C.M. 2005/Del.04/05/01

Report of the
Bureau Working Group Data Development Project

2–4 February 2005
ICES, Copenhagen

International Council for the Exploration of the Sea

Conseil International pour l'Exploration de la Mer

Palægade 2–4 DK–1261 Copenhagen K Denmark

Telephone + 45 33 15 42 25 · Telefax +45 33 93 42 15

www.ices.dk · info@ices.dk

Contents

1 Terms of reference

4DEL09 An *ad hoc* Bureau Working Group on Data Development (Chair: First Vice-President) will meet for 2-4 February 2005 (back to back with MCAP) at Council expense to:

a) Outline the data needed for ICES to produce the current advice and the type of advice, which are likely to be requested during the next five years;
b) Outline the database and software tools necessary to support efficient QA procedures and workflows;
c) Plan documentation and material to be available to the meeting suggested in STEP 1 of Document Del04/11/1;
d) Discuss and amend as appropriate STEP 1 to 4 in Del04/11/1;
e) Discuss and suggest priorities for ICES Secretariat work plan 2005 with regard to the IT and database work;
f) Develop a synopsis for an ICES IT strategy and business plan to be developed in collaboration with the ICES Secretariat before June 2005.

In addition to the Chair, the Group will consist of the Chairs of ACFM, ACE, ACME and CONC, the co-chairs of SGMID, the Data Centre Manager, and one additional member with expertise in database development and handling. The Group will work in close consultation with relevant Client Commissions, and will report on their view on data coordination mechanisms, IT needs and financing in relation to any proposed strategy.

The group will take note of constraints of Member Countries' national IT systems.

1.1 List of Participants

Niels Axel Nielsen (Chair)	Denmark
Peter Wiebe	U.S.A.
Chrisopher Zimmerman	Germany
Poul Degnbol	Denmark
Paul Keizer	Canada
Harald Loeng	Norway
Helge Sagen	Norway
Julie Gillin	(Date Centre Manager) ICES
David de G. Griffith	(General Secretary) ICES
Adi Kellermann	(Head of Science Programme) ICES
Hans Lassen	(Head of Advisory Programme) ICES

2 Outline the data needed for ICES to produce the current advice and the type of advice which is likely to be requested during the next five years

2.1 Introduction

Section 2 outlines the data needed for ICES to produce its current and future advice. This is determined by developments in societal needs, and by requests for scientific advice on human activity and environmental questions in the marine area. Development in marine science is responsive to these societal needs, and the scientific advances themselves also feedback to requests for advice. In order to give a prospective description of the data needed for ICES advice, it is therefore relevant to consider trends in societal needs for advice, and in trends in the scope and methodology of marine science.

Societal needs for advice and monitoring are to a large extent given by international conventions on the use and protection of the marine area.

The landscape and conventions where ICES is operating?

The most prominent conventions are the "Bonn Convention on Migrating Species" (1980) and the "Convention on Biological Diversity" (Rio, 1992). The primary target of the Bonn Convention is conservation of species. It includes i.a. ASCOBANS and ACCOBAMS (small cetaceans), the AEWA (African-Euroasian Migratory Water Birds, 1999) and the Trilateral Seals Agreement in the framework of the Trilateral Cooperation on the Protection of the Wadden Sea. Implementation of these conventions requires monitoring and reporting obligations for small cetaceans, birds and seals, conservation and mitigation measures and management plans, e.g. for seals in the Wadden Sea. The data comprise distribution maps of species across a vast scale of space and time, and additionally key parameters as indicators for the state of populations (growth, recruitment, abundance, diseases). They are very similar to the data required (and collected) to serve fisheries management.

The Biodiversity convention aims at conservation of ecosystems and sustainable use of resources as well as the equitable share of burdens and benefits of sustainable use. There are no explicit monitoring or research obligations. In addition to the ecosystem approach to management, member states are obliged to set up a network of Protected Areas. The data required cover distribution and abundance information on species and communities, as well as on resources and their use which involves a variety of socio-economic data. The information requirements are best described in the context of the DPSIR framework. Information is required on the Drivers, largely human activities, the Pressures, the activity that interacts with the environment, the present Status of the environment, and lastly the measurement of the Impacts and monitoring of the Response to changes in management actions.

Among the directive regulatory system of the EC, the directives most relevant for the marine sector are the "Flora-Fauna-Habitat Directive" (FFH) and the "Water-Framework-Directive" (WFD). FFH or NATURA 2000 comprises a network of protected areas using species and habitats as criteria for designation, listed in two annexes of the EC Bird Protection Directive (79/409/EWG). Its goal is protection of habitats and species. It explicitly includes monitoring and reporting obligations. Implementation will result in a broad spectrum of species, communities and habitat data, mostly on a high-resolution geographical scale. Under the Water Framework Directive (2000/60/EC), long-term processes are initiated including surveying the present state of European waters (ending 2004), establishing protection and mitigation measures (final management plan by 2009), and establishing a monitoring network (operational by 2006) to survey compliance with environmental objectives (by 2015). Additionally, chemical and physical (morphology, area) data will be required.

The European Commission is currently developing a marine strategy for extending management of European Seas beyond coastal areas. It is an ongoing activity the details of

which are worked out by the EMMA process (European Marine Monitoring and Assessment). Issues of concern are eutrophication, hazardous substances, shipping and oil discharges, resources management, biodiversity and habitat degradation, climate change, marine litter and radionuclides.

What are current mainstreams, trends in marine science (and society) relevant for ICES?

Rapid advances in technology such as miniaturization of sensors, continuous data collection technology and enhanced computing power set new norms for the amount of data used in marine science. Additionally, new areas are included in the marine science. Examples of mainstream issues are:

- Increased use of techniques for synoptic surveys
- Genetic techniques, e.g., in stock identification
- Assimilation and analysis of long-term data sets to detect changes in ecosystems and habitats (background: climate change)
- Climate change and oceanographic forcing (e.g., related: UV/CO2/carbonate system)

Since ICES is committed to the ecosystem approach to management, what is required to comply?

The ecosystem approach has recently been introduced as the preferred approach to management of marine activities via the Biodiversity Convention. There, the ecosystem approach is elaborated and specified in the "Malawi-Principles". While some principles are fairly general and geared towards practical problem management, others can be transformed into monitoring and management units as well as assessment criteria. Relevant elements in the present context (there are more) are:

- Ecosystem structure and function
- Spatial and temporal scales of these (and of Management)
- Change is part of the system
- Integration and balance between conservation and use
- Economic context

The resulting set of data is the most comprehensive among those considered hitherto since the inclusion of human use covers all economies from fisheries to pharmaceutical exploitation of resources. It also covers all aspects of fauna, flora and landscapes as well as the cultures of the local inhabitants. These data are a prerequisite for the undertaking the research required to support the ecosystem approach.

2.2 What type of data is necessary to fulfil advice expectations

2.2.1 Environmental Advice and Science to support it

The ICES Secretariat has several contractual agreements - existing and pending - to act as a data centre for a variety of types of environmental data (see Annex 1). In addition to these ongoing services, the ICES Data Centre (ICES-DC) performs data services specified in annual work programmes or agreements for OSPAR, HELCOM, and other clients.

Oceanographic data is needed to support the development and verification of circulation and ocean climate models that describe the physical forces acting on marine biological systems. As ICES and the client commissions move towards an ecosystem approach to the management of marine resources, the value of these data as well as their scientific interpretation increases. In particular, there will be a need to access and interpret the data based on the eco-regions that have been identified for management purposes.

This section provides an overview of the data that ICES-DC presently holds and that it expects to have in the future.

OSPAR

Joint Assessment and Monitoring Programme (JAMP)

The Secretariat has agreed to act as the data centre for specific information collected under the JAMP wherein data are submitted to and exchanged with ICES. Specifically, ICES-DC serves as the thematic data centre for the Coordinated Environmental Monitoring Programme (CEMP, Agreement 2001-8). Data submitted are in accordance with the Requirements for the Submission of National Comments to ICES when Submitting Monitoring Data (Agreement 2001-9). Furthermore, for CEMP data, OSPAR Contracting Parties ensure that when their national institutions provide data to the thematic data centre, there is agreement that the data may be released by the thematic data centre as soon as the centre has completed the necessary quality control procedures. Details of this arrangement are found in Annex 2.

Data from the CEMP provide the basis for general assessments, i.e., quality status reports (QSRs), under the JAMP. These general assessments include:

> "… a statement of the whole or part of the current knowledge of the health of the environment of a defined maritime area and its coastal margin. A complete statement includes an analysis of the region's hydrodynamics, chemistry, habitats, and biota with an evaluation of the impact of humans over space and time against this background of natural variability. All aspects of human influence on the maritime area concerned should be examined. This should include discharges, emissions, and losses of contaminants, nutrient, and radioactive substances occurring in that maritime area, or reaching it from the catchments draining into it or by airborne transport. It should also include inputs, concentrations, and environmental effects of contaminants, nutrients, and radioactive substances, dumping, transport, and the exploitation of biological and non-biological resources. The evaluation of the effectiveness of measures taken and planned for the protection of the marine environment and the identification of priorities for action should also form part of it."

The CEMP encompasses the following elements (see details in Annex 2):

- contaminant concentrations in biota and sediments
- measurements of biological effects
- data from the implementation of the Nutrient Monitoring Programme
- data on phytobenthos, zoobenthos, and phytoplankton species

with the possibility of expansion in future to cover data handling of:

- environmental concentrations and/or effects of further OSPAR substances for priority action (see details in Annex 2)
- monitoring related to Ecological Quality Objectives (EcoQOs)
- monitoring related to threatened and/or declining species and habitats;
- information on non-indigenous species

ICES has also been asked to contribute its considerable expertise in the assessment of these data with respect to trend analysis. Additionally, ICES has made significant contributions to the recent OSPAR initiative on integrated monitoring, i.e., chemistry and biological effects.

Ecological Quality Objectives (EcoQOs)

With respect to EcoQOs, ICES has collaborated with OSPAR on the development of a list of Ecological Quality Elements with associated objectives. ICES has provided extensive advice to OSPAR on the appropriateness and formulation of the EcoQOs (ACE reports, 2002 to 2004). The present list includes:

Issue	Ecological quality element
1. Commercial fish species	i) Spawning stock biomass of commercial fish species in the North Sea
2. Threatened and declining species	i) Presence and extent of threatened and declining species in the North Sea
3. Sea mammals	i) Seal population trends in the North Sea
	ii) Utilisation of seal breeding sites in the North Sea
	iii) By-catch of harbour porpoises
4. Seabirds	i) Proportion of oiled Common Guillemots among those found dead or dying on beaches
	ii) Mercury concentrations in seabird eggs and feathers
	iii) Organochlorine concentrations in seabird eggs
	iv) Plastic particles in stomachs of seabirds
	v) Local sand-eel availability to black-legged Kittiwakes
	vi) Seabird population trends as an index of seabird community health
5. Fish communities	i) Changes in the proportion of large fish and hence the average weight and average maximum length of the fish community
6. Benthic communities	i) Changes/kills in zoobenthos in relation to eutrophication
	ii) Imposex in dogwhelks (*Nucella lapillus*)
	iii) Density of sensitive (e.g., fragile) species
	iv) Density of opportunistic species
7. Plankton communities	i) Phytoplankton chlorophyll *a*
	ii) Phytoplankton indicator species for eutrophication
8. Habitats	i) Restore and/or maintain habitat quality
9. Nutrient budgets and production	i) Winter nutrient (DIN and DIP) concentrations
10. Oxygen consumption	i) Oxygen

This list, like the CEMP, is subject to frequent update as scientific knowledge and understanding develop. A number of other EcoQOs have already been suggested by ICES. For example, in 2003, the Working Group on Zooplankton Ecology (WGZE) evaluated possible biological indices of ecological significance for the fisheries and environmental assessment groups. The group proposed that an index to serve as an EcoQ status indicator (by satisfactorily meeting the 8 criteria) was the Zooplankton Species Abundance and Diversity Index. WGZE noted that the basic measures required are taxonomic identification and species counts, which also allow calculation of other suggested indices and relationships with other ecological measures of environment such as phytoplankton chlorophyll. WGZE recommended coverage of the seasonal changes (sampling at least once per month) and the use of a fine mesh to collect the samples (53 μm) in order to make sure that the zooplankton size classes relevant to fish larvae recruitment were collected.

Other expert groups and the advisory committees have made similar contributions to the development of EcoQOs.

HELCOM

MONAS - COMBINE

The ICES Secretariat is contracted by HELCOM to function as the thematic data centre for the HELCOM COMBINE Programme (oceanographic, biological community, and contaminant data and data products: see Annex 3). Since 1992, COMBINE has ensured that relevant monitoring data from different national programmes are shared and integrated through a common system. The programme provides for continuous international monitoring of:

- natural fluctuations in the marine environment
- amounts and effects anthropogenic nutrients
- levels and effects of contaminants in ecosystems.

The aim of the programme is to evaluate the influence of human activity on the Baltic Sea with regard to the effects of environmental policies.

Arctic Council

Arctic Monitoring and Assessment Programme (AMAP)

The ICES Secretariat has also agreed to serve as one of AMAP's thematic data centres. Specific tasks are contracted individually by the Arctic Council. Currently, there are no outstanding contracts. AMAP is in the process of integrating both monitoring and assessment activities in order to:

- produce integrated assessment reports on the pollution status and trends of the conditions of Arctic ecosystems;
- identify possible causes for changing conditions;
- detect emerging problems, their possible causes, and the potential risk to Arctic ecosystems including indigenous peoples and other Arctic residents;
- recommend actions required to reduce risks to Arctic ecosystems.

To prepare its assessments, AMAP:

- designed and implemented a coordinated monitoring programme to monitor the levels of pollutants and assess the effects of pollution in all compartments of the Arctic environment (atmospheric, terrestrial, freshwater and marine environments, and human populations);
- instituted an assessment process to produce assessment reports. The AMAP assessments are performed in accordance with agreed guidelines and are based on: i) data already published in scientific literature, ii) data obtained from AMAP's monitoring programme, and iii) traditional knowledge.

AMAP's priorities include the following contaminant groups and issues:

- Persistent organic contaminants (POPs)
- Heavy metals (in particular mercury, cadmium, and lead)
- Radioactivity
- Acidification and Arctic haze (sub-regional context)
- Petroleum hydrocarbon pollution (sub-regional)
- Climate change (environmental consequences and biological effects in the Arctic resulting from global climate change)
- Stratospheric ozone depletion (biological effects due to increased UV-B, etc)
- Effects of pollution on the health of humans living in the Arctic (including effects of increased UV radiation as a result of ozone depletion, and climate change)

Combined effects of pollutants and other stressors on both ecosystems and humans

2.2.2 Fisheries advice and science to support it

As a result of changes in the international environment described above as well as requests from clients, ICES' advice for fisheries is changing in several aspects. ICES is also receiving

more one-off requests for advice typically on very specific issues regarding a fishery or an area. These trends have implications for the data requirements for the advisory process.

Fisheries basis for advice Fisheries management is about managing fisheries - not stocks - and clients are asking for advice on that basis. This has implications for advice regarding mixed demersal fisheries in particular. It raises a number of issues with implications for data types required:

- The identity of fisheries – in order to characterise and categorise fisheries, information on vessel, gear characteristics and fishing practices is required. Fishing practices are currently derived from catch data – for lack of better information.
- Prognosis by fisheries – in order to make predictions on the basis of fisheries, recent catch and discards compositions and effort data are required on the basis of fisheries.
- Fisheries are dynamic: they respond to management measures, markets and changing fishing opportunities in terms of local abundance. When fisheries-based advice is required, catch (including discards) and effort data with fleet resolution must be updated with the same temporal and spatial resolution, and with no longer time delay as overall catch and effort data. (Currently, information on fleet level is compiled through a separate compilation exercise which is delayed and which is seen to some extent as a one-off project. For discard data, it may not be realistic to have ongoing observer-based sampling which can give new annual estimates on fleet level.)
- Resolution of catch and effort data on fleet basically requires vessel data to be available at least to national laboratories. This raises issues of confidentiality.

Ecosystem approach. The ecosystem approach requires data on environmental drivers for fish stock dynamics and productivity, and on fisheries impact on the ecosystem. A central problem here is that while such interactions may (at best) be known on general level, the resolution is too low or the processes are not sufficiently known to enable use in catch predictions or management strategies. The WGRED process will identify data needs in this respect.

Strategic advice. The change from annual catch predictions advice to management strategy advice (with short-term implications) does not have immediate data implications. However, the evaluation of management strategies must consider a range of issues which are not a part of ICES' usual disciplinary profile, namely economic and sociological information relating to management implementations and adaptations of fleets to regulations. Whether these issues should be addressed by taking expertise (and thus data requirements) onboard in an ICES process or whether ICES should seek cooperation with other organisations to supplement with this expertise has not been resolved.

ICES must also develop its technical capacity to provide advice by better capturing changes in the environment, fisheries and management in its advice, and by better addressing data quality problems:

Address changing conditions in fishery. ICES is requested to provide advice on emerging and changing fisheries, prime and high profile examples being the development of deep sea fisheries and the changes in mixed demersal fisheries as a response to restrictive management measures on cod. In order to advise on the consequences of such changes, ICES needs data for fisheries and areas that may not be covered by ongoing sampling schemes and surveys. As the situation is dynamic by nature, such data must be delivered with short delay. Specifically, for deep-water fisheries, catch and effort data are needed with very high resolution in time and space because these fisheries impact habitats which are on a smaller scale than squares.

There are changes in fisheries technology and in technical regulations, as well as technical creep in efficiency. ICES does not have any system to monitor these types of changes which do comprise important information for evaluation of the effects of technical regulations and of technological creep as required for effort management. Such monitoring would not result in large amounts of hard data, but rather descriptive data on current technological practices. This could be a new and important task for the Fisheries Technology Committee.

Address changes in objectives and management instruments. Commissions are changing their emphasis on management tools from nearly sole reliance on TAC type measures to strong emphasis on effort-based management and closed areas. Pre- and post-evaluations of such measures are requested, but require effort and catch data with high resolution in time and space. Access to VMS data is limited due to red tape issues. The evaluation of gear regulations should change from an emphasis on pre-evaluations based on idealised experiments to post-evaluations including the fisheries practice/adaptations to regulations. This means that there is a need for databases on selectivity which are not only based on specific experiments but also relate to fishery practices: this would involve considerable extension of existing selectivity data – both in volume and types of data.

Maintain (or establish) basic data quality. Data quality is increasingly of concern since many stocks are subject to very restrictive management measures. Data owners also influence the availability, utility and quality of basic data by withholding certain data or resolutions.

- Discard data have not been made fully available to ICES.
- Discard sampling is not done with sufficient resolution to enable dynamic changes in discard practices to be captured.
- Misreporting & Nonreporting has for some stocks reached an extent which makes assessments based on reported catches useless. Attempts to estimate the extent of mis- and nonreporting have limited value because such estimates cannot stand up to requirements for transparency. An increasing number of assessments are based on survey data only.
- Species composition data are poor in some fisheries – Baltic pelagic fisheries and deep water fisheries.
- There is increasing focus on sensitive species bycatch. Sensitive species may be so rare that a very high sampling density is required to detect bycatches that may still represent an important harvest.

Maintain data preparedness and fast track access to data. Advice is increasingly requested on an ad hoc basis for specific situations which are not covered by standard advice and which may not be analysed using existing data types. This means that ICES must keep its standard data updated on an ongoing basis with short delay, and that processes must be established to acquire special data in a fast track when non-standard data are required.

Data type	Use	Required resolution	Who collects	Who controls	Database	ICES access	Weaknesses
Landings	TAC forecast	By stock and fishery	ND (landings) NRI (biological info such as ageing)	ND NRI	Dis-aggregated data – NRI	Only aggregated data	Poor documentation of aggregation procedures on national level. Low tractability. Mis and non-reporting Poor species composition data Primary ageing data not available internationally
Landings	Area and effort regulations	+ high resolution space and time	ND (logbooks)	ND			
Discards	TAC forecast	By stock and fishery	NRI – observer programmes	NRI	NRI	Only aggregated and not all data available	Discard data not systematically assembled and reported to ICES – red tape issues. Low coverage
Discards	Area and effort regulations	+ high resolution space and time	Not collected	na	na		
Effort	TAC or effort forecast	By stock and fishery	ND (logbooks)	ND		Not available	Effort data not reported routinely
Effort	Area and effort regulations	+ high resolution space and time	ND (VMS)	ND		Not available	Not reported
Technology	Evaluation of gear measures and effort	By fleet and area	Not collected				
Environmental drivers	Ecosystem approach, catch predictions		NRI, U (Research projects)				
Fisheries impacts	Ecosystem approach		NRI, U (Research projects)				

ND: National directorates in charge of monitoring, control and surveillance

NRI: National research institutes

U: universities

2.2.3 Oceanographic Data

Oceanographic information should be used to support fish stock assessments. It is already used for modelling and predicting the fate and distribution of contaminants. This information is generally provided as assessments of ocean circulation and ocean climate. These assessments are derived from models and time series analyses that are based on extensive observations of temperature, salinity and currents. As the European community moves towards a requirement for ecosystem-based advice, information regarding lower trophic levels, i.e. phytoplankton and zooplankton, and essential nutrients will be needed. At present, ICES-DC does not hold this type of data.

ICES-DC currently maintains an extensive physical oceanographic database. While there is no specific contract for these data, they often provide the foundation for and/or supplement more specific data, such as contaminant data. Physical oceanographic data are essential to the development of an ecosystem-based approach to managing marine resources since they provide information on the essential forcing functions of ecosystem function and the foundation of ecosystem structure.

Physical, chemical, and biological oceanographic data are insufficient to enable the evaluation of environmental status and its change because of their sparseness in time and space (the oceans are grossly under-sampled). The use of hydrodynamic and biological models has advanced significantly over the past two decades. Models are increasingly able to provide the means to accurately hindcast existing data sets and are now beginning to provide a means of predicting future system states. An important aspect of modelling is the need for and use of high quality field data to set model boundary conditions and to verify model outcomes. In particular, assimilative modelling which depends heavily on field data has become an important tool in the ecosystem approach.

2.2.4 Integrated Bio-Physical Models – Data Requirements

Models must be compared to observations in order to identify model deficiencies, and improve the description and understanding of processes. Considering the recent changes moving the ecosystem towards new extremes, these modelling efforts eventually result in better simulation capabilities covering a wider range of climates, i.e. better predictive skill. Model simulations and developments gain additional value by close links to the ICES area. Sensitivity studies of the bio-physical system will help to evaluate the relative importance of processes, and enhance identification and understanding of processes and mechanisms. Furthermore, exploration of sensitivities supports the design of monitoring systems by suggesting locations and variables for additional observation. Besides classical sensitivity studies, the adjoint method is applied to provide the sensitivity of particular aspects (e.g. ocean climate) to the model state variables, to boundary and initial conditions, as well as to internal model parameters. The adjoint method provides a means of adjusting model outcomes by using field data to minimize the error between the model and field values. In the future, ocean observing systems will provide data needed for both hydrodynamic and biologic assimilative modelling .

Improved oceanic reanalysis fields for the ICES areas give dynamically consistent integrated datasets which are suitable for domain-wide analysis of key parameters and processes, and for improved model validation. These products could provide the best currently possible 4D fields for the area in focus. Furthermore, dependencies within the system illuminating forecasting possibilities, e.g. remote forcing of properties by North Atlantic winds and possible coupled interaction can be explored in such an area-wide, all-variable setup. This leads to improved understanding of the system's inherent predictability. Both regional operational ocean-atmosphere models and ocean-ice climate models will be utilized to demonstrate or to carry out regional reanalysis of physical, chemical, and biological variables. Various types of observations at their proper time and site can be utilized by the adjoint approach of data assimilation. Specific data to assimilate are e.g. sea surface data and vertical profiles.

Bio-physical models, like their purely physical counterparts, can ultimately benefit from the assimilation of Eulerian and Lagrangian data, especially when such models are used for hindcasts. Such bio-physical assimilation is being done with current technology. Data which could be assimilated into bio-physical hindcasts include moored current meter and bio-optical data, altimeter data, drogued drifter data, ADCP data, plankton and fish survey data.

Centralized access to the meta-data by investigators with a variety of skills and interests from many different institutions or agencies is critical to the study of ecosystem processes. Policies and procedures for uniform data management and rapid exchange are essential to foster linkages between studies on different time and space scales. Also, timely, free flow of information to all investigators is vital for planning experiments, data analyses, and modelling

efforts. It is also necessary to determine the relationship between and among various biological, physical, and chemical data which have been collected.

Compilation of a bio-physical meta-database of varied, disparate historical data sets (hydrography to higher predator abundances) would offer several benefits to any ecosystem-related project. The meta-data should prove extremely useful to individuals undertaking field work or designing field sampling programs, designing process studies, validating simulation models, or collecting data for retrospective analyses. The meta-data should facilitate comparisons of past, present and future biological processes and their coupling to the physical and biological structure and variability of the environment. Decadal data sets of biotic and abiotic variables would allow the question of climate scale variability to be addressed. Moreover, it would allow individuals to monitor changes and provide baselines for formulating and testing hypotheses to advance understanding of the interactive processes that regulate ecosystem production. It is anticipated that the products of this approach will become more valuable as time passes. Finally, the meta-data would support regional comparisons, and better definition of regional differences in forcing and response within the ICES areas and of the extent to which long term changes are regionally focused or coherent throughout the area.

Data from models

The recent availability of atmospheric reanalyses like NCAR/NCEP (The National Center for Atmospheric Research/National Centers for Environmental Prediction) and ECMWF's ERA-40 (European Centre for Medium-Range Weather Forecasts) of the last 40-55 years makes it possible to do model-based ocean reanalyses on the North Atlantic up to the global scale for the same period. Such hindcast experiments may be pure ocean model simulations or may incorporate observations by data assimilation. In both cases, model-generated data can supplement observed data with valuable information.

Requirements for data storage are substantial. As an example, to save daily averages from a 20 km model for the North Atlantic requires 45 GB per year. Storing and exchanging datasets of this magnitude is not trivial. Centralized storage of several such datasets is not practical.

Results from a hindcast experiment may take years to be published. It is therefore difficult to keep track of the work being done. A centralized meta-database with information on modelling groups, the experiments they have performed including details on the model set-up, publications, availability of the results, contact persons, etc. could be useful for the marine research community.

3 Outline of the database and software tools necessary to support efficient QA procedures and workflows

3.1 ICES Data Centre holdings and systems

Historically, the ICES community met at the Secretariat, and collated and shared data during these meetings. This required local data storage facilities which formed the basis for centralised ICES databases. Value was added simply by aggregating data from different sources. New databases evolved with little coordination – usually simply due to some group's desire for such facilities.

At present, ICES-DC therefore holds a number of uncoordinated databases with very different inventories. ICES-DC stores and processes large amounts of data. With the exception of hydrographic and some contaminant data, none of the datasets comprises a comprehensive set of a specific data category, but rather well-defined subsets. For example, for fishery-independent data used for assessments, there is a functional database for 3 surveys in 2 areas out of possibly ten's of different surveys in 8 areas. Some of the databases have been developed according to agreements with partner commissions/clients.

ICES has worked within three disciplines: Environment, Oceanography and Fisheries. These three disciplines have co-existed with little interchange. This is reflected in dispersed, widely differing data formats, systems and processes. For example, separate discipline- or function-specific systems are used such as DATRAS for trawl data and REC12 for preliminary catch statistics; fixed environmental data formats versus free oceanographic formats; and, various degrees of data verification and various approaches to data correction.

Marine science disciplines should increasingly be seen as interrelated, interacting elements of the ecosystem approach. Information from the three disciplines synergizes when considered collectively. For example, integrated data is necessary for cross-data comparisons. Thus, there is incitement to link and integrate the data via common programs and processes.

An integrated system is also more efficient to operate, maintain and support than several independent, specialized systems serving the same purposes. However, development of integrated systems usually proves to be more challenging.

ICES-DC is moving toward integrated systems. New multi-disciplinary databases such as DOME (Database on Oceanography and Marine Environment) are being developed to supplement and/or replace existing discipline-specific, specialized databases.

ICES-DC has also developed a number of tools for validating, processing and presenting data. Some examples are:

- **RECO**: integrated reference code system used by all disciplines and new systems. Referenced by several databases.
- **DATSU**: data checking program for universal application. Currently used by DATRAS. Planned for use by InterCatch and DOME.
- **OCEANPC**: QC, visualization, management & documentation of Oceanographic data
- Formats for data reporting: free formats are coming!
- Web interface to preliminary catch statistics (under development)

Currently most tools are used in-house for maintaining databases and supporting users. But more recent developments focus on systems for working group users as well as data submitters in accordance with the growing demand for immediate, direct access to data.

All databases – existing and planned – are listed in the following table. A general idea of the amount of data is indicated by their size (no. of records/stations and file size). Previously, databases were structured hierarchically in ASCII files. Recent developments are all as relational MS SQL or Access databases. More detailed information is found in Annex 4.

Discipline	File type	Description	Format	Size	
				No. stations	MB
Environmental	Contaminants in Biota		ASCII files	7,000	72
	Contaminants in Sediments		ASCII files	6,000	26
	Contaminants in Seawater		ASCII files	32,000	54
	Fish Disease Database		ASCII files	4,000	60
	Biological Community Data		ASCII files	1,000	5
	Intercalibration results		ASCII files/paper	3,000	234
Oceanographic	High resolution CTD	conductivity, temperature, depth	Indexed ASCII files	184,000	4123
	Hydro-Chemistry data	incl. low resolution CTD	Indexed ASCII files	930,000	1097
	Surface data	temperature & salinity during steaming	Indexed ASCII files	1,757,000	76
	Pump data	stern-inflow samples	Indexed ASCII files	531,000	3
	ROSCOP	Cruise data	ASCII files	31,000	14
Fishery	DATRAS	IBTS, BITS and BTS trawl surveys	MS SQL Server	18,000	530
	Rec-12	preliminary Catch statistics	MS SQL Server	variable	2
	Assessment Summary	fish stock assessment summary tables	Access	n/a	2
	NewIFAP	Catch aggregated to international level	MS SQL Server	n/a	under development
Integrated (multi-disciplinary)	Accessions	Log and audit trail of all submissions	Jet engine/Access	n/a	2
	ICES Integrated Inventory	Summaries of data holdings by area, time and parameter	MS SQL Server	n/a	440
	DOME	3-phase project to integrate oceanographic, environment, and, possibly, fishery data	MS SQL Server	n/a	under development

3.2 ICES role in relation to databases collected and held by large international projects

Since the International Decade of Ocean Exploration in the 1970's, there has been an increasing number of large disciplinary and multi-disciplinary projects of regional and global scope. Examples are the Coastal Upwelling Experiment (CUE), the Warm-Core Rings Program (WCR), the World Ocean Circulation Experiment (WOCE), the Joint Global Ocean Flux Study (JGOFS), the Global Ocean Ecosystem Dynamics Program (GLOBEC) and CLIVAR (part of the wider World Climate Research Program - WCRP). Most of these programs were international in scope, and generated large, complex datasets. In the ICES region, there have been other (multi-)disciplinary programs involving multi-investigators/institutions such as the European Subpolar Ocean Program "Variability of Exchanges in the Northern Seas" (VEINS), Trans-Atlantic Study of *Calanus* (TASC), and Wind-driven diapycnal mixing in the deepwater of the Baltic (DIAMIX).

In the early years, processing and storing data required mainframe or mini-computers, and often substantial portions of data were not effectively archived. Today, almost all data can be handled electronically; data storage is becoming easier, although improved sensors and data acquisition systems are still taxing the present data storage systems. Most nations or unions sponsoring marine research now insist that data is made public within a limited time (6 months to 2 years) and, for most large programs, data management is an integral part of the program and provides a mechanism for making the data public. There is usually a mandate for both metadata and actual data to be archived in a national oceanographic data centre (NODC). In general NODCs do not, however, have a good record of archiving all multi-disciplinary data generated by the projects - especially not data associated with biological process studies. Thus, a major challenge facing the oceanographic community is to develop the means to preserve data from large national and international programs in their entirety.

A second problem stems from the many small federally funded projects lasting from one to five years which produce useful short-term data sets. While the meta-data are now often required to be submitted to the sponsoring agencies, the actual data often reside only on investigators' computers and can be eventually lost.

In the past, ICES has served as data manager and data repository for oceanographic research programs such as the Fladenground experiment (FLEX '76), VEINS, and the Skagerrak Experiment (SKAGEX). Details of some such projects are given in Annex 5.

With regard to projects, there is the basic question of where their data will ultimately be archived. ICES as suggested above could be part of the solution. A significant reason for ICES to host or data is the long-term stability of the database systems and the ICES organization itself. This can be achieved for current and future projects by including ICES data services in the project proposal. For historic data, the problem is much more complex and must be considered on an individual project basis. This was, of course, outside the mandate and scope of the BWGDDP.

3.3 Data available outside ICES

Data sources outside ICES-DC

Some scientific data are becoming available through the Internet. Data are often stored in relational or hierarchial databases with access tools freely available on Internet. Data can be screened using the available tools (often an internet browser) or downloaded for inclusion in local analysis tools. Large amounts of data are, however, still protected from free download. Sometimes they are available through registration at the website - but not always. Users may need to contact the data owner to request access to the data.

Web portals

Web portals are widely used as focal points for data. A "web portal" is a website that provides a starting point, a gateway, or portal to other resources on the Internet or an Intranet. Intranet portals are also known as "enterprise information portals" (EIP). Portals typically provide personalized capabilities to their users. They are designed to use distributed applications, different numbers and types of middleware, and hardware to provide services of a number of different sources. In addition, business portals are designed for collaboration in workplaces. A further business-driven requirement of portals is that the content must work on multiple platforms such as personal computers, personal digital assistants (PDAs) and cell phones. A number of relevant web portals are described in Annex 6.

Metadata

Data can be made available through metadata catalogues (examples in Annex 7). Metadata describes the attributes of an information bearing object (IBO), namely, a document, dataset, database, image, artifact, collection, etc. Metadata acts as a surrogate representation of the IBO. A metadata record can include representations of the content, context, structure, quality, provenance, condition, and other characteristics of an IBO. Metadata represents the IBO to a potential user for discovery, evaluation of fitness for use, access, transfer, and citation. Examples of metadata format include:

- MARC format used by the library community
- Content Standards for Digital Geospatial Metadata developed by the Federal Geographic Data Committee
- Directory Interchange Format (DIF) used by NASA's Global Change Master Directory
- Government Information Locator Service (GILS)
- Dublin Core set of attributes for electronic resources developed with the lead of the Online Computer Library Centre (OCLC).

The ISO 19115 standard defines core metadata components, recommended components and allows community based profiles to be described as extensions to the standard.

Parameter dictionaries

A normalised database is based on parameter codes. Parameter codes need a parameter dictionary. A set of codes may be managed in many different ways, for example in a text file, a spreadsheet or a database, and describe the parameters in the related database. The managed set of codes is called a parameter dictionary.

4 Vision of database architecture and software that will support the work of ICES

The regulatory framework in which ICES works, the needs of users - especially the foreseeable needs of the advisory process, and finally the current status of ICES databases provide the background for this section.

BWGDDP agreed that collating data on an international level still adds value to the data and serves the scientific community, as it has for the last 100 years. The ICES-DC is considered to be the ideal body to organise collation and quality control of data. Also, ICES Working Groups and committees regularly require data as basis for their work and advice. The procedure leading to ICES' advice must be highly transparent. Thus, all data on which advice is based must be completely and properly documented to enable reproduction.

Advice is increasingly delivered in the ecosystem context, which requires a higher level of data integration. Technical possibilities in terms of database structures have significantly changed over the last decade. They can now reduce redundancy, enable access to data through the internet, administrate individual user access patterns, and provide powerful public search engines, to name just a few features. In principle, it is no longer necessary to store all data in a central database to ensure data integrity and quality. With this in mind, BWGDDP has developed the following vision.

a. ICES will remain a focal point for marine data in the North Atlantic.

As an intergovernmental, permanent organisation promoting science and science coordination, ICES is already in an excellent position to collate and aggregate international data. We envision ICES-DC's future role as a data hub easily accessed by the science community at large, but especially by working groups and committees. Data could either reside physically at ICES (centralised system) or be dispersed in various databases elsewhere (distributed system). Advantages and disadvantages of both types of systems are given in Annex 8. Distributed systems are preferred, but there are a number of exceptions as follows:

1. *Databases already residing at ICES which require little maintenance.* These databases are valuable for their contents and associated expertise, which in turn facilitate their supplemental use as sample data structures and testing material. This is especially applicable for databases used to fulfil agreements between ICES and partner commissions/clients. In such cases, data should be as complete as possible.
2. *Aggregated data used as the basis for advice.* With regard to transparency of the advisory process, all data needed to reconstruct advice should be documented. Today, this is done, for the most part, in the form of printed reports. Electronic storage would be preferable.
3. *Anonymous data may need to be maintained centrally.* Some of the national data currently used in the advisory process contain legally or economically-sensitive information and are therefore brought to the Working Group as confidential information. Examples include vessel monitoring system tracks for individual vessels, and unreported/misreported catch for individual fleets. Some of these data need to be stored (anonymized) to be available for documentation and re-runs.
4. *Valuable data at risk of extinction.* Some data produced by small projects might be lost after the project ends. ICES should offer to support the storage of valuable project data which might otherwise be lost. This can be done either by transferring the data to centralised databases, or by linking to or developing a (distributed) database elsewhere. In both schemes, the data must be well-defined and well-documented. In the case of transfer to a centralised database, only storage would be offered. User facilities to access, process and/or extract may not be available. Long-term storage via distributed systems is discussed in the next section.

b. ICES will create a portal serving as a hub for distributed data *(data not physically stored in-house).* In the future, much of the data needed within the ICES community may be stored in distributed systems. ICES will thus promote the development of such databases (e.g. databases for ICES coordinated surveys currently stored in spreadsheets on an individual coordinator's computer), as well as the development of web interfaces for the databases. To do this, ICES will

- Provide assistance to set up databases at national laboratories or coordinating bodies, according to ICES standards. This could be done by providing a template database, which would then only need to be adapted to the individual survey needs. Comprehensive documentation and platform independence is essential. Open source software is strongly recommended for most situations. (See below.)
- Define standards, transfer protocols, and interfaces to access distributed databases through the ICES portal, and encourage managers of distributed databases to use them. In the future, whenever possible, ICES-DC development work should focus on a distributed configuration, rather than new centralised databases. Some of the necessary software might already be available, such as DiGIR (Distributed Generic Information Retrieval).
- Develop quality control rules and procedures. This includes data consistency checks; means of communicating error and fix reports between data users and sources; and procedures to promote timely correction of errors.
- ICES will contribute data to other relevant distributed databases and provide links to those through our portal.

Timely, open access to data will add value and is thus crucial for the usability of the portal. ICES must develop a clear strategy on how to access distributed data. This involves policy as well as technical considerations. First-of-all, ICES must update its data policy, which is more than a decade old and no longer fully reflects the way the scientific community handles data. A proposal for updating the ICES data policy will be drafted by SGMID following its meeting in April, 2005. The policy must acknowledge that ICES is the repository and/or distributor of the data, but not the owner with whom data rights ultimately remain. The policy may require negotiations with data owners and, since there are variable proprietary periods, it will undoubtedly presume a complex, secure system for administrating access.

On the technological side, distributed data can be accessed in many ways:

i.) External sources could be integrally linked into the ICES system.

ii.) In cases where part of the data is not publicly available, ICES could access or upload the available subset of the data (e.g. national databases), and channel it to the requester after validation.

iii.) Aggregated and/or meta-data could be integrated into the ICES system. The ICES portal would then deliver only pointer data, guiding the user to the relevant data sources.

BWGDDP agreed that it would not be useful to develop search engines for data: excellent commercial engines are universally available. ICES-DC should focus on accessing raw data. However, data which are not well-defined or readily available on the web might be difficult to locate or access data through commercial search engines. For those data, ICES could provide valuable storage of their meta-data and location.

Regarding **prioritisation**, ICES should put special emphasis on the access of data required for the advisory process. This includes long time series, throughout which consistency is assured. While the collection of fish stock assessment input data will certainly be promoted by the development of the new ICES fishery assessment programme (InterCatch), more effort in gathering environmental data is needed.

c. ICES data, documentation and systems will be universally and efficiently accessible to the science community..

1. Data & Documentation will be freely and openly available as soon as possible after collection.

Fast(er) turnaround times. Data should be available as soon as possible after collection. To prevent backlogs of the data, which ultimately cause time lags between when the data are reported and when they are available in useful form, data quality control procedures should be altered to make it possible to serve data almost immediately. The data should be supplemented with quality information on the application and outcome of quality control measures.

With the exception of data marked confidential, all data should be publicly available within the legislated time period. Aggregated data used in the advisory process must be publicly available immediately.

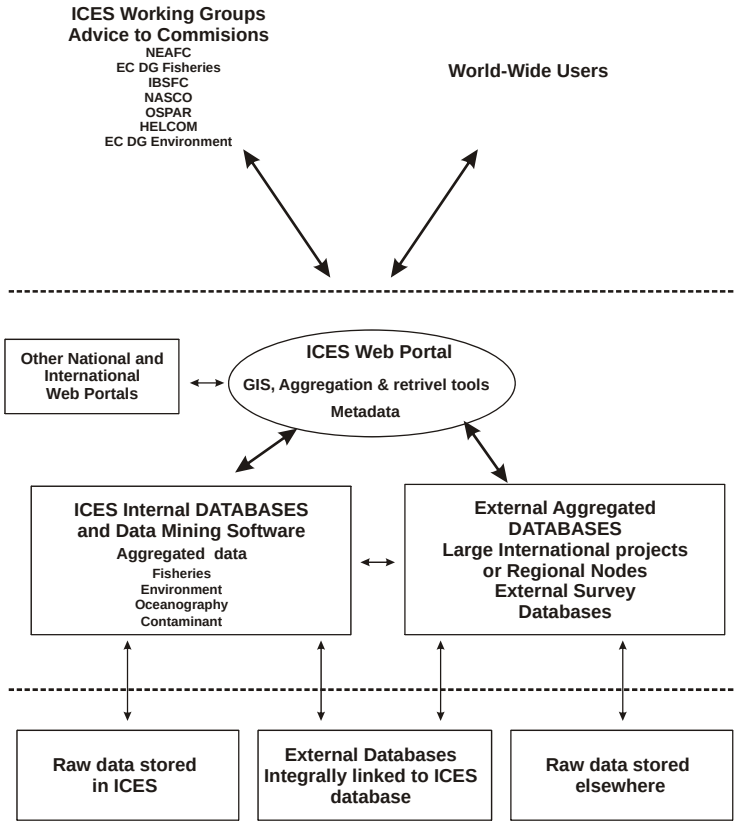
Most of the vision presented here requires **restructuring** the way in which data are currently handled by ICES-DC. Additional **funding** might be needed to complete these tasks, and should be negotiated with the partner commissions. ICES should encourage members and national institutes to take part in the restructuring and development of databases and tools. This would save resources for the Secretariat, and strengthen the ICES community by closer cooperation.

2. Systems will be readably operable

The ICES web portal will become more attractive to the science community.

ICES DC as well as expert group's own tools should be platform independent and use open source software whenever possible to minimize technical and financial limitations to their use. Furthermore, tools shall be designed corresponding to their user computer literacy.

Useful facilities for providing and presenting raw and processed data will be provided. For example, mapping tools/GIS systems for the visualisation of data; aggregation/gridding, validation, and browsing tools.



5 Stakeholders Involvement

A fundamental precept is that ICES should have access to the data used to formulate advice, either at ICES or via assured external data sources. It is recognized that ICES needs more data that is reliable, and easy and flexible to access in order to fulfil its advice and science obligations. In Section 2, we discussed the need for data supporting ICES fishery management advice to be more timely, flexible and of a higher quality. This section also address data needed to develop integrated ecosystem advice. Even though advisory tools for the ecosystem approach are still developing and data requirements are still uncertain, we already have a broad idea of the growing need for data.

Section 4 outlines a vision for database architecture and software tools to make vast amounts of data useful and available to the scientific community and for advisory purposes. The design and tools must take account of current and future data policy requirements on confidentiality and security as set by (national) data owners. To gradually develop such an infrastructure, and to fund and invest in the relevant technology, ICES and members state institutes need firm commitment from ICES, data owners in members states, and Partner Commissions on a data development strategy.

5.1 External Stakeholders

Therefore BWGDDP recommends meetings between Partner Commissions and ICES to discuss:

- Data needs for current and future advice and research

- Availability, reliability and accessibility of such data
- Overall design of databases and software (centralised/decentralised/data policy/confidentiality)
- Tasks and responsibilities in a roadmap to improve data quality and accessibility
- (Reconfirm if necessary, data agreements between ICES and partner commissions)
- Funding plan

Plans for the meeting and documentation

BWGDDP suggest that ICES makes use of the planned for MCAP-Commissions Coordinating meeting in April. All (or most) partner commissions will attend, so the initiative can be introduced there. Assuming the support of the Commissions, BWGDDP suggests arranging a number of bilateral meetings. At these meetings, particular problems in relations to specific regions and/or Commissions can be detailed and discussed.

Primary topics for the fishery commissions are:

- Reliability and accessibility of data
- Overall design of databases and software (centralised/decentralised/data policy/confidentiality)
- Possible synergies with software developments and databases in other organisations

The main topics for the environmental commissions are:

- Data needs for current and future advice and research
- Availability and accessibility of data
- Overview of completeness and utility of certain groups of data
- Working relations and future role of national data centers and ICES DC.

Documents for the meetings will be prepared by the ICES secretariat based on this report.

5.2 Internal Stakeholders

The users' view of which data and systems ICES should supply is important. One way to determine this would be a questionnaire to all ICES expert groups. The responses could then be used as a guideline for further developments. Questions should be on a general level, for example:

- To what extent do you use ICES' data in your ICES work? in other work ?
- How should it be improved in order be more useful ?
- What kind of data are you missing?
- What data processing software is most useful ?
- What data processing software are you missing ?

BWGDDP recommends that CONC, at its September 2005 meeting, discuss how such a survey could best be arranged. One possibility is to install a "hearing process" for the ICES strategy and business plan within ICES committees. In October 2005, the Council is expected to discuss and agree to the overall content of the ICES data strategy and business plan. In practice this could be coupled with a questionnaire on data and software needs, pending its completion by the ICES secretariat in January 2006 (tabled at Bureau 2006).

6 ICES data strategy and business plan

In collaboration with the ICES Secretariat, the BWGDDP should develop a synopsis for an ICES data strategy and business plan before June 2005. The strategy should encompass all data activities - not only databases and related software.

The BWGDDP recommends the following actions and process:

- Survey the usefulness and prioritization of ICES-DC data holdings by the ICES scientific community, and discover other needs which ICES-DC should fulfil. (To be developed by BWGDDP at May meeting and presented to CONC at their autumn meeting.)
- WGMDM to provide instructions, training and assistance to institutes that do not conform to WGMDM guidelines, and to advise on processing and prioritization of ICES-DC's activities on oceanographic data.
- On the basis of the recommendations, vision and priorities in this report, the ICES Secretariat to draft a data strategy.
- In late April/early May, BWGDDP will receive MCAP feedback, a Data Policy proposal from SGMID, advice on processing and prioritization of ICES-DC oceanographic data activities, and a proposal for a data strategy from the ICES Secretariat.
- The BWGDDP will meet for one day in May 2005 (preferably back-to-back with another meeting where the majority of the participants are already present) to comment and feedback on the proposed data policy and strategy, so that the Secretariat can complete a final draft before Bureau meeting June 2005.

7 ICES Data Centre Workplan 2005

The BWGDDP commends the clarity of ICES-DC's workplan, and supports the use of resources for development of software to aid fishery and environmental assessments. The main activities which are allocated the majority of resources are:

- InterCatch (New ICES Fishery Assessment Program): aggregation and accessibility facilities.
- DOME (Database on Oceanography and Marine Environment) integrates data from all 3 marine science disciplines, thus advancing the ecosystem approach.

Other important endeavours include:

- DATRAS integrates 3 trawl survey databases: BITS, IBTS and BTS.
- Accessions, DATSU, and RECO: generic utilities which support all systems – regardless of discipline.

The BWGDDP support this prioritization of activities. They also recognize the importance of and need to maintain ICES-DC's role as thematic data centre for OSPAR, HELCOM and AMAP, as a means towards realizing the great potential within the marine environment area. All activities are listed in Annex 9.

The BWGDDP also acknowledges past contributions of data activities in the Oceanographic area. However, today's proliferation and accessibility of hydrographic databases in general, and local application of quality control in specific, may render ICES-DC's processing of oceanographic data redundant. Furthermore, speed of data availability is now a major quality criterion: extra processing slows data availability.

The WGMDM is thus requested to optimize quality control by providing instructions, training and assistance to institutes that do not conform to WGMDM guidelines. Furthermore, their advice on processing and prioritization of ICES-DC's activities on oceanographic data is requested.

ICES and project participation and outsourcing

ICES has also participated in a number of projects and collaborations. Three oceanographic projects are described in Annex 5. For all three projects, ICES participated as the data centre responsible for quality control and distribution of data.

Other models of collaboration are also being used:

- ICES-DC is currently developing software supported in part by IMR, DIFRES, OSPAR, HELCOM, CEFAS. In one case, we are also considering collaborating with another institute so that to extend and enhance a system to meet our needs – rather than starting from scratch.
- In 2003, the ICES Secretariat outsourced some activities related to the maintenance of the fish catch databases by entering into an agreement with EuroStat. This model has proven beneficial for both parties.

8 Summary and Recommendations

If ICES is going to continue its role as the organisation that coordinates and promotes marine research in the North Atlantic, it must cope faster with broader, larger datasets in the future. At the same time, ICES must maintain its vital roles and services such as thematic data centre to partner commissions, advisor to fishery commissions, and data provider for the scientific community.

Commissions as well as the scientific community have growing expectations for accelerated access to data. Technically, this is possible by advances in IT and database technology such as distributed databases. But in order to ensure appropriate, advantageous data access, ICES must strengthen its role and capacity by exploiting new possibilities and technologies. Practically speaking, to achieve this ICES first must evaluate and prioritize client and user data needs together with commissions, national data centres and users; then determine the best way to achieve them via strategic planning; and, last but not least, build up capacities to meet the challenges.

In sections 2 and 3, current needs were described and future needs were predicted based on agreements, expectations, technological possibilities, etc. These are summarized below. The needs provide the basis for the vision described in detail in section 4 and outlined below. Sections 5 and 6 recommend specific actions, namely stakeholder involvement and strategic planning, to follow-up on the vision. These are also briefly outlined below.

Environmental data

ICES has agreements with a number of agencies to provide data management services for environmental data. In the past, these data were almost exclusively used to provide advice to client commissions. Demand for assessments is increasing with the move toward more integrated advice and ecosystem-based approach to management of human activities in marine waters. Assessments are based on the combined use of actual data and sophisticated models. Information includes variables directly impacted by human activity such as contaminant levels and nutrient concentrations in coastal waters, as well as variables that facilitate the assessment of ocean circulation and climate.

ICES serves as a thematic data centre for the OSPAR and HELCOM monitoring programmes, and stores and processes large amounts of data. However, individual datasets are sometimes uncoordinated and do not comprise a comprehensive set of the respective data type.

Fishery data

ICES also has contractual agreement with a number of Fisheries Commissions for advice on fisheries management for a large number of stocks and fisheries. This advice requires *timely,*

reliable fishery data time series disaggregated on matching time and area bases. In recent years, requests for advice have often involved a multitude of management measures such as direct effort control, technical measures and catch limitations. Furthermore, advice on ecosystem impact of fisheries requires high-resolution geographical information – due both to the nature of the problem and to the requested advice. Finally, data dependability has deteriorated for some stocks and must be restored.

Only top level aggregated fishery data are held by ICES, with little area, fishery or time breakdown. This imposes a lack of flexibility which can hamper stock assessment work at the working groups. ICES does have some databases for fishery-independent data that are used for assessments, for example, the DATRAS database holds data for 3 surveys in 2 areas - out of possibly ten's of different surveys in 8 areas. Some of these databases have been developed according to agreements with partner commissions/clients.

Integrated and distributed data

The Data Centre is moving toward integrated systems. New multi-disciplinary databases are being developed to supplement and/or replace existing discipline-specific, specialized databases.

For widely dispersed or voluminous data, distributed databases may be more feasible. Here, ICES is uniquely positioned to serve as a data hub, guiding data owners to structure and organize their data in compatible ways, and helping data users locate and access the data through the use of aggregated and meta-databases, and links.

ICES should also explore the possibility of using its database and data management expertise to host and archive large international project data to ensure that valuable data are preserved and documented. ICES could thereby also assist and promote data integration and sharing among these projects.

The vision

a. ICES will remain a focal point for marine data in the North Atlantic.

b. ICES will create a portal serving as a hub for distributed data.

c. The ICES web portal will become more attractive to the science community.

Recommended next steps

To embark on this vision, the BWGDDP recommends 2 parallel activities:

i. Develop a Data Strategy and Business Plan, including a new data policy, based on the vision and rationale presented herein.

ii. Clarify and agree on short-term and long-term actions with stakeholders such as partner commissions and the ICES science community.

Annex 1: Data Activities on ICES Secretariat's Agreements

Sponsor/Partner	MoU/ Contract	Date	Activity	Man-days	Deadline
AMAP	AMAP data handling	-	-	0	-
CEFAS	CEFAS-DOME	Feb-05	Design, development & maintenance of integrated database	0	31-Dec-05
CEFAS / RIVO (6th Framework programme)	ICES-Fishmap	04-Jun-04	Delivery of data (IBTS)	(see requests)	
			Review work, prepare web-site	2	
			Website assistance	20	01-Oct-05
			Phase 2 requires more DC input since the atlas will be transferred to our server with a direct link to DATRAS.	tbd	starts in 2006
EC (DG-Fish, DG-Environment)	Basis for cooperation w/ EC	2003	Statlant 27	-	-
			Evaluate progress	?	01-Jun-05
EEA	EEA MoU	17-Aug-04	regular data exchange (once per year)	6	26-Nov-05
EuroStat	STATLANT programme	27-May-03	Receive & vet data	0	
			Continue REC12	20	
			Publish FishStat+ databases	10	
			Hold annual review meeting (secretariats)	3	asap
			Hold joint statistics meetings (every 2-3 years)	?	27-May-06
FAO	FAO MoU	27-Mar-96	-	n/a	-
GIWA			one-time task performed by Neil & Henrik	n/a	completed
HELCOM	HELCOM COMBINE Data	14-Sep-04	data processing	0	
			conversion of 2.2 data	60	
			Free formats		
			data products		annually
	HELCOM MoU	31-Aug-99	-	n/a	-
IBSFC	IBSFC MoU		-	n/a	-
IMR	IMR-NewIFAP	Nov-02	NewIFAP	595	31-Aug-04
IOC	IOC MoU	Sep-04	cooperation on data	15	

Sponsor/Partner	MoU/ Contract	Date	Activity	Man-days	Deadline
NASCO	NASCO MoU / Contract	22-Mar-01		0	
NEAFC	NEAFC MoU	09-Mar-04	-	n/a	-
OSPAR	OSPAR MoU	21-Sep-95	JAMP data handling		
	OSPAR Work Programme	2004	GIS evaluation	25	
			CEMP data handling	?	
			CEMP data products	20	31-Dec-05
PICES	PICES MoU	24-Nov-98	-	n/a	-
QUASIMEME	QUASIMEME (no contract)		Advisory Board	5	
			Website assistance	5	
SAHFOS	SAHFOS Digitisation of Plankton data	18-Aug-04	review reports; assess feasibility of further digitisation; write new contract	3	30-Jun-05
SCOR			Scientific Cooperation	n/a	
UNEP			Random data deliveries	n/a	-

Annex 2: Agreements between the ICES Secretariat and OSPAR

A. Extract from the ICES-OSPAR Memorandum of Understanding of 21 September 1995:

Data handling

2. The Council's secretariat will serve as data centre for data collected under the Commissions' Joint Assessment and Monitoring Programme (JAMP) for monitoring the marine environment, such as:

    - data on contaminants observed in the compartments waters, sediment and biota of the marine environment; and
    - data resulting from biological monitoring (including biological effects monitoring).

    The data centre will:

    a. receive monitoring data submitted under the Commissions' programme and according to the Commissions' format agreed upon;
    b. transfer that data onto suitable information-technology equipment provided by the Council and undertake its validation, in cooperation with Contracting Parties concerned;
    c. prepare the data sets to be used for assessment in time before the relevant assessment meetings;
    d. in cooperation with the Commissions' secretariat, arrange for the assessment of the data sets by appropriately qualified subsidiary bodies of the Commissions;
    e. prepare appropriate data products including output of statistical analyses and maps; and
    f. provide the Commissions with master copies of the assessment products in computer-readable form.

3. .

    a. Where a database is maintained exclusively for the work fo the Commissions, the Commissions will be responsible for the costs (including a reasonable share of overheads) associated with that data base.
    b. Where a data base is maintained partly for work of the Commissions and partly for the work of the Council other than their assistance to the Commissions, the costs (including a reasonable share of overheads) associated with that data base will be divided between the Commissions and the Council in proportion to the agreed assessment of the extent of the use made of it for each purpose.
    c. Data submitted for the Commissions' Joint Monitoring and Assessment Programmes will be handled in accordance with the Council's standard data security procedures for environmental data, namely, that raw data will not be provided to third parties without the prior permission of the originator (that is, the country or laboratory that has submitted the data).
    d. Data products may be provided to third persons after adoption of the relevant products by the Commissions.

B. Extract from the 2005 ICES Work Programme

9 Evaluation of a geographic information system to enable a geographic presentation of the results of the assessment of environmental data.

9.1 This would evaluate the potential of suitable Geographic Information Systems to enable environmental data/information to be represented so that the following types of products could be prepared.

a. maps and figures to provide an overview of where and when data/information are available and spatial presentation of trends;
b. integrated models that can provide general statements about trends between data types (if there are such general trends) and between stations.

9.2 This evaluation work should take into account the needs of the ongoing assessments for completion in 2005 of OSPAR environmental (CEMP), riverine and direct input (RID) and atmospheric deposition (CAMP) data, and the assessment needs for all other types of OSPAR environmental data (e.g. radioactive substances, offshore oil and gas, biological species and community data). It should also take account of the development of the environmental monitoring ans assessment elements of the European Marine Strategy, and as such should also be compatible with the needs of partner marine Conventions.

9.3 The deliverable would be a recommendation on which GIS system to use and how to proceed.

C. OSPAR CEMP specification for 2004-2005

Detailed specifications for the CEMP can be found on the OSPAR web site as Agreement 2004-16. The updated OSPAR List of Chemicals for Priority Action (Update 2004) can also be found as Reference number 2004-12.

Annex 3:

CONTRACT

between

Baltic Marine Environment Protection Commission - Helsinki Commission

(hereinafter referred to as HELCOM)

as contractor

and

International Council for the Exploration of the Sea

(hereinafter referred to as ICES)

as consultant

for the management of Baltic Sea COMBINE Data

1. Commission

1.1 Task

In 1998, the 19th Meeting of HELCOM decided to give the Baltic Monitoring Programme (BMP) data handling consultancy to ICES, which has a long established international marine data centre. Further, HELCOM adopted Recommendation 19/3 on the Manual for the marine monitoring in the COMBINE programme of HELCOM. The primary objective of this task is for ICES to centrally compile the COMBINE data for use in the HELCOM assessment process. Through its fulfilment of HELCOM data tasks since 1998, ICES has been established as HELCOM's Thematic Data Centre (TDC) capable of collecting the COMBINE data over a long period, assuming the continuation of funding for this activity.

Following the conclusion of the original 3-year contract terminating 30 June 2004, this agreement is instituted to continue the operation of the HELCOM COMBINE database by ICES for a further three-year period starting on 1 July 2004.

ICES is contracted to function as the TDC for the HELCOM COMBINE Programme (oceanographic data, biological data and contaminants data); to develop, operate and maintain appropriate databases and data handling systems for the reporting, compilation, archiving, and quality control[1] of relevant data, and for preparing data products; according to the conditions and specifications detailed below.

[1] Data Centre responsibilities relating to "quality control" are described in Section 4.2

1.2 Annexes to the Contract

This Contract includes an Annex which constitutes an inseparable part of the contract. Its content may only be changed as a result of bilateral agreement between HELCOM and ICES as the parties to this Contract. If the content of the Annex differs from that of the Contract, the agreement is determined by the content of the Annex.

1.3 Organization of Work

The Annex describes the organization of the work, including specifications for the work tasks and the products/deliverables requested.

ICES shall not sub-contract any part of the work without prior agreement of HELCOM.

1.4 Contacts and Questions relating to Cooperation

Responsible person at HELCOM is: Executive Secretary

Responsible person at ICES is: General Secretary

Contact shall be maintained between HELCOM and ICES regarding progress of the work under this contract.

1.5 Duration and Time Schedule for the Contract

This contract is effective for the period 1 July 2004 - 30 June 2007.

It may be extended, subject to the agreement of both parties, to include additional work tasks and/or for additional time periods.

The work tasks shall be completed according to the time schedule in the Annex.

1.6 Reporting

ICES will present reports on data submissions and/or progress to HELCOM and/or its subsidiary bodies on request of HELCOM. Whenever possible, ICES will inform HELCOM of delays in data processing which can affect agreed deadlines. ICES and HELCOM will maintain an open communication to ensure identification, investigation and improvement of problems in data submission, quality and/or handling,

2 Fees and Payment

2.1 Economic framework

The contract will be financed from the HELCOM budget.

2.2 Price and Accounting

The contract budget amounts to a total of € 50,000 (excluding VAT) for the budget year 1 July 2004 – 30 June 2005 and € 60,000 per year for the subsequent two budget years pending the approval by the Commission. Budgeted costs include all salaries and overheads required to complete the contract as described. No travel is included in the price of this contract.

If HELCOM requests the participation of ICES in meetings not covered by the Memorandum of Understanding, ICES may charge HELCOM for the associated travel costs and per diem.

2.3 Payment Conditions

Funds will be transferred within 30 days of the receipt of invoice, subject to HELCOM's acceptance of the work (see 3.2). 50% of the annual payment will be invoiced no later than 1 August. The remaining 50% of the annual payment will be invoiced not later than 1 February of the following year.

3 Rights and Acceptance

3.1 Rights

The rights to the data management systems, defined as the software applications which are developed for the management of the data referred to under this contract, remain the property of ICES. Non-exclusive, non-transferable rights to use these applications are transferred to HELCOM under this contract.

Access to data reported to ICES as HELCOM COMBINE data is subject to agreements between HELCOM and the parties providing the data. HELCOM must provide ICES with the details of all such agreements in writing. There are no such limitations on access to HELCOM COMBINE data which are reported to ICES both for HELCOM purposes and as contributions to other national and international programmes for which ICES has data management responsibilities.

HELCOM has ownership and copyright on all products of the contract, including reports (paper and electronic media) and information arising in connection with the project, where these are neither covered by the rights of the data originators nor by the rights of ICES to the data management systems.

ICES retains the right to charge for work and/or data delivered to other parties for purposes outside the scope of the COMBINE.

3.2 Acceptance

Acceptance of the work under this contract is subject to HELCOM approval in accordance with the conditions of this contract and the specifications described in the Annex.

4 Responsibilities

4.1 Responsibilities of HELCOM

HELCOM is responsible for coordinating the submission of data required for the HELCOM assessment and for deciding on priorities with respect to data sets to be submitted.

HELCOM is responsible for providing ICES, either directly or indirectly, with essential information required for handling of the HELCOM COMBINE data, e.g., providing contact names and addresses, etc.

HELCOM is responsible for informing ICES in writing of any restrictions on the access, use or distribution of the data.

Data must be delivered in electronic format.

National HELCOM data coordinators are responsible for the submission of data from their respective countries according to the standards for COMBINE data submission to ICES. Data shall be subject to quality assurance (QA) prior to its submission to ICES.

Data originators are responsible for ensuring good quality (both scientific and technical) of delivered data. If the quality of the data is poor or data submissions do not conform to the agreed specifications, i.e. to the extent that they cannot be correctly interpreted and stored in the ICES Data Centre's data management systems, ICES has the right to return the data to the data originator and request correction/ resubmission.

HELCOM is responsible for any consequences arising from delays in delivery of data or data submissions which do not comply with the agreed formats.

HELCOM shall inform ICES of any changes in HELCOM's data policy.

4.2 Responsibilities of ICES

ICES is responsible for establishing appropriate systems to handle the HELCOM data submissions, archive the data, and prepare appropriate data products.

ICES is responsible for specifying media and reporting formats to be used for data submissions.

The specification of reporting requirements shall include procedures for validation of submitted data and correspondence with data originators on questions relating to submissions. ICES must take into account the preliminary schedule for reporting of data as indicated in the Annex, and will recognize that some institutes reporting data for the first time may experience problems with use of reporting formats.

ICES is responsible for quality control (QC) of data format; correct loading of data into databases; and, if necessary, for returning copies of data to originators for validation and follow-up with data originators on questions relating to data submissions.

ICES is responsible for QC in relation to the scientific validity of the data using systems/expertise available at the data centre, including application of relevant screening programs, and simple statistical and graphical applications to check data for valid ranges of values, inter-component consistency, time series consistency, and consistency with other data sets received.

ICES is responsible for archiving HELCOM/COMBINE data, and for data security.

ICES is responsible for preparing products to be defined in consultation with data assessors (see the Annex). Data products shall be based on a consistent treatment of data submissions. Basic data products will comprise standard tabulations, overviews and summaries of data, univariate statistical products, exportable (ASCII) data files, and simple graphical and mapping presentations.

ICES is responsible for completing the contract as specified within the given time and budget.

4.3 Joint responsibilities

HELCOM and ICES shall agree on further development of the databases and data handling procedures relevant to the COMBINE Programme. When such developments are outside the budgetary and/or scheduling limitations of this contract, they must be prescribed by a separate contract.

5 Delays – Force Majeure

Delays which can be attributed to ICES alone may result in reduction or postponement of the annual contracted payment. In such a case, HELCOM and ICES will negotiate the possible reduction or postponement.

Delays that cannot be attributed to ICES, such as late submission or non-submission of data by the Contracting Parties to HELCOM, cannot be the basis for reduced or postponed payment.

In the event of a situation arising which falls under the definition of Force Majeure, this contract is not considered to apply whilst the Force Majeure situation exists. If this situation results in delays which cannot be accommodated within the existing schedule of activities, a new agreement for the time schedule and continuation of the project shall be agreed in the form of a supplementary annex to this contract.

6 Cancellation

Either HELCOM or ICES may propose a withdrawal from this contract. Any such proposal shall be made in written form at least 12 months before the withdrawal shall come into effect.

In the case that HELCOM significantly alters the content or scope of the project, ICES can resign the contract on a 2-month notice, subject to completion of work already begun.

7 Disputes and Governing Law

Should disputes arise out of this contract, which cannot be solved amicably, they shall be submitted to the City Court of Helsinki subject to Finnish law.

This contract is completed in two originals.

Helsinki Commission	International Council for Exploration of the Sea
Date: 2004	Date: 2004
Anne Christine Brusendorff	David Griffith
Executive Secretary	General Secretary

Annex: Organization, Tasks and Schedule for HELCOM/COMBINE Data Work

HELCOM/COMBINE national data coordinators will act as the primary contacts at the national level, in relation to national submissions of data, and coordinate work at the national level to compile quality assured data from relevant projects and transmit them to ICES in accordance with agreed protocols and time schedules.

Data originators are defined as the individual/group responsible for submission of data to the data centre: depending upon data reporting procedures adopted within different countries, data originators may comprise NDMs, data quality assurance experts associated with a given assessment activity, national/institute data centres, project leaders, etc. ICES shall seek to establish and maintain contact with data originators.

Data originators will be responsible for notifying ICES of any restrictions concerning access to their data which are additional to the restrictions of the HELCOM data policy.

Upon receipt, data will be processed by ICES, incorporated in ICES databases, and flagged as HELCOM/COMBINE data. Data validation including any necessary correspondence with data originators will be an integral part of this data handling process.

HELCOM/COMBINE national data coordinators will coordinate work at the national level to ensure timely responses on data validation questions, etc.

On request from HELCOM, ICES will make available the contents of the HELCOM/COMBINE data holdings.

Type of data	**Data Originator and Data coordinator tasks**	**Scheduled for**	**Data centre tasks**	**Scheduled for ***
COMBINE hydrographic and hydrochemical data	Data submission (for year **x**) according to agreed formats. Data subject to QA prior to submission. National data coordinator's responsibilities: • collect data sets from each national laboratory • check that all required information has been submitted • perform basic quality control checks on data • submit data to the data centre.	as soon as possible but no later than 1 May of the following year (**x**+1)	Data receipt and Inventory of submitted data	Continuous
			• quality control submitted data • return copies of data to originators for validation (if necessary) • follow-up (if necessary) • load data into database	31 December of year **x**+1
			Prepare standard data products and summaries annually according to specifications developed during the MONAS autumn meetings.	28 February of year **x**+2
			Develop and prepare additional data products such as indicator-based assessment products. If such products are not within the schedule or budget of this contract, requirements for them shall be developed on a project basis, as a HELCOM project with proposed funding	As agreed between HELCOM and ICES.
COMBINE biological data and data on harmful substances		as soon as possible but not later than 1 September of the following year (**x**+1)	Other work tasks as above	Same dates as above
*) This schedule applies only for data submitted on time. Late submissions will be handled, but only after on-time submissions and as ICES resources allow.				

Annex 4:

Databases in ICES

ICES secretariat

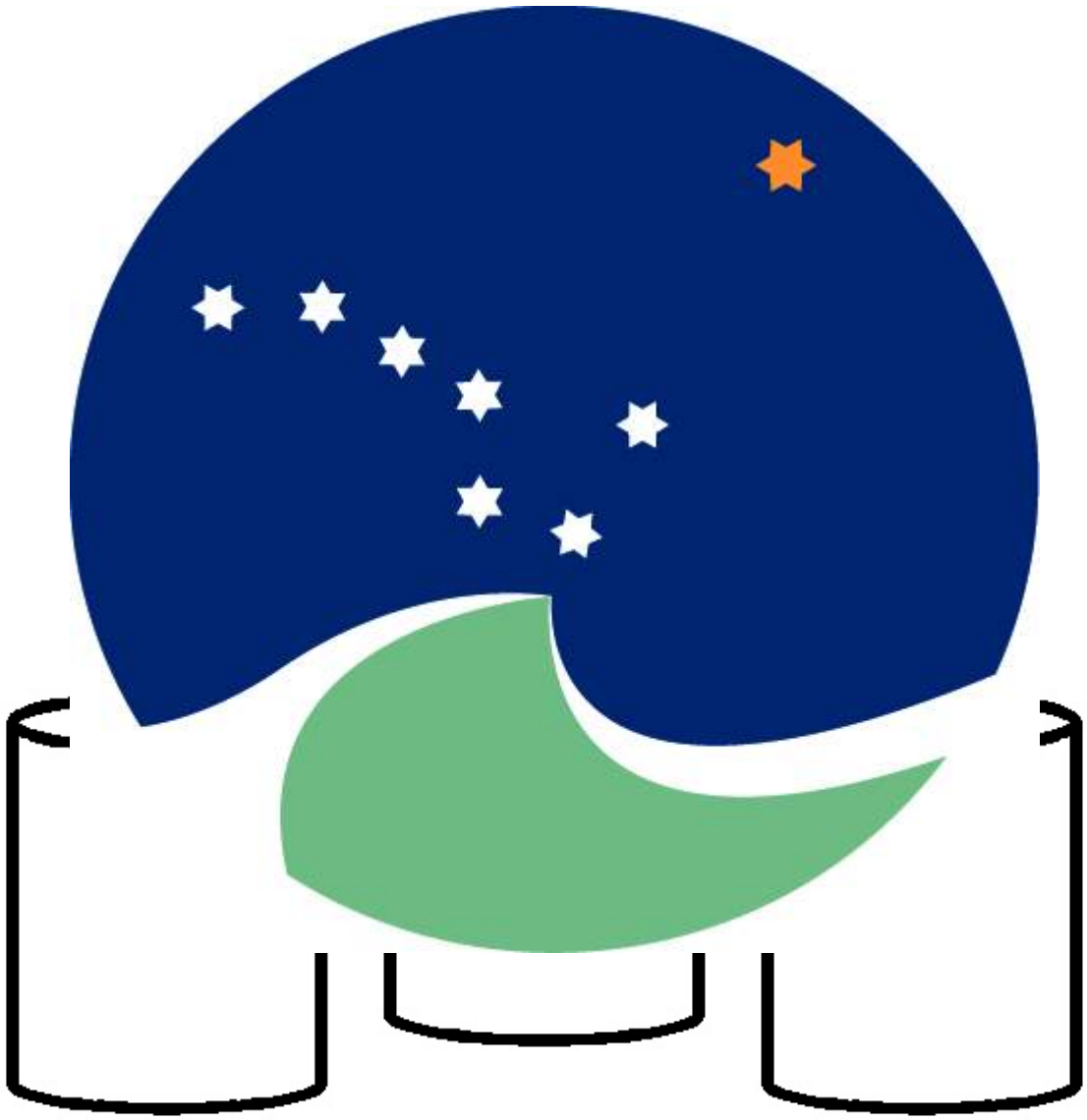


TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ICES Databases

The ICES Secretariat stores and processes large amounts of data. Traditionally, ICES has worked within 3 disciplines: Environment, Oceanography and Fisheries. These 3 disciplines have co-existed with little interchange. This is reflected in dispersed, widely differing data formats, systems and processes.

Today, marine science disciplines should be seen as interrelated, interacting elements of the ecosystem approach. The information of the 3 disciplines synergizes when considered collectively. Thus, there is incitement to link and integrate the data via common programs and processes.

An integrated system is also more efficient to operate, maintain and support than several independent, specialized systems serving the same purposes. However, development of integrated systems usually proves to be more challenging.

The ICES Secretariat is moving toward integrated systems. New multi-disciplinary databases are being developed to supplement and/or replace existing specialized databases. These are described herein. Related systems and processes also being developed, such as data verification and calculation, are outside the scope of this document.

Environmental Data

Data Types

Environmental data comprise three data groups:

- **Contaminants and biological effects** in Biota (CF), Seawater (CW) and Sediments (CS)
- **Fish Disease** (DF): prevalence of fish diseases
- **Biological Community**: abundance/biomass of phytoplankton (PP), zooplankton (ZP), phytobenthos (PB) and zoobenthos (ZB).

Compared to the other disciplines, there is relatively little data. Environmental data is, however, much more structurally complex.

Environmental data are collected by member countries and sent to ICES every year in separate files according to data type. Upon receipt, the data are checked and, if acceptable, stored in their original format. Environmental data have previously been supplied in pre-defined data formats in ASCII (text) files (format version 2.2), or comma-separated (CSV) files (format version 3.1). An integrated format (version 3.2) for all data types is now being used (see “Current Format”).

Table 1 summarizes the environmental data stores.

Environmental data types		
Name	Data format	Size (MB)
Contaminants/effects in Biota	ASCII files (2.2)	72
Contaminants/effects in Sediments	ASCII files (2.2)	26
Contaminants/effects in Seawater	ASCII files (2.2)	54
Fish Disease Database	ASCII /binary files (2.2)	60
Biological Community Data	ASCII files (3.1)/ Helcom format	5/20
"Integrated" submissions	CSV files (3.2)	13
Intercalibration result database	ASCII files	234

Table 1. Environmental Data Stores

All environmental data except the version 2.2 biological effects data will eventually be migrated to DOME (see "Integrated Databases"). Intercalibration results will be linked to DOME.

Current formats

An integrated data reporting format, version 3.2, was recently implemented for all environment data types. This version offers many advantages in relation to versions 2.2 and 3.1, including support of the latest biological effects techniques which replace and expand on the "old" biological effects reporting. Its records are listed in Table 2, and data hierarchies are shown in Figures 1 & 2.

- Figure 1 shows the hierarchy for contaminants and biological effects in biota (CF), sediment (CS) and seawater (CW)
- Figure 2 shows the Biological Community data hierarchy which contains measurements of the abundance/biomass of phytoplankton (PP), zooplankton (ZP), phytobenthos (PB) and zoobenthos (ZB).

Still, any format can be burdensome for data suppliers, therefore we are also conducting a free format trial to determine if allowing data submissions in any well-defined format could benefit data submitters and still be feasible for us.

Environmental Data Format 3.2			
Record identifier. Desc. (Group)	Biota incl. Fish Diseases	Seawater Sediment	Biological Community data
00. Information (Country, reporting institute)	X	X	X
90. Sampling Platform	X	X	X
91. Station / Sampling event	X	X	X
92. Station / Site description	X	X	X
03. Contaminant and biol. effects sample	X	X	
20. Sampling method	X	X	X
04. Biota Specimen	X		
10. Measured parameter data	X	X	
21. Analytical method	X	X	X
93. Reference material (QA)	X	X	
94. Intercomparison info. (QA)	X	X	X
34. Biol. Community sample			X
38. Biol. Community Abundance/Biomass			X

Table 2. Environmental Format 3.2 Data Records

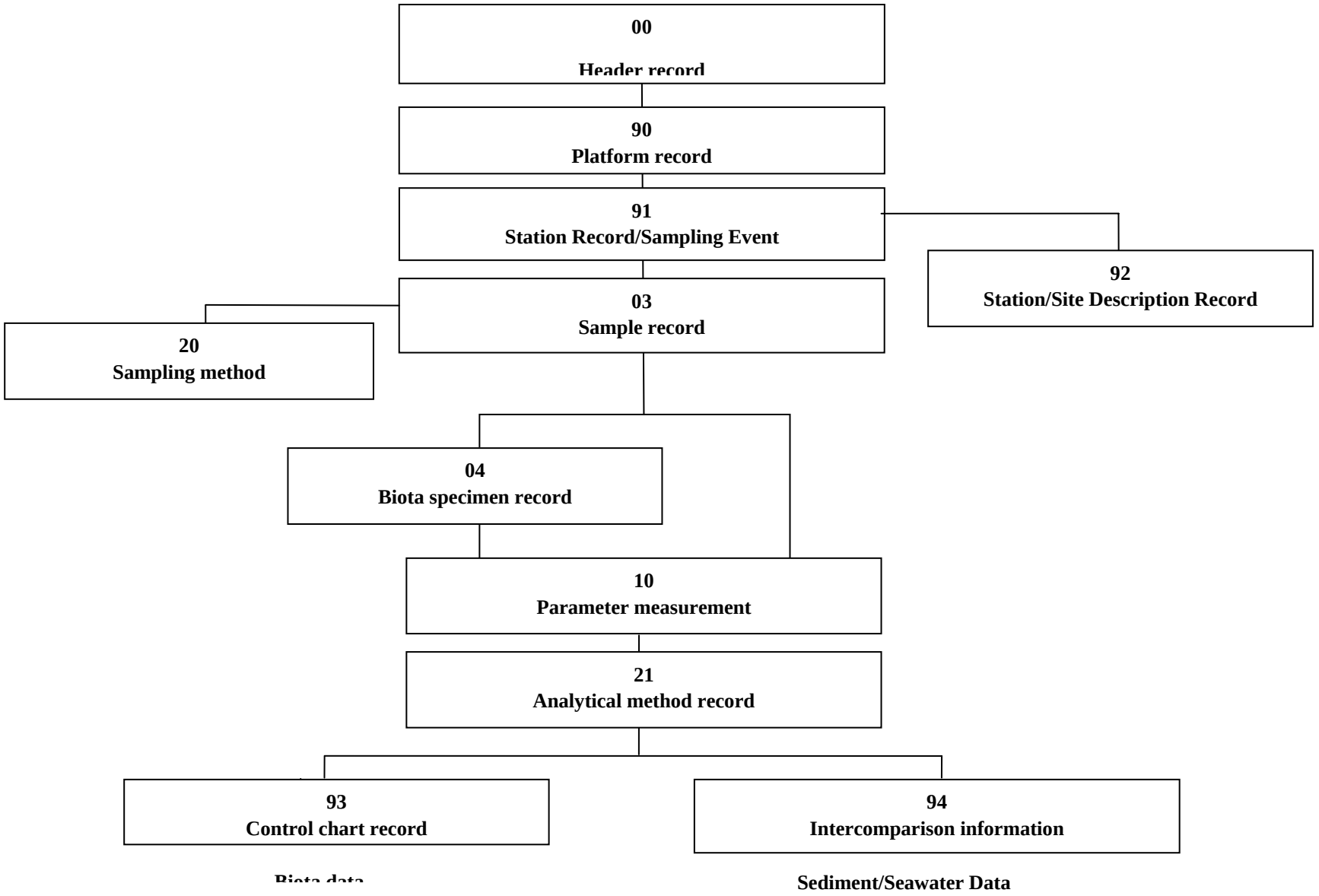


Figure 1. Contaminants in Biota, Sediment/Seawater: Format 3.2 Record Hierarchy

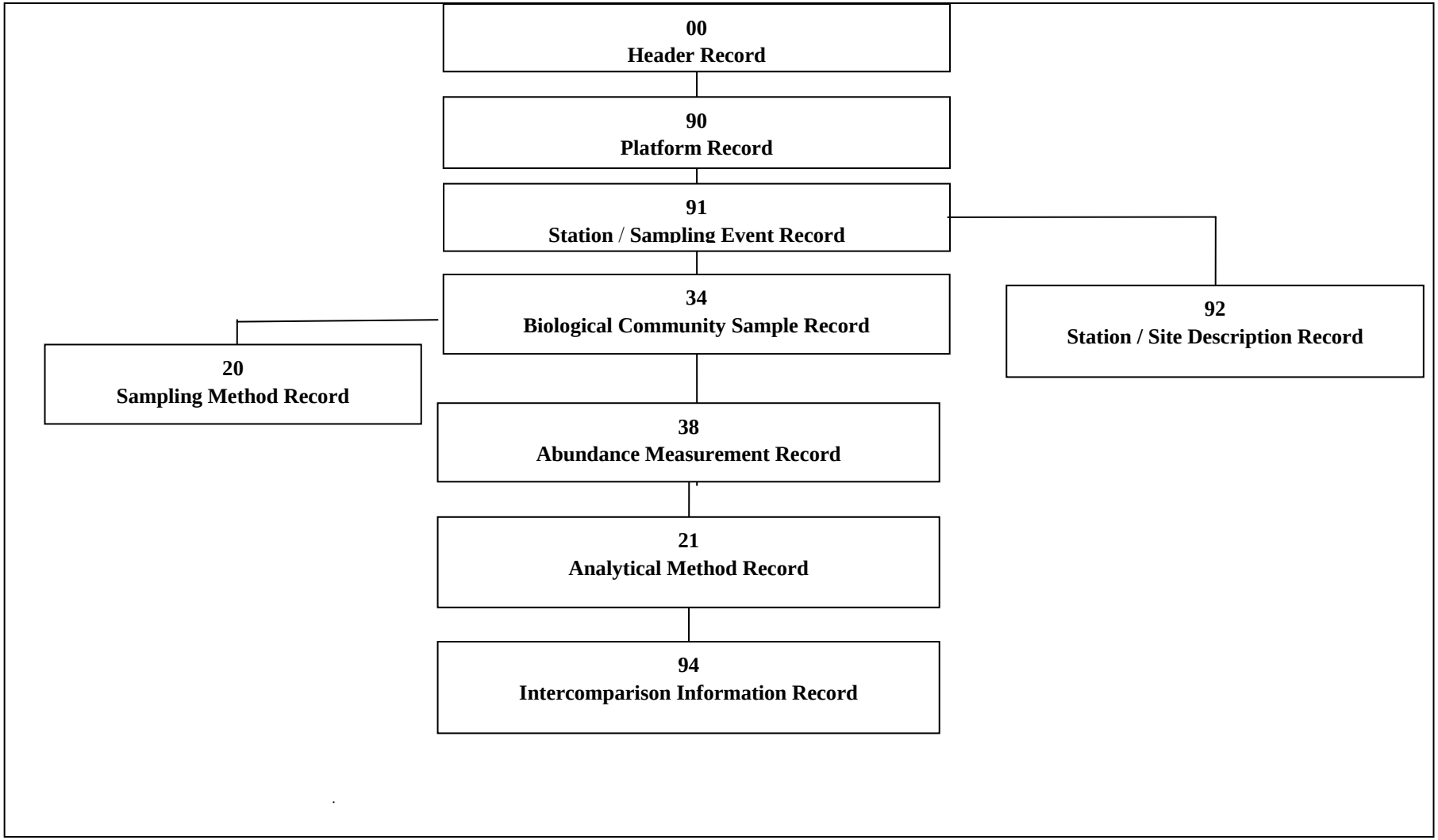


Figure 2. Biological Community: Format 3.2 Record Hierarchy

Oceanographic Data

Oceanography data are voluminous, but relatively simple in structure.

Data are submitted on regular basis from all national data centres within ICES member countries, and occasionally from individual scientists.

All oceanographic files except ROSCOP contain measurements of physical and chemical parameters (nutrients ect). The data are stored as indexed ASCII files which make them reasonably fast to search and extract.

Table 3 summarizes the oceanographic data.

Oceanographic data			
File type	**Description**	**Data format/DB**	**Size (MB)**
High resolution CTD	conductivity, temperature, depth	Indexed ASCII files	4123
Hydro-Chemistry data	incl. low resolution CTD	Indexed ASCII files	1097
Surface data	temperature & salinity during steaming	Indexed ASCII files	76
Pump data	stern-inflow samples	Indexed ASCII files	3
ROSCOP	Cruise data	ASCII files	14

Table 3. Oceanographic Data Stores

CTD data are, by far, the most voluminous. High resolution CTD and hydro-chemistry data are often measured at the same cast. During the down-cast, the CTD instrument records **C**onductivity, **T**emperature and **D**epth data frequently – thus its high resolution. On the up-cast, a number of water bottles are shot, taking water samples for laboratory analysis of nutrients, tracers, contaminants etc.

Data are submitted in a variety of formats. Upon receipt, they are assigned an enquiry number which links them to cruise information in the ROSCOP file. Data are converted to the comprehensive “ICES Oceanographic Format” (IOF), quality controlled and merged into the data base.

The IOF is comprised of 80-character long records in which positions 79-80 identify the record type. Table 4 lists oceanographic data record types. Figure 3 shows the hierarchy of low resolution CTD data.

Record ID	Information	CTD		Pump	Surface	Occurrence
		High Res.	Low Res.			
0J	Date, time, coordinates, ship, country, code to cruise	X	X	X	X	Mandatory
03	Depth/Pressure, temp., salinity, method of salinity and oxygen.	X	X	X	X	Optional
76	Fixed number of measured hydro parameters	X	X	X	X	Optional
0Z	Any parameter&analytical method defined in BODC Parameter Dictionary.	X	X	X	X	Optional

Table 4. Oceanographic Data Records

"0J" contains time and coordinates, and thus serves as the master record for all oceanographic data.

In some cases, only one data type occurs. However, if more than one data type occurs, the hierarchy in Figure 3 must be maintained in the file. A typical case is an "0J" record followed by a series of "03" records, followed by one "76" record, followed by a series of "0Z" records. *Note: leading digits represent the record identifier; a dashed frame indicates that records of the same type can occur in the series.*

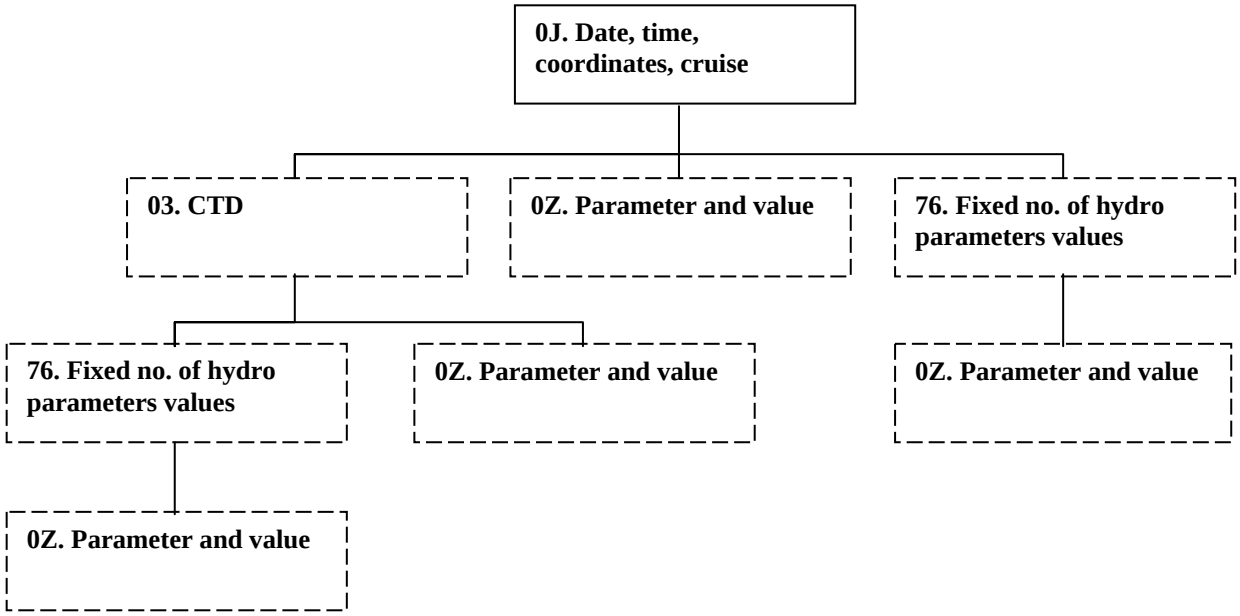


Figure 3. Low Resolution CTD data record Hierarchy

Fishery Data

Fishery data are very different from environmental and oceanographic data. Both environmental and oceanographic data are stored in "flat" ASCII text files, whereas fisheries data are stored in relational databases. Table 5 summarizes the fishery data.

Fishery Data			
Name	**Description**	**Data format/DB**	**Size (MB)**
DATRAS	IBTS, BITS and BTS trawl surveys	MS SQL Server	530
Rec-12	preliminary Catch statistics	MS SQL Server	2
Assessment Summary	fish stock assessment summary tables	Access	2
'NewIFAP'	Catch aggregated to international level	MS SQL Server	under development

Table 5. Fishery Data Stores

DATRAS

DATRAS is an SQL Server database consisting of data from 3 trawl surveys, namely, IBTS, BITS and BTS. Survey type and area are listed in Table 6.

DATRAS		
Survey	Survey type	Carried out in
IBTS	Bottom trawl	North Sea, Skagerrak, Kattegat
BITS	Bottom trawl	Baltic Sea and Belts
BTS	Beam Trawl	North Sea

Table 6. DATRAS Surveys

Catch in number/hour by species, length, and individual trawl haul are submitted in ASCII files. Each record/line starts with a 2 character record ID which identifies its type of information. Record types are listed in Table 7. The data hierarchy of IBTS and BITS data is shown in Figure 4.

Record ID	**Information**	**Occurrence**

HH	Date, time, coordinates, ship, country, haul data	Mandatory
HL	Species, length class, number caught for the spec. length class	Mandatory
CA	Species, length class, age for spec. length class	Optional

Table 7. Fishery Data Records

An "HH" record is followed by a series of "HL" records which may be followed by a series of "CA" records at the end of the file.

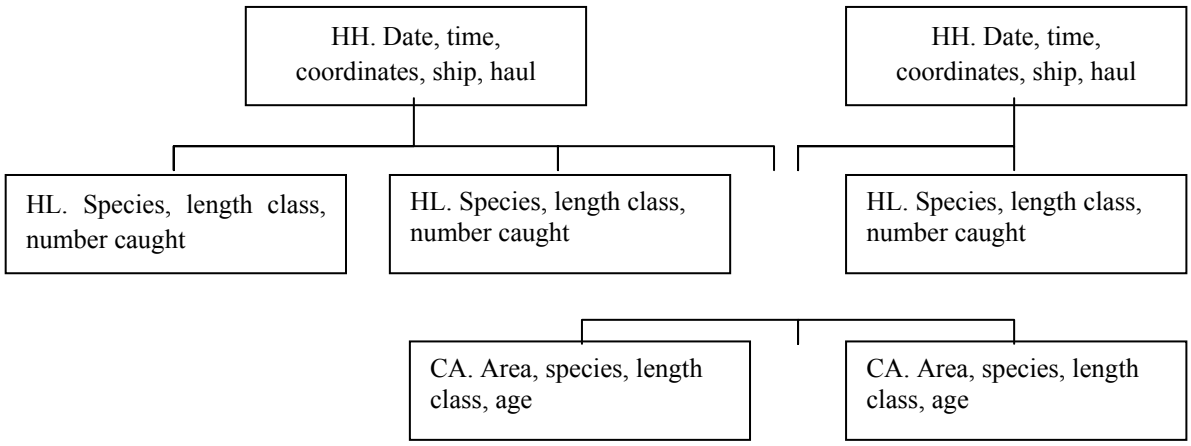


Figure 4. Fishery Data Record Hierarchy

Rec-12

Rec-12 is also an SQL Server database. It contains preliminary catch data which are reported once a year by each country.

Data are collected only for ICES Divisions, and are thus are limited to the Northeast Atlantic. Data are imported from ASCII files or entered manually.

Assessment Summary database

The Assessment Summary database is built in Access, and contains the summary table data from the fish stock assessment. For each year and stock, the Assessment Summary database contains landing, recruitment, spawning stock biomass, total biomass and fishing mortality. This data forms the basis for the standard graphs in the 'Reports of the ICES Advisory Committee on Fishery Management'.

New IFAP

The NewIFAP (IFish/ICatch) system will be developed in collaboration with DIFRES in 2005. It will contain national commercial catch and aggregated biological sampling information, all inputs for stock assessments (including effort and fishery independent information), and will facilitate the manipulation of data (e.g., for raising and aggregation to international catch) and the documentation thereof.

Integrated Databases

The term “integrated” is used to describe databases containing data from more than one discipline. Integration supports a multi-disciplinary, ecosystem approach. As can be surmised from the various data types and hierarchies within each discipline, integrating data within and across the disciplines presents many challenges.

Name	**Data format/DB**	**Size (MB)**
Accessions	Jet engine/Access	2
ICES Integrated Inventory	MS SQL Server	440
DOME	MS SQL Server	Under development

Table 8. Integrated Databases

Accessions

All data submissions of all types are logged in Accessions, an Access database.

ICES Integrated Inventory (III)

The inventory summarizes ICES data holdings by listing amounts of data for almost all data types and parameters (see Table 9). It is primarily used by researchers to determine how much data ICES has for a given area, period and parameter(s). Free and open access to the inventory is provided via the internet.

Discipline	Data
Environment	Biota
	Seawater
	Sediments
	Fish Disease
Oceanographic	High res. CTD
	Hydro- chem. Data
	Surface Data
	Pumpdata
Fisheries	IBTS
	BITS
	BTS

Table 9. ICES Integrated Inventory Coverage

DOME

Ideally, all data would be completely integrated into a single database. The ICES secretariat has embarked on a 3-phase project to do so. It is called the Database on Oceanography and

Marine Ecosystems (DOME) and aims to integrate as much measurement data as relevant and possible. Table 10 summarizes the development plan.

ɔntal		phic	
Contaminants and Biological Effects in Biota (including Fish Disease), Seawater and Sediments		mistry	
Community		CTD, surface and pump data	
			under consideration

Table 10. DOME Development Phases

Annex 5: Overview of Oceanographic projects which the ICES Service Hydrographique has participated in since 1996.

For all three projects listed below, the ICES Secretariat has acted as data centre, and been a member of the project's steering committee. Specifically, ICES quality controlled the data and made it available in common formats to the partners. ICES was also responsible for providing the partners with an up-to-date overview of their outstanding obligations to collect data. Finally, at the end of the projects, ICES published the project's datasets on a CDROM.

ESOP-II (a continuation of ESOP-I and the Greenland Sea Project)

Title	Thermohaline Circulation in the Greenland Sea
ICES web page	http://octopus.ices.dk/ocean/project/esop
Project Office	University of Bergen
Duration	3 years (1/1-1996 to 31/12-1998)
Funded by	EU MAST-III (MAS2-CT95-0015)
Approximate ICES budget	1.4 Million DKK
Project funding percentage	50%

Partners:

UOB-Univ. of Bergen, Norway (co-ordinator)
PML-Plymouth Marine Laboratories, UK
UCAM-SPRI-Scott Polar Research Institute, Cambridge University, UK
LODYC-Université Pierre et Marie Curie, France
GU-Gothenburg University, Sweden
AWI-Alfred Wegener Institut, Germany
IFM-Inst. für Meereskunde, Hamburg Univ., Germany
DHI-Danish Hydraulic Institute, Denmark
CSNSM-Centre de Spectrom. Nucleaire et de Spectrom. de Masse, ORSAY, France
UCAMB-DAMTP-Dep. of Applied Mathematics and Theoretical Physics, Cambridge University, UK
DMI-Danish Meteorological Institute, Denmark
TUD-Technical University of Denmark, Denmark
IMGA-CNR, Modena, Italy
IMR-Institute for Marine Research, Norway
NTH-Univ. of Trondheim, Norway
NERSC-Nansen Environmental and Remote Sensing Center, Norway
MRI-Marine Research Institute, Iceland
NPRI- Norwegian Polar Research Institute, Norway
IMO-Iceland Meteorological Office, Iceland
UKMETO-UK Meteorological Office, UK
ICES, Denmark

VEINS

Title	Variability of Exchanges in the Northern Seas
ICES web page	http://octopus.ices.dk/ocean/project/veins
Project Office	University of Hamburg (Jens Meincke)
Duration	3 years (1997 to 1999)
Funded by	EU MAST-III (MAS3-CT96-0070).
Approximate ICES budget	1.75 Million DKK
Project funding percentage	50%

Partners:

Alfred-Wegener-Institut für Polar- und Meeresforschung (AWI)
Centre for Environment, Fisheries & Aquaculture Science (CEFAS)
Det Norske Meteorologiske Institutt (DNMI)
University of Dundee Dept. of Civil Engineering (DUCE)
Finnish Institute of Marine Research (FIMR)
Geophysical Institute, University of Bergen (GI)
Göteborg University Marine Research Centre (GUMC)
Göteborg University Department of Oceanography (GUDO)
Institut für Meereskunde Universität Hamburg (IfMH)
International Council for the Exploration of the Sea (ICES)
Institute of Marine Research Bergen (IMR)
Institute of Oceanology Polish Academy of Sciences (IOPAS)
Instituto Sperimentale Talassographico die Trieste (IST)
Marine Research Institute Reykjavik (MRI)
Norsk Polar Institute Tromsoe (NPI)
Plymouth Marine Laboratory (PML)
Department of Meteorology Stockholm University (SUMO)
University of East Anglia (UEA)
University Courses of Svalbard (UNIS)

TRACTOR (ESOP-II follow up)

Title	TRAcer and Circulation in The NORdic Seas Region
ICES web page	http://octopus.ices.dk/ocean/project/tractor
Project Office	University of Bergen (Bjerkness Centre)
Duration	February 1, 2001 to January 31, 2004
Funded by	EU Fifth Framework Programme Contract Nr. EVK2-2000-00080
Approximate ICES budget	1.5 Million DKK (~1 MDKK spend)
Project funding percentage	50%

Partners:

University of Bergen (Co-ordinator) UoB	Norway
University of East Anglia UEA	United Kingdom
University of Göteborg UGOT	Sweden
Nansen Environmental and Remote Sensing Center NERSC	Norway
Norwegian Polar Institute NPI	Norway
Institute of Marine Research IMR	Iceland
International Council for Exploration of the Seas ICES	Denmark

Coming projects

New oceanographic projects are continuously being funded by EC. The 6th framework program emphasizes very large projects and such a project (named for Europe, abbreviated 4EU) has been proposed by many of the same partners who participated in VEINS. The object of the projects is to increase the ability to predict changes in the Thermohaline circulation. This will be done by improving models and data assimilation projects. A major part of the project deals with observations much in the same way as VEINS. ICES-DC has the experience necessary to coordinate the collection, quality control, exchange and safeguarding of data in such a project.

Other projects in smaller scale are also coming up. E.g. the Galathea 3 expedition (http://www.galathea3.dk/)

Annex 6: Web Portals

IOC/IODE OceanPortal http://www.oceanportal.org/

Ocean Portal is a high-level directory of Ocean Data and Information related web sites. Its objective is to help scientists and other ocean experts in locating such data & information. Interested users can submit a new URL, modify an existing link or report a broken link. When submitting a new link the user has to select the Category that best describes the new site. New category names can also be suggested.

SeaSearch Portal http://www.sea-search.net/

The SeaSearch project is a EC funded thematic network project, running until end of September 2005. It has been active first as the "Mast Data Committee and later as the EC funded concerted action project named EURNODIM. The continuation of the activity is proposed in a new project named SeaDataNet, (http://www.seadatanet.org/), which is not funded yet. SeaSearch describes itself to be "Your gateway to Oceanography and Marine Data & Information in Europe". The website has been set up and is being operated and further developed to provide an effective navigation tool to data and information sources in Europe, to oceanographic data and information, managed by European centres, and to centres in Europe with expertise and skills in oceanographic and marine data & information management.

U.S. GLOBEC Data Hub http://globec.whoi.edu/jg/dir/

The U.S. GLOBEC (GLOBal ocean Ecosystems dynamics) Data Hub makes use of the JGOFS (The Joint Global Ocean Flux Study) Software System. Large oceanographic programs such as JGOFS require data management systems which enable the exchange and synthesis of extremely diverse and widely spread data sets. A distributed, object-based data management system for multidisciplinary, multi-institutional programs have been developed. It provides the capability for all JGOFS scientists to work with the data without regard for the storage format or for the actual location where the data resides. The approach used yields a powerful and extensible system (in the sense that data manipulation operations are not predefined) for managing and working with data from large scale, on-going field experiments.

Search engine Google http://www.google.com/

Web Portals created on a specific topic like data management can probably never be as efficient as the huge search engines available on the Internet. Search engines like Google are specially designed to be efficient and fast and therefore run "robots" in background to index web pages. ICES WGMDM did a survey on the Internet to look for the Guidelines published by the group. It turns out that one has to put effort in how web pages are created. Three criteria are important, key words, links and updates. If the web pages fulfil these criteria they will be given high relevance and priority by the search engines and thereby be on the top of the search result list. To be found by search engines information has to be available somewhere and web linked several places. The use of Web portals helps to make information available also through search engines like Google.

Annex 7: Metadata Catalogues

Metadata

EDMED http://www.bodc.ac.uk/services/edmed/edmeds.html

The European Directory of Marine Environmental Data (EDMED) was initiated in 1991 by the British Oceanographic Data Centre within the EC-MAST framework and has established itself as a de-facto European standard for indexing and searching datasets relating to the marine environment. It covers a wide range of disciplines and is a high level inventory, describing both Datasets and Data Holding Centres. At present, EDMED already describes more than 2300 Datasets from over 500 Data Holding Centres across Europe.

Click above to search the EDMED

All national directories are assembled into a single centralised system managed by the BODC and are made available via the Sea-Search website. Activities are underway by the European Sea-Search partners to update their national EDMED entries and to develop and install an innovative infrastructure for updating the EDMED database by means of the Internet.

EDIOS http://www.edios.org/

EDIOS is the European Directory of the Ocean-observing System, a unique searchable metadatabase. The EDIOS directory provides a new internet-based tool for searching information on observing systems operating repeatedly, regularly and routinely in European waters. The EDIOS directory contains metadata on European observing systems such as platforms, repeated ship-borne measurements, buoys, remote imagery, etc. EDIOS is an initiative of the European Global Ocean Observing System (EuroGOOS). The directory was developed during the EDIOS project, co-funded by the European Commission Research Directorate General. The EDIOS directory currently holds well over 8,500 data entries, which are regularly updated.

To make a full search in the EDIOS directory access the URL: http://www.edios.org/full-search.htm

ISO 19115

The International Organisation for Standardization (ISO) Technical Committee (TC) 211 has developed an international metadata structure called ISO 19115. Approved in July 2003, the standard provides detailed descriptions of the entities and attributes (which compromises over 300 elements) covering the following topics: data set access constraints, data set maintenance frequency, raster – vector spatial representations, spatil-temporal reference system, distribution details (fees, availability, media, ..), spatial extent of the data set and citation, contact and responsible party information.

The ISO 19115 standard defines core metadata components, recommended components and allows community based profiles to be described as extensions to the standard. It is widely seen as the international standard for metadata descriptions. ISO 19115 is a geo-reference metadata standard. As such, ISO 19115 does contain all the necessary fields to adequately describe ocean data sets. These fields will need to be constructed by the ocean data

community and made compliant with the ISO 19115 via the user extension capability of the standard.

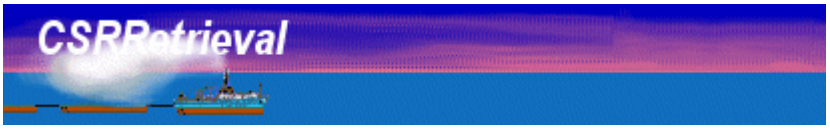
CSR Cruise Summary Report http://www.ices.dk/ocean/roscop/

ROSCOP (Report of Observations/Samples collected by Oceanographic Programmes) was conceived by IOC in the late 1960s in order to provide a low level inventory for tracking oceanographic data collected on Research Vessels etc. The ROSCOP form was extensively revised in 1990, and was re-named the CSR (Cruise Summary Report), but the name ROSCOP still persists because of its love-hate relationship with many marine scientists. Most marine disciplines are represented in ROSCOP, including physical, chemical, and biological oceanography, fisheries, marine contamination/pollution, and marine meteorology.

Traditionally, it is the Chief Scientist's obligation to submit a CSR to his/her National Oceanographic Data Centre (NODC) not later than two weeks after the cruise. With this, a first level inventory of measurements and samples collected at sea are provided. NODCs send CSRs to ICES, where this information is compiled. Up to now the information had to be entered via a paper sheet, and provision of CSRs decreased. Within the initiative of the "European Network for Oceanographic Data & Information Management" (EURONODIM)-project, the Deutsches Ozeanographisches Datenzentrum (DOD) developed and installed an innovative infrastructure for entering, searching and presenting the CSR/ROSCOP database by means of the internet. CSRONLINE was presented amongst other groups to the "ICES Working Group on Marine Data Management" (MDM) and to the IOC Committee on International Oceanographic Data and Information Exchange (IODE), comments for improvements were taken on board. After further testing, CSRONLINE is operational and in use by several scientists. It is password protected and the current password is Username: csronline Password: jellyfish

A password is needed for access by members.
Non members supply their data in a different way.

The entered cruise information is put onto the DOD web pages. This board is also valid as a data tracking system, and is used as such intensively. The online retrieval system for CSR data is linked to the existing database. For the sake of uniform feel and go within the Sea-Search system surfaces, the retrieval system was adopted from the British Oceanograpic data centre. The **CSRRETRIEVAL** has been finalized and can be viewed by the Sea-Search partners only by clicking the banner.



The Marine Metadata Interoperability website http://www.marinemetadata.org/

The Marine Metadata Interoperability project wants to make it easier to work with marine science data. The project's work -- including this site -- is for scientists and their data management staff who want the best advice and materials for doing data management right.

The MMI site identifies best practices to make your science data easy to distribute, advertise, reuse, and combine with other data sets. And, it's making those other data sets easier for you to find, access, and use. By simplifying the incredibly complex world of metadata into specific, simple guidance, MMI allows scientists and data managers at all levels to apply good metadata practices from the start of a project.

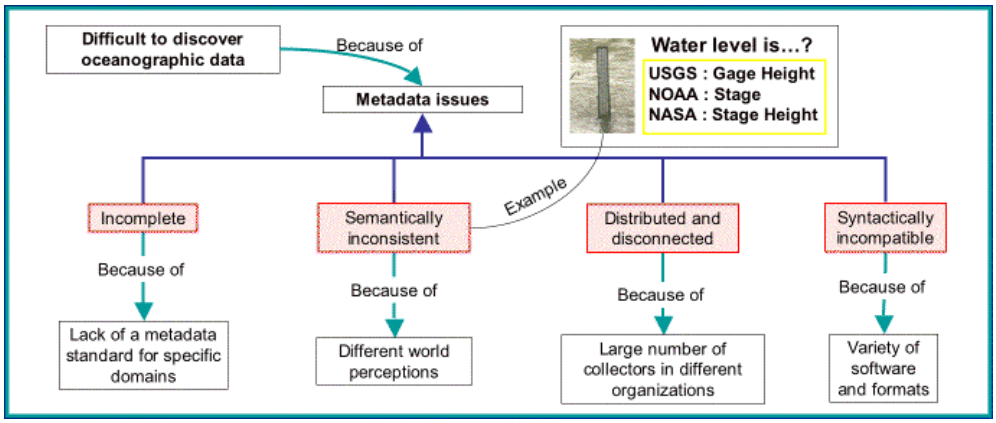


Image adapted from "HOW: Hydrologic Ontology for the Web". Luis Bermudez, Michael Piasecki, Dec, 2003. (AGU Poster.)

For marine data management professionals, the MMI will provide useful and detailed resources -- information, tools, standards, cookbooks, and working

Others

A large amount of other metadata registries, some examples are:

MEDI – Marine Environmental Data Information Referral Catalogue system (IOC)

http://ioc.unesco.org/medi/index.html

The International Inventory of Moored Current Meter Data

The inventory comprises about 22500 entries relating to current meter data collected in 15 countries.

http://www.bodc.ac.uk/ Data Services.

GBIF – Global Biodiversity Information Facility http://www.gbif.org/

The mission of the Global Biodiversity Information Facility (GBIF) is to make the world's **primary data on biodiversity** freely and universally available via the Internet.

Parameter dictionaries

A normalised database is based on parameter codes. Parameter codes need a parameter dictionary. A set of codes may be managed in many different ways, a text file, a spreadsheet or a database. The set of codes represent a description of the parameters being considered. The managed set of codes is termed a parameter dictionary. BODC, British Oceanographic Data Centre, came to develop a fully normalised database structure based on parameter codes for the atmospheric, water column and biogeochemical data, because no existing parameter coding system could fully satisfy the specification required. BODC has therefore developed a parameter dictionary to enable a multi-disciplinary normalised database to be used for project data management. The dictionary describes the nature, sphere and methodology of measurands using text terms built by automated concatenation of a semantic model built from controlled vocabularies.

The dictionary is made publicly available and is currently used by NIOZ - The Royal Netherlands Institute for Sea Research, RIKZ - National Institute for Coastal and Marine Management in the Netherlands, Fisheries and Oceans Canada and the ICES database in addition to its use in BODC projects. It is far the most extensive parameter directory available to the oceanographic community, containing over 16.500 codes (December 2004). The ICES SGXML, Study Group on the Development of Marine Data Exchange Systems using XML early recognized the importance of parameter codes when exchanging oceanographic data using XML. The SGXML recommends that the BODC parameter dictionary be adopted as the marine ocean community standard, including the use of the BODC dictionary in any developed marine XML.

Annex 8: Distributed vs centralised databases (including a review of existing tools and formats/protocols useful for setting up integrated databases)

The working group discussed the benefits of "centralized" vs. "distributed" database models. The ICES Study Group on Management of Integrated Data, SGMID, made an overview of the subject on its meeting in 2004. The information is reused from their report.

A *centralized* database is where all data are physically located at and served from a single location, as is the current ICES database. A *distributed* database is where the data can be located at various geographically distributed nodes (*e.g.*, multiple institutions and/or local data centres) and may be accessible from a single interface. The management issues and attributes of each model are summarized in Table X.X

Each model has its own advantages and disadvantages. The working group did not conclude if either of the choices are ideal for ICES, neither does SGMID. Some applications favours the centralized model (*e.g.*, large, data-heavy hydrographic databases) whereas others favours the distributed model (*e.g.*, complex biological data types that are better kept closer to the expertise that originally collected it). The two models do not have to be exclusive. For example, ICES could ideally have a single centralized database for certain data types and distributed sub-databases which serve the special needs or data of local institutions. The SGMID thought that the distributed systems might be developed faster and for less expense than centralised systems.

Management Issues / Attributes	**Centralized Model**	**Distributed Model**
Computer & Technical Issues	Centralized hardware and computer staff *(better support and hardware for less or same money, +)* Easier to keep entire (single) data system online and to provide fast internet access (+)	Each node requires capable hardware and staff (which might be present already) (quality of both may differ between nodes, -) Rural nodes may suffer poor internet access and/or power losses (-) *Note: A centralized data archive is still required for long term data archiving.*
Data Access & Re-Formating[1]	Reformatting and indexing effort done at initial load *(quicker access and response, +)*	Data are reformatted and/or indexed on the fly for each request, plus travel time from the nodes *(slower access and response, -)*

Management Issues / Attributes	Centralized Model	Distributed Model
Quality Control	Data managers not as familiar with data and separated from its original collector (-) Corrections to original data require reloading into central database, slowing update (-) Easier to perform multi-data cross-variable statistical and quality control checks (+)	Data closest to the original investigator expertise (+) Corrections to local data are immediately available (+) Data may only be checked against themselves (-) Less risk of data duplication (+)
Apparent Data Ownership and/or Credit	Focus on centralized center may over-shadow (or omit) collecting institutes and parties (-) Investigator participation and submission interest may be lower (-)	Data stays local, offering better possibility of credit and ownership recognition (+) Possible better investigator participation & submission (+)

1 Assuming that effort of reformatting to centralized database is approximately equivalent to programming an OPeNDAP protocol and/or minor reformatting required for the distributed database.

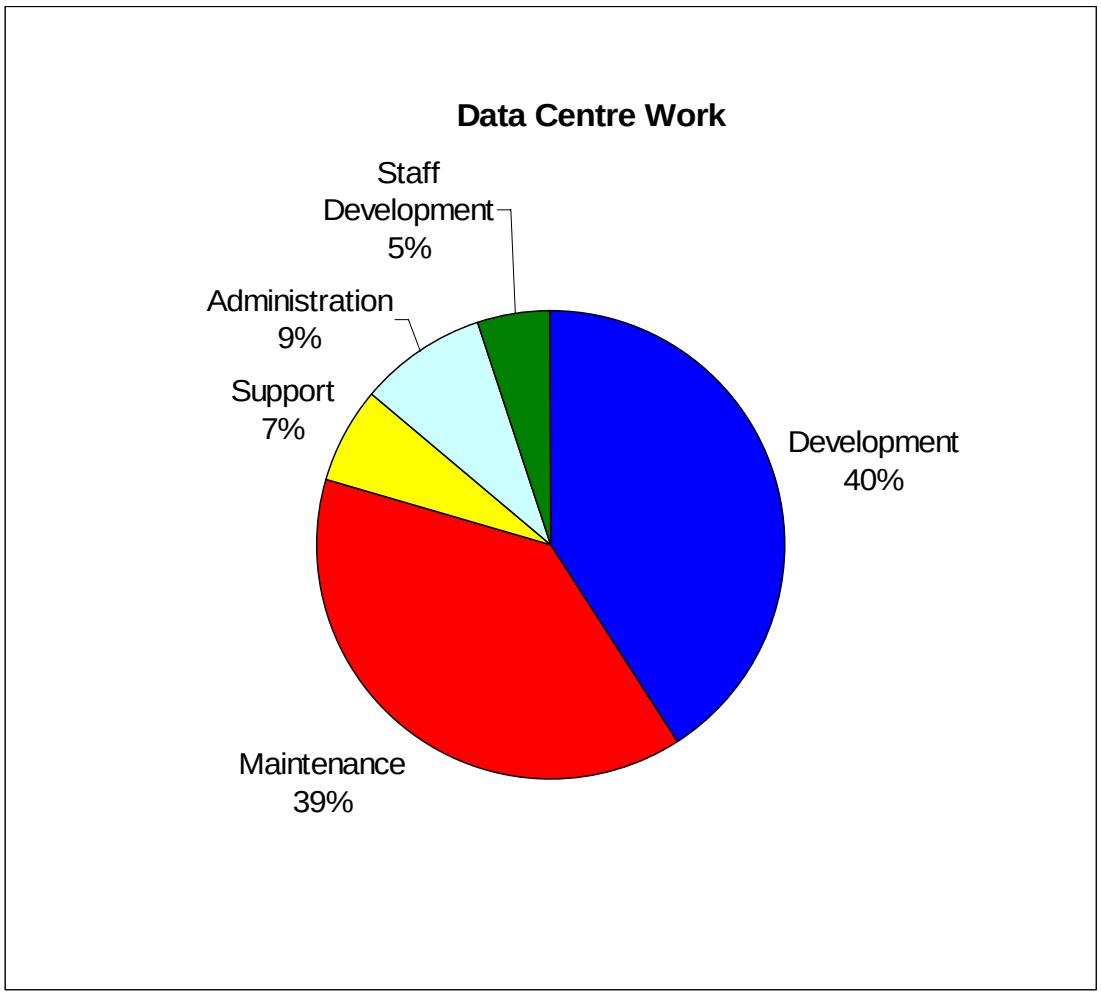
Table. Issues and Attributes of Centralized and Distributed Database Systems.

Annex 9: Data Centre Workplan 2005

Activities	Total Man-days
MoU's / Contracts	
AMAP data handling	0
CEFAS-DOME (see Internal Tasks-DOME)	0
ICES-Fishmap	22
Basis for cooperation w/ EC	0
EEA MoU	6
STATLANT programme	32
FAO MoU	0
HELCOM COMBINE Data	60
HELCOM MoU	0
IBSFC MoU	0
IMR-NewIFAP	595
IOC MoU	15
NASCO MoU / Contract	0
NEAFC MoU	0
OSPAR MoU	0
OSPAR Work Programme	45
PICES MoU	0
QUASIMEME	10
SAHFOS Digitisation of Plankton data	3

Activities	Total Man-days
RESOLUTIONS	
MCAP	0
SGQUA	23
HAWG	0
WGBFAS	45
WGNSSK	10
WGMHSA	0
WGNEW	0
WGSTAL	0
WGRED	0
SGQAB	1
SGQAC	4
SGQAE	1
WKIMON	4
SGMID	20
WKEUT	4
WGFAST	1
WKSAD	7
PGHAC	0
WGPE	0
WGZE	0
WGPBI	0
NORSEPP	0
WGOH	8
WGMDM	21
IBTSWG	7
REGNS	0
WGMS	0
WGSAEM	0
MCWG	5
SGNSBP	2
WGMHM	1
WGEXT	0
WGBEC	11
SGINC	0
WGPDMO	0
WGFE	0
WGBEAM	6
SGSIMUW	0
WGBIFS	0
SGMAP	0
Other-4ACE02	5
Other-4C03	0
Other-4D04	0
BWGDDP	20
Other-2CSY01	0
Other-2ASY08	0

Activities	Total Man-days
INTERNAL TASKS	
Development (new products)	
Screening/data check programme	90
Direction of Oceanography	50
DOME Phase 1	300
DOME Phase 2	0
DOME Phase 3	0
Free Format data load trial	25
Acoustic data in DATRAS	0
Rec12 web-enabling	15
Publications db	5
Action plan & resolutions db	0
CONC audit db	0
Maintenance (existing products)	0
Environmental data load	85
Oceanographic data check & load	600
Historical Fish Catch data	95
DATRAS data loading	125
DATRAS - updates of web pages	15
Retirement of BITS, IBTS	15
NewIFAP maintenance (Standard Graphs)	20
ICES Integrated Inventory maintenance	10
IT: SQL administration	10
Address db	5
Misc. administrative system support	25
Support	
Support to WGs	15
Misc. data requests	33
Fish Assessment CD	5
ASC support & participation	15
Administration	
ICES meetings	21
Management of Data Centre	50
DC meetings	64
Oceanography meetings	16
Environmental meetings	16
Staff Representative	20
Staff Development	
ICES seminar	11
DC seminars	21
Orientation of new employees	5
Staff Training	53
Staff one-to-one meetings	58
Total planned work	2886


Data Centre Work
Staff Development 5%
Administration 9%
Support 7%
Development 40%
Maintenance 39%