking the origin of the unlikely records to year and country and feeding this information to the national labs for checking and correcting if possible.

iii) The maps should also be reported to the relevant survey working group, for collating an agreed list of valid species by ICES Division. DATRAS should update their flags accordingly.

iv) If there is still doubt as to whether entries are valid based on the confirmed geographic range or otherwise, these data should be flagged within DATRAS.

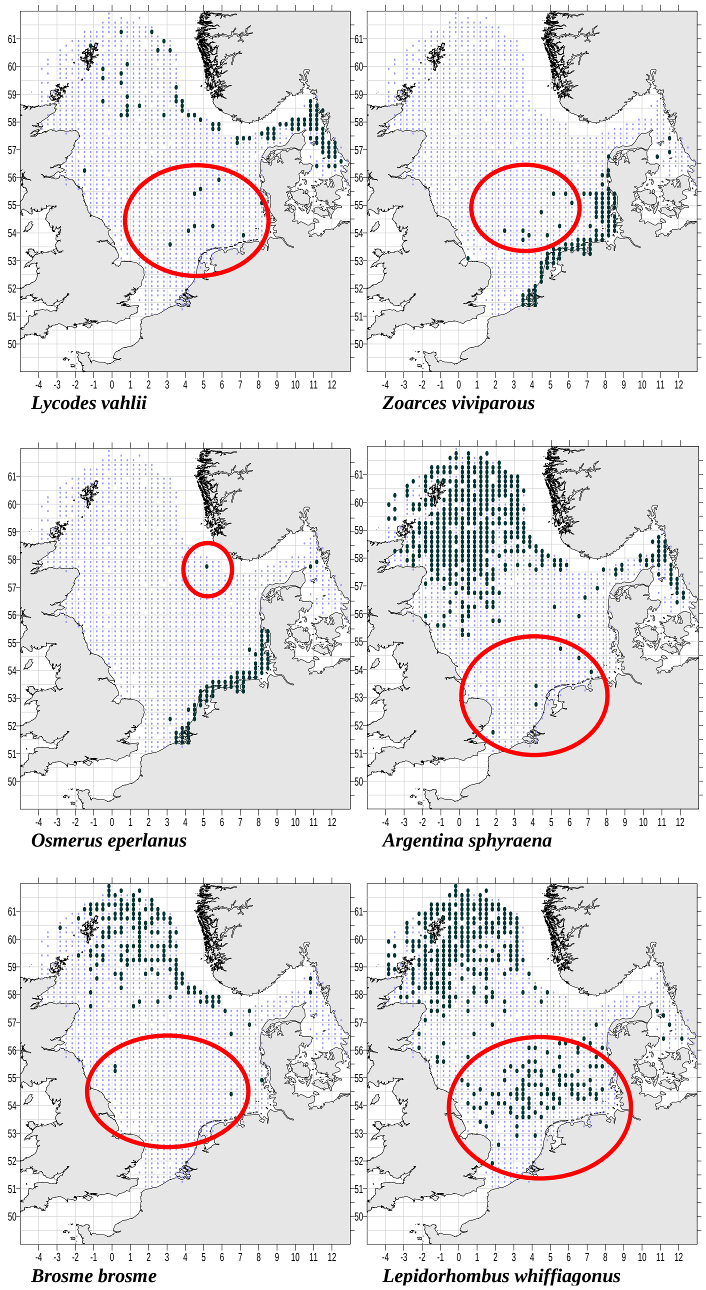

Figure 3.4.1. Examples of geographical outliers among observations as recorded in DATRAS (presented as presence/absence plots).

3.5 Problematic taxa

If all species would have been identified correctly, one might expect some consistency in the numbers reported by countries sampling the same areas in the same year. Thus, the comparison between data obtained by the individual countries is useful to identify discrepancies that may be traced back to misidentifications. Although it may not be directly obvious which component should be considered the correct one, the additional information on size compositions reported may help to identify the sets that are most likely incorrect. Although a comprehensive analysis is still lacking, some major inconsistencies have been identified in the past and new ones are still emerging. We will provide two examples here, but additional ones are presented in Daan (2001)m and ter Hofstede and Daan (2006a). After the examples we provide a first outline of a protocol that may be useful in dealing with problematic taxa, but we feel that this has to be tested in practice and may need to be improved after more experience is gained.

3.5.1 Some examples

3.5.1.1 Rocklings

Six species of rocklings are reported in DATRAS: *Rhinonemus cimbrius* (four-bearded r.), *Gaidropsarus vulgaris* (three-bearded r.), *G. mediterraneus* (shore r.), *G. argentatus* (Arctic r.), *Ciliata mustela* (five-bearded r.) and *C. septemtrionalis* (northern r.). Of these, *R. cimbrius* should not cause much of a problem. However, the other five may be easily confused, especially the smaller individuals. In the Netherlands, catches of *C. septemtrionalis* have only been positively ascertained since 2004 and there is a strong suspicion that these have been formerly mistaken for either *G. mediterraneus* or *C. mustela*.

Table 3.5.1.1 provides the reported length distributions of the five species. Of the 20 specimens reported by Germany, six exceeded the Lmax reported in the literature and these were all recorded in years (1983, 1988 and 1991 when *G. vulgaris* were notably absent in the catches. We consider that these must have been mid-identified. The six *G. argentatus* were reported by Norway from a single haul in 1995 and are likewise not to be trusted completely. The other length compositions stay within the Lmax with the exception of two *C. mustela* reported by the Netherlands in 1976, which was excessively large (53 cm compared to an Lmax of 30 cm). Furthermore, compared to catches of other countries, eight exceptionally small *G. vulgaris* reported by Denmark in 1988, which appear to need checking.

Table 3.5.1.1.1. Reported rocklings by size class and taxa (inconsistencies highlighted).

	Gaidropsarus	*G. vulgaris*	*G. mediterraneus*	*G. argentatus*	*Ciliata mustella*	*C. septemtrionalis*
2	.	2			.	.
3	.	4			.	.
4	.	2			2	.
5	.	.			5	1
6	1	4			9	5
7	.	16			34	12
8	.	39			82	20
9	.	43			108	31
10	1	28			78	10
11	.	8			59	8
12	.	18			82	2
13	1	19			118	4
14	1	47			142	1
15	6	46			92	2
16	7	103			146	.
17	6	116	2		127	.
18	15	139			85	.
19	4	167			78	.
20	13	140	2		76	.
21	12	113			121	.
22	8	75	6		42	.
23	.	71	2		27	.
24	3	40		6	11	.
25	1	31	2		10	.
26	1	33			.	.
27	1	13	2		2	.
28	.	16	2		.	.
29	.	.	2		.	.
30	.	4			.	.
33	.	1			.	.
34	.	2			.	.
35	.	1			.	.
36	.	6			.	.
38	.	2			.	.
39	.	4			.	.
41	.	2			.	.
44	.	2			.	.
46	.	3			.	.
47	.	1			.	.
53	.	.			2	.

The numbers of the three problematic species reported by year reveal some major discrepancies among data sets (Table 3.5.1.2). We consider that the number of G. vulgaris especially may have been overestimated, while at least the number of C. septemtrionalis reported by the Netherlands has been underestimated before 2004. These rockling data need careful checking.

Table 3.5.1.1.2. Reported catches of three problematic rockling species by country and year (exceptionally large numbers highlighted).

	GAIDROPSARUS VULGARIS	CILIATA MUSTELLA	C. SEPTEMTRIONALIS
DEN 1987	..	2	.
DEN 1988	.23	.	.
DEN 1991	..	13	.
DEN 1992.	4	.	.
DEN 1994-2005.	.	6	
ENG 1977.	43	.	.
ENG 1982.	4	2	.
ENG 1983.	.	18	.
ENG 1984.	4	.	10
ENG 1985.	.	12	2
ENG 1986.	10	2	2
ENG 1987-1988.	.	7	.
ENG 1989.	.	30	4
ENG 1990-2001.	.	16	.
ENG 2002.	2	10	.
ENG 2003.	352	40	.
ENG 2005.	150	.	.
FRA 1991-1996.	17	.	.
FRA 1997.	20	30	.
FRA 1998.	20	.	.
FRA 1999.	.	36	.
FRA 2001.	.	63	.
FRA 2002.	2	20	.
FRA 2003.	18	186	.
FRA 2004.	.	48	.
FRA 2005.	6	74	36
GFR 1990	36	.	.
GFR 1992.	.	46	.
GFR 1995.	18	.	.
GFR 1997.	.	10	.
GFR 1998.	.	14	.
GFR 2000.	4	16	.
GFR 2003.	2	2	.
NED 1970-1981.	9	15	.
NED 1984-2003.	6	35	.
NED 2004.	.	7	28
NED 2005.	.	7	7
NOR 1976-2005.	30	.	.
SCO 1970-1975.	26	.	.
SCO 1976.	.	1	.
SCO 1977-1983.	2	.	.
SCO 1984.	2	1	1
SCO 1985.	5	1	1
SCO 1987.	1	.	.

	Gaidropsarus vulgaris	Ciliata mustella	C. septemtrionalis
SCO 1990.	2	.	.
SCO 1996.	.	.	1
SCO 1999-2002.	7	.	.
SWE 1993.	4	.	.
SWE 1997-200.	.	11	.

3.5.1.2 Sandeels - Ammodytidae

Sandeels represent a notoriously difficult group. They are reported in DATRAS under 8 taxa, of which 5 are at the species level and 3 refer to the genus or family level (Table 3.5.2.1). The species composition reported by individual countries is highly inconsistent, with some countries reporting predominantly one or the other (highlighted). While it is highly unlikely that one country could catch only one species and another only another one, there appears to be a major problem with the higher taxa as well. Ammodytidae formally include *Ammodytes*, *Gymnammodytes* and *Hyperoplus*. Thus, if large numbers of *Hyperoplus* are reported, one might expect that these have been properly identified (the protrudable mouth is a good characteristic), so that Ammodytidae in this case should represent only *Ammodytes* and *Gymnammodytes*. Some countries seem to account for this by reporting *Ammodytes* (although these could include *Gymnammodytes*).

Table 3.5.1.2.1. Overall numbers of sandeels by taxon reported by individual countries (outliers highlighted).

	Ammodytidae	*Ammodytes*	*A. tobianus*	*A. marinus*	*G. semisquamatus*	*Hyperoplus*	*H. lanceolatus*	*H. immaculatus*
DEN	393798	9914	1406	4435	36461	0	111506	0
ENG	2	220171	90	2	4	0	117797	4
FRA	3472	922	32608	6	0	8687	9877	55193
GFR	2	6889	0	508403	0	0	45107	2
NED	446696	1515	94	146	0	0	54936	272
NOR	25427	464004	9952	681	0	0	4539	0
SCO	355	0	3698	59726	906	0	33752	2841
SWE	777	1544	3206	24	18	0	22828	0

The problem is even worse, because there are also huge inconsistencies in reporting within countries among years (Table 3.5.2.2). For many countries, reports of non-*Hyperoplus* species jump all over the place. We suggest that none of these records can be trusted, with the exception of *Hyperoplus* sp. and *Ammodytes* sp. (in the understanding that these might include *Gymnammodytes*). Ammodytidae should be reserved for cases where no *Hyperoplus* is reported. All records should be modified accordingly.

Table 3.5.1.2.2. (Average) number of sandeels reported by taxon by individual countries and years.

	Ammodytidae	Ammodytes	A. tobianus	A. marinus	G. semisquamatus	Hyperoplus	H. lanceolatus	H. immaculatus
DEN 1971-1986		1652						
DEN 1987-1997	6703						306	
DEN 1998	19405	.	.	.	135	.	5532	.
DEN 1999	1172	.	1404	.	.	.	13406	.
DEN 2000	214	.	.	3437	.	.	6126	.
DEN 2001	277324	.	.		.	8038	.	42935
DEN 2002-2003	7458						4253	
DEN 2004	1624	.	2	.	21292	.	9637	.
DEN 2005	5413	.	.	998	6996	.	21993	.
ENG 1978-1990		14837						
ENG 1991-1993		48157					1264	
ENG 1994	2	1136	.	.	.	.	2165	.
ENG 1995	.	237	.	.	.	.	99435	.
ENG 1996	.	28463	.	2	.	.	448	.
ENG 1997-2000		764					1413	
ENG 2001	.	27160	86	.	4	.	3675	.
ENG 2002	.	614	.	.	.	.	2103	.
ENG 2003	.	196	4	.	.	.	528	.
ENG 2005	.	.	.	.	.	.	.	4
FRA 1985-1992	355							
FRA 1993	1328	.	.	.	.	.	1560	.
FRA 1994	16	.	4	.	.	.	2414	40586
FRA 1995	.	.	.	.	.	.	225	136
FRA 1996	.	.	.	.	.	8563	3869	13739
FRA 1997	.	.	.	.	.	.	365	569
FRA 1998-1999			432				104	36
FRA 2000	.	.	8	.	.	124	300	56
FRA 2001	.	922	30714	.	.	.	154	17

	Ammodytidae	Ammodytes	A. tobianus	A. marinus	G. semisquamatus	Hyperoplus	H. lanceolatus	H. immaculatus
FRA 2002	.	.	613	.	.	.	332	.
FRA 2003	.	.	103	.	.	.	118	2
FRA 2004	.	.	302	.	.	.	166	.
FRA 2005	.	.	.	6	.	.	166	16
GFR 1977-1987		1148						
GFR 1988	.	.	.	.	.	.	2	.
GFR 1989	2	.	.	.	.	.	106	.
GFR 1990	.	2	.	4	.	.	70	.
GFR 1991-2002				42330			3555	
GFR 2003	.	.	.	20	.	.	1452	2
GFR 2004-2005				210			407	
NED 1970-1981	641						137	
NED 1982	.	14	.	.	.	.	.	.
NED 1983	.	10	70	.	.	.	14	.
NED 1984	.	660	4	146	.	.	16	.
NED 1985-2001	25824						3346	
NED 2002	.	52	20	.	.	.	.	272
NED 2003-2005		260					46	
NOR 1971	.	1	.	.	.	.	.	.
NOR 1975	.	460357	.	.	.	.	6	.
NOR 1976	.	42	1625	.	.	.	.	.
NOR 1977-1978	44							
NOR 1979	2	.	342	.	.	.	.	.
NOR 1981-1982		116						
NOR 1984	18	2	.	.	.	.	.	.
NOR 1985	22	.	.	.	.	.	.	.
NOR 1986-1989			238					
NOR 1990	.	10	.	.	.	.	2	.
NOR 1991	.	693	2	.	.	.	2	.

	Ammodytidae	Ammodytes	A. tobianus	A. marinus	G. semisquamatus	Hyperoplus	H. lanceolatus	H. immaculatus
NOR 1992	36	116	.	14	.	.	94	.
NOR 1993	.	2032	.	.	.	.	2	.
NOR 1994	496	16	.	.	.	.	108	.
NOR 1995	2586	.	.	220	.	.	506	.
NOR 1996	2525	499	7018	329	.	.	274	.
NOR 1997	13748	.	.	.	.	.	2	.
NOR 1998	4	.	.	.	.	.	2	.
NOR 1999	5840	4	.	42	.	.	857	.
NOR 2000	.	.	.	10	.	.	2111	.
NOR 2001	.	.	2	60	.	.	32	.
NOR 2002	32	.	.	6	.	.	506	.
NOR 2004	30	.	13	.	.	.	35	.
SCO 1971-1982	29							
SCO 1983	41	.	3182	8118	.	.	1	.
SCO 1984-1991				460			1367	
SCO 1992	.	.	.	5351	.	.	4404	2822
SCO 1993-1994			194	5401			1950	
SCO 1995	.	.	.	24103	.	.	10305	2
SCO 1996	.	.	.	384	.	.	28	.
SCO 1997	.	.	129	1574	8	.	512	1
SCO 1998-1999				1690			130	2
SCO 2000-2001				957			1157	
SCO 2002-2003				16			140	5
SCO 2004	.	.	.	194	.	.	120	.
SCO 2005	.	.	.	194	898	.	696	2
SWE 1972	1	.	.	.	.	.	.	.
SWE 1975	10	.	.	.	.	.	4	.
SWE 1977-1981	3							
SWE 1983	.	.	2	.	.	.	.	.
SWE 1986	.	.	302	.	.	.	4	.

	Ammodytidae	Ammodytes	A. tobianus	A. marinus	G. semisquamatus	Hyperoplus	H. lanceolatus	H. immaculatus
SWE 1991	.	.	.	.	18	.	540	.
SWE 1992	.	.	.	.	.	.	49	.
SWE 1993	466	.	1724	.	.	.	11002	.
SWE 1994	82	.	105	.	.	.	483	.
SWE 1995	.	.	1071	.	.	.	2596	.
SWE 1996-1998	59						108	
SWE 1999	.	.	.	.	.	.	1566	.
SWE 2000	.	.	.	8	.	.	4	.
SWE 2001	10	.	2	2	.	.	72	.
SWE 2002	22	60	.	.	.	.	10	.
SWE 2003	.	342	.	.	.	.	20	.
SWE 2004	2	14	.	.	.	.	8	.
SWE 2005	.	1128	.	14	.	.	6146	.

3.5.2 Protocol for dealing with problematic taxa

The problems already identified should be addressed by the IBTSWG in 2007, ensuring that national labs examine the inconsistencies reported and submit their experiences and conclusions to the 2008 meeting. However, we suggest that all responsible working groups set up a more systematic approach to check consistency for problematic taxa throughout the data set. As this potentially turns out to be a time consuming process, we recommend a stepwise approach, concentrating in the first instance on the period 2000-2006, ensuring that we build up a recent set of reliable information on abundance and distribution, which can help to review and repair the older information.

The following protocol may aid in setting up this checking process:

i) Prepare lists of the numbers by taxon reported by individual countries by year for comparison across countries and across years to identify major inconsistencies. Also plots of the spatial distribution of all problematic taxa by country and year will help to identify clear outliers.

ii) National laboratories should update any records that are obvious input or processing mistakes.

iii) Real misidentifications should only be amended after the relevant survey working group has reviewed the proposed changes, so as to ensure a consistent approach. Joint analyses of problematic taxa should be examined at meetings of e.g. WGBEAM and IBTSWG in order to establish an appropriate correction procedure before these data are corrected by national labs.

iv) If there is still doubt as to whether entries are valid, then these data should be flagged within DATRAS

v) DATRAS should ensure that flagged records are reported at a higher taxonomic level when data are made available for public access.

After all nations have checked 2000-2006 data, then a programme for the checking of earlier data can be developed.

Problematic taxa in North Sea, southern and western IBTS include:

- Lampreys (Petromyzontiformes)
- Smoothhounds (*Mustelus* spp.)
- Skates and rays (Rajidae)
- Shads (*Alosa* spp.)
- Argentines (*Argentina* spp.)
- Rocklings (Gadidae, Lotinae)
- Clingfishes (Gobiesocidae)
- Sticklebacks (Gasterosteidae)
- Seahorses and pipefish (Syngnathidae)
- Redfish (*Sebastes* spp.)
- Scorpion fish (*Scorpaena* spp.)
- Sea scorpions (Cottidae)
- Horse mackerel (*Trachurus* spp.)
- Sea breams (Sparidae)
- Mullets (Mugilidae)
- Wrasse (Labridae)
- Eelpouts (Zoarcidae)
- Snake blennies (Stichaeidae)
- Blennies (Blennidae)
- Sand eels (Ammodytidae)
- Dragonets (*Callionymus* spp.)
- Gobies (Gobidae)
- Topknots (*Phrynorhombus* sp. and *Zeugopterus* sp.)
- Scaldfish (Arnoglossus spp.)
- Soles and tonguefishes (e.g. *Bathysolea* and *Diclogoglossa*)

Additionally, those surveys operating on the edge of the continental shelf may sample many other problematic taxa, such as deep-water sharks (Squalidae), rat-tails (Macrouridae), myctophids (Myctophidae), hatchet fish (*Sternoptychidae*), Beryx spp. and Hoplostethus spp.

4 ToR (b): Development of protocols for ensuring the appropriate treatment of data reported at higher taxonomic levels

Raw DATRAS data cannot be used for certain types of diversity studies without prior data manipulation to standardise treatment of species reported at different taxonomic levels. If for instance, one country reports *Argentina* spp., while others have properly identified *A. sphyraena* and *A. silus*, an analysis of species richness would require that either all species are combined under *Argentina* spp. (species information is de facto rejected) or all records of *Argentina* spp. would have to be decomposed in the constituting species on the basis of records of similarly sized fish with reliable identifications in the same year and area. Any quantitative results of the analysis will thus depend on the actual routine used to combine or to split taxa and therefore, it would be appropriate to develop a standard protocol for using DATRAS information in the context of community studies, because only this can ensure that the analyses give similar results when they are repeated in the future. Here we provide a first draft of such a standardized algorithm for treatment of taxonomic data (adapted from Daan *et al*., 2005).

- i) Define taxa that should always be grouped into higher taxonomic units (at least until more reliable data on species compositions become available). This should presently apply for instance to sandeels (*Ammodytes* spp.), smelts (*Hyperoplus* spp.) and the gobies belonging to Pomaschistes spp.;
- ii) Assign all records with a flag because of unreliable species identification to the next higher taxonomic level;
- iii) Use Lmax as a distinguishing criterion;
- iv) If the species has not been assigned, try to split records at the family or genus level into the constituting genus and species, respectively, by trying to match, using size class (cm) as a distinguishing feature, with reliable information on the constituting taxa using a hierarchical approach:
  - a) for the same rectangle and year;
  - b) for the same rectangle averaged over all years;
  - c) for the same year averaged over all rectangles;
  - d) averaged over all years and rectangles;
- v) If the species has not been assigned (because the particular size class had not been recorded in any other year, use the relative abundance of the constituting species to split the catch using the same hierarchical approach as under iii.
- vi) Round always to the nearest integer to avoid fractional numbers caught (inappropriate for diversity indices).

In the case catches are reported at the family level, the algorithm may have to be repeated for the genus level. If the distribution areas of the different constituting taxa differ substantially, this should be incorporated by averaging over the rectangles in these areas rather than over all rectangles. This approach should allow all higher taxa to be suitably decomposed in their constituent taxa.

While it would be appropriate to develop such a standard algorithm for use by ICES working groups and ICES products such as ICES FishMap, it might be cumbersome to actually apply this algorithm whenever DATRAS data are made avialable for external users. Rather attention should be drawn to the problem that higher taxa need to be decomposed in their constituent species before the data can be used for studying trends in species affected by problematic idenitification and an appropriate SAS routine to do this might be given upon request. It is of course up to the specific survey working group to decide upon these matters.

5 ToR (c): Development of improved protocols to ensure that species identification in trawl surveys is appropriate for fish community studies, including the development of photo-ID keys for nations participating in surveys

Many potential methods for improving the quality of data collection in the field have been identified. The survey working groups should review these and implement those considered most appropriate:

- Survey managers should attempt to ensure that all staff carrying out groundfish surveys comprise of at least one person (or one person per shift for cruises working 24 hours) who has a high level of experience regarding identification of species, their distribution, and life history.
- Electronic ID guides and photo catalogues should be further developed for all problematic taxa (see Annex 2). These guides should be checked in the field by the various fishery institutes and updated, improved and expanded where appropriate.
- Electronic distribution (or benthimetric) maps should be produced, and regularly updated, based on reliable DATRAS information, excluding flagged data (e.g. through ICES FishMap).
- Survey reports should include a list of species recorded during the survey (including the number of specimens) with special notes clarifying the details of any records of rare and unusual fish and shellfish species. Key taxonomic features (if possible supplemented by photographs) for records of unusual species should be included. These reports should be presented at the relevant survey Working Group for quality control.
- If a species new to a survey is recorded, the specimen should be retained and deposited in a museum collection (in the case of large, obvious species or those that are rare, photographic evidence would suffice).
- Deck personnel should be instructed to photograph and retain any species that cannot be identified with certainty at sea for subsequent verification at the laboratory. Similarly, any species of unusual size or place of capture or any aberrant specimens (e.g. differing colour pattern) should also be retained and photographs added to the catalogue. For problematic taxa, additional photographs of key taxonomic features should be collected.
- National museums may be able to verify and hold a reference collection of the fish species taken during the various surveys. Similarly, institutes should consider keeping and maintaining a reference collection, if resources are available.
- The IBTS manual should update its section on fish identification, including protocols for ensuring improved data collection, as well as standardised and correct use of higher taxa (genus/family).
- Fishery institutes should attempt to maintain a certain level of taxonomic expertise, and also improve staff training for species identification (e.g. more in-house training, staff exchange programmes, participation in ring tests etc.). It was noted that several nations had fish identification courses (see Annex 2) that have been announced), and there may be a role for ICES in providing such courses.
- Any data data entry routines for use at sea should try to include filters for size ranges (Lmin and Lmax) etc. to ensure that species of unusual size can be flagged and checked as soon as possible (note that Lmin is also needed as there are several records of elasmobranchs that are less than the size at birth/hatching). These filters also help to reduce input errors in respect of size rather than errors in species identification!
- During the 2007 Q3 and Q4 surveys, one voucher specimen of selected problem taxa could be retained for a one-day workshop to be convened in 2008, so that national participants can compare their actual species identification. Such a workshop could be convened, for example, prior to IBTSWG or WGBEAM.

It is noted that there are several national initiatives on their way to produce identification guides for the marine fauna (see Annex 2)

6 ToR (d): Development of protocols for (i) improving quality control during the submission of data to DATRAS and (ii) the future checking and quality assurance of DATRAS data.

6.1 (i) Improving quality control during the submission of data to DATRAS

6.1.1 Warning messages for dubious records during submission

A sound approach to improve taxonomic quality control of newly submitted data is to use existing information on the size distribution and spatial distribution of the different species. In order to avoid the input of incorrect data into DATRAS, a warning message should be given during the data-submission process when the new data deviate from the historic data to allow a timely validation of the identification.

DATRAS is already equipped with such a warning system, but it needs to be thoroughly updated. Therefore, a list of fish species known to occur in a particular survey area has to be produced that includes information on their appropriate minimum and maximum lengths. These lists should be made for each area separately, i.e. the Baltic Sea (ICES Sub-divisions 22-32, the North Sea/Skagerak/Kattegat (Sub-area IV and Division IIIa), the western area (ICES Sub-areas VI and VII) and the southern area (Sub-areas VIII and IX). The responsible working groups, i.e. WGBIFS (Baltic International Fish Survey Working Group, Baltic sea), IBTSWG (International Bottom Trawl Survey, North Sea, Western and Southern Area) and WGBEAM (Working Group on Beam Trawl Surveys, North Sea) should provide DATRAS with these lists.

These lists should also be regularly updated by these working groups, based on the annual information collected on newly recorded species and on verified recordings of sizes exceeding the reported maximum length in the literature.

6.1.2 Recording/screening data on invertebrates.

The issue of including selected invertebrate species in the standard sampling schemes was also raised. At present there is no agreed procedure for recording invertebrate data. Some countries appear to provide data on some species, but there is not an agreed list and the range of species varies among countries and years. As a consequence, absence of a species in a particular haul may not mean that that species has not been caught. Therefore, at present the invertebrate data reported to DATRAS are largely useless.

The Manual (revision VII) contains a sentence (Sect. 3) on how to measure carapace length in crustaceans and provides figures for *Cancer*, *Maja* and *Nephrops*. In most IBTS surveys, the commercially important invertebrates, e.g. shrimp, edible crab, Norway lobster, squid and octopus appear to be recorded, but a formal requirement to do so and exactly for which species is lacking.

In order to improve the consistency in reporting data on invertebrates we recommend that the various working groups clarify which taxa are to be consistently included in the records. Also, DATRAS should reject any data on additional species reported.

6.2 (ii) Future checking and quality assurance of DATRAS data

6.2.1 Standardized annual check by survey working groups

Standardized output of the data collected during the latest survey should be extracted from DATRAS on an annual basis and these should be thoroughly checked during the working group meetings for obvious errors and unlikely records. The creation of this output should be simple and straightforward and the check has to be practical and not time consuming.

Although the working groups are responsible for defining a convenient format for such a standard annual check, we suggest that the output is arranged by a limited number of regions within the survey area and includes:

- A list of all species recorded with the number of positive hauls, total number caught (actual rather than raised to one hour fishing), Lmin, Lmax, and the total number of hauls by region;.
- Length-frequency distributions of all species caught over the entire survey area;
- Distribution maps with numbers reported by species;
- A cross-table showing numbers of all species caught versus all countries, in order to allow the comparison of species identification across countries. Countries with an overlap in survey area should be marked.

These annual files should be compared with the average results for the preceding 5 years as extracted from DATRAS to track possible drifts in species identification.

The working group should identify any inconsistencies or irregularities in the dataset that need to be checked and/or changed, and document these in the WG-report. Action should be undertaken by the national institutes to adjust their data if necessary and resubmit the set to DATRAS. It should be emphasized that only national laboratories have the authority to change DATRAS data. However, the survey working groups do have the authority to flag any dubious records in the data base and they should take the responsibility to do so, if they feel that the wrong conclusions might be drawn if these records are taken for granted.

6.2.2 Process of correcting errors in DATRAS

Daan (2001) was the first to highlight that the North Sea IBTS data base as maintained by ICES at that time contained serious input errors and mis-identifications. The results of his analysis were alarming because of the large number of taxa providing unrealistic information. He concluded that the scientific credibility of ICES was at stake, because it seemed a waste to maintain a flawed database at ICES Headquarters. Therefore, he made a strong plea for a comprehensive check of all data in the existing database, with an appropriate correction of uncertain identifications to higher taxa to reflect inherent uncertainties, and to embark on a serious quality control programme to ensure that future surveys yield only reliable information at the species level.

However, it took until 2006 before the issue was put on the table again. A working document of ter Hofstede and Daan (2006a) listed problematic data available through DATRAS (section "North Sea International Bottom Trawl Survey") for the years 1965-2005 and provided recommendations for changes. The IBTSWG and WGFE (Working Group on Fish Ecology) took the matter serious enough to recommend this one-off workshop to address taxonomic data quality issues in the existing DATRAS database (ICES, 2006a; ICES 2006b). After the ICES Annual Science Conference in 2006, the Resource Management Committee recommended to convene such a workshop in 2007. This means that 6 years have past since the first warning was announced. We strongly recommend that the protocols designed during the workshop are implemented in the near future for all surveys entering DATRAS, particularly because similar exercises for other surveys than the North Sea IBTS have not yet been performed.

7 References

Daan, N. 2001. The IBTS database: a plea for quality control. ICES CM 2001/T:03.

Daan, N., Gislason, H., Pope, J. G., and Rice, J. 2005. Changes in the North Sea fish community: evidence of indirect effects of fishing? - ICES Journal of Marine Science, 62: 177-188.

ICES. 2006a. Report of the Working Group on Fish Ecology (WGFE), 13–17 March 2006, ICES Headquarters, Copenhagen. ICES CM 2006/LRC:06. 154 pp.

ICES. 2006b. Report of the International Bottom Trawl Survey Working Group (IBTSWG), 27-31 March 2006, Lysekil, Sweden. ICES CM 2006/RMC:03, Ref. ACFM. 298 pp.

Ter Hofstede, R., and Daan, N. 2006a. Quality check surveys: DATRAS - North Sea IBTS. In ICES. 2006. Report of the International Bottom Trawl Survey Working Group (IBTSWG), 27-31 March 2006, Lysekil, Sweden. ICES CM 2006/RMC:03, Ref. ACFM. p 216-292.

Ter Hofstede, R., and Daan, N. 2006b. Green light needed for quality check DATRAS; examples from the North Sea IBTS. CM 2006/M:38.

Annex 1: List of participants

Name	Address	Email
Angelino Meerhaeghe	MUMM Gulledelle 100 B-1200 Brussels Belgium	A.Meerhaeghe@mumm.ac.be
Anne Sell	BFAFi ISH Palmaille 9 D-22767 Hamburg Germany	anne.sell@ish.bfa-fisch.de
Barbara Bland	SBF P.O. Box 4 SE-453 21 Lysekil Sweden	Barbara.bland@fiskeriverket.se
Brian Harley	Cefas Lowestoft Pakefield Road NR33 0HT Lowestoft Suffolk United Kingdom	brian.harley@cefas.co.uk
Dave Kulka	DFO P.O. Box 5667 A1C 5X1 St John's, NF Canada	KulkaD@dfo-mpo.gc.ca
David Stokes	MI Rinville Co. Galway Oranmore Ireland	david.stokes@marine.ie
Francisco Velasco	IEO Promontorio San Martin, s/n P.O. Box 240 E-39080 Santander Spain	francisco.velasco@st.ieo.es
Jim Ellis	Cefas Lowestoft Pakefield Road NR33 0HT Lowestoft Suffolk United Kingdom	jim.ellis@cefas.co.uk
Lena Inger Larsen	ICES H. C. Andersens Boulevard 44-46 DK-1553 Copenhagen V Denmark	lena@ices.dk
Niels Daan	IMARES Haringkade 1 NL-1976 CP IJmuiden Netherlands	niels.daan@wur.nl
Remment ter Hofstede	IMARES Haringkade 1 NL-1976 CP IJmuiden Netherlands	remment.terhofstede@wur.nl

NAME	ADDRESS	EMAIL
Rudolf Kafemann	BFAFi, ISH Palmaille 9 D-22767 Hamburg Germany	rudolf.kafemann@ish.bfa-fisch.de
Sten Munch-Petersen	DIFRES Charlottenlund Slot DK-2920 Charlottenlund Denmark	smp@difres.dk
Franz Uiblein	Institute of Marine Research P.O. Box 1870 N-5817 Bergen Norway	franz.uiblein@imr.no
Gorden Henderson	FRS Marine Laboratory 335 Victoria Road Aberdeen AB11 9DB	g.i.henderson@marlab.ac.uk

Annex 2: Information on taxonomic guides and training courses

Workshop on Taxonomic Quality Issues in the DATRAS Database

ICES Headquarters, 23–25 January, 2007

Ifremer, France

Ifremer has developed a reference taxonomic database (REFTAX-SIH) that includes some 5000 taxa comprising of fishes, crustaceans, molluscs, echinoderms and others. This list of valid names is consistent with up-to-date taxonomic nomenclature and internationally validated lists (Fishbase for fishes, CLEMAM for molluscs, MarBEF for crustaceans, echinoderms and others groups).

REFTAX-SIH is an ACCESS database. The main table contains all taxa gathered from tables issued from different surveys (IBTS, EVHOE), while some satellite tables provide a systematic classification, a list of synonyms, a list of vernacular French names and another one for English names (at the planning stage). Some other tables are in preparation, such as for special external codes (NODC, FAO,…), which will permit to connect our data with data of other countries.

Identification guides written by Ifremer scientists:

Quéro J. C., Porché P. et Vayne, 2003. Guide des poissons de l'Atlantique européen. Les guides du naturaliste, Lonay (Suisse), Delachaux et Niestlé (Paris), 465 p. + 85 p. de pl. en couleur. ISBN 2-603-01271-1.

Garren F., Vérin Y. et Dufour J.L., 2003. Fiches d'aide à l'identification. Poissons et céphalopodes de Manche et de mer du Nord. Ifremer, Boulogne-sur-Mer, Rapport interne Ifremer DRV/RH/DT/03-07, 23 p.

Martin J., 2007. Les invertébrés du golfe de Gascogne. Manuel d'aide à l'identification des espèces capturées par chalutage. Éditions QAE, Brest, 240 p.

Souplet A., 2007. Manuel pratique d'identification des espèces démersales de la Méditerranée française. Basé sur les principales espèces de la campagne MEDITS (MEDiterranean International Trawl Survey). Version 01 du Rapport interne LHMT 2007/01, 60 p.

Training courses for people taking part in surveys have been planned with the tutoring taken care of by the National Museum of Natural History. The first one dealing mainly with fish species will take place at the end of March 2007, while a second is planned for invertebrates. Other courses for new colleagues will follow. One purpose of such courses is the appropriate use of identification keys for families, genera and species when uncommon species are caught. This will require a good knowledge of the anatomy.

IMR, Norway

A series of fish-taxonomy workshops, "Training and Guidance in Fish Species Identification", is being organised to (1) enhance expertise in fish identification on IMR cruises, (2) improve the use of taxonomic literature and other tools for fish identification, (3) improve recording of species on board, and (4) identify the needs for collecting material for future workshops and additional taxonomic research.

The 3-days workshops are held twice a year in direct connection with research vessel cruises: a pre-cruise part of 2 days duration and a post-cruise part of one day. They are held at IMR, the Chemical Laboratory and the Benthos Laboratory. The scientist in charge is Franz Uiblein (Fisheries and Stocks Research Group, IMR, Bergen; franz@imr.no). The workshops are financed by the IMR Fish Systematics Project (FU), with additional support from cruise or area-related IMR projects as well external projects. The language will be Norwegian.

The next workshop is planned for two days (pre-cruise) in mid-June 2007 and one day (post-cruise) in late August/beginning of September.

Participants comprise of technicians, scientists, students, and guests belonging to the staff carrying out the surveys, eventually supplemented with other interested persons. The maximum number of workshop attendees is 15.

Programme:

Pre-cruise part: General introduction; practical work with reference collection, use of literature and keys; on-board collection and preparation/preservation of taxonomic reference material; exercise with Fishbase.

Post-cruise part: Reporting of on-board identification work using recorded data, notes, photographs, and collected material.

During the cruises, remote advice and further guidance from workshop organizers can be provided.

Comments: The most important prerequisite for the workshop is the IMR Fish Reference Collection consisting of frozen material that is continuously updated through collections from cruises, and formalin/ethanol-preserved material that partly derives from earlier workshops and serves also for training or for scientific work.

The three workshops in 2006/2007 related to Barents Sea cruises but the second one in 2007 shall also be related to cruises in the North Sea and Norwegian Sea.

IMARES, The Netherlands

IMARES is building up an electronic collection of photographs and distribution maps together with key characteristics of Northeast Atlantic fish and benthos species, with the emphasis on species that have been positively identified during various surveys. This reference set (ZEUS) is freely available for all interested.

DFO, Canada

NAFO requires the production of a photo-ID key for elasmobranchs occurring in the contiguous NAFO area, where there is a substantial overlap in species. Such a high-resolution digital photo guide/ID materials for elasmobranchs might be produced as one component of a larger project in a North Atlantic (ICES/NAFO area - Carolinas to N. Africa) that could be interactive and web-based and linked to FishBase and other existing lists. The attached is a draft list of skates and rays compiled by David Kulka from the Caribbean around to Portugal, which should be pretty close to complete for that area.

For the Grand Banks to Greenland, we have produced a series of skate photo ID sheet handouts for our area and are looking closely at the difficult species (i.e. where tooth count is required to confirm ID for example). We have tens of thousands of good digital photos to draw on. Stephen Clifford is also doing some genetic work to aid in species identification.

The problem and solutions of course go beyond just identification materials. A key issue is appropriate taxonomic training for all those involved in sampling, including how to use the ID materials. We produce stuff on an ad hoc basis for both sea-going technicians and fishery observers. We are also producing materials for the fishing industry (Canada and NAFO) for selected species at issue. Another key issue is allocation of enough time on board to allow staff sufficient time to properly identify all species, including a specimen return system when there is uncertainty about the ID. Presently, all skates that come on board during DFO surveys are bagged and tagged and returned to the lab for confirmation and sampling. Spending more time on board interferes with the “traditional” way of doing things, because a few commercial species are sampled exhaustively to the exclusion of other species. Ways to resolve this are to design a sampling protocol that starts with species ID (focusing on how to handle difficult species such as some skates) and allocates maximum sample sizes for commercial species that still produce reliable information rather than the other way around.