ICES SGMID Report 2005

ICES ADVISORY COMMITTEE ON ECOSYSTEMS
ACE:03

Report of the Study Group on Management of Integrated Data (SGMID)

11-13 April 2005

Lisbon, Portugal

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

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer
H.C. Andersens Boulevard 44-46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

Recommended format for purposes of citation:
ICES. 2005. Report of the Study Group on Management of Integrated Data (SGMID), 11-13 April 2005, Lisbon, Portugal. ACE:03. 39 pp.

Contents

1 Introduction

1.1 Terms of Reference (ToR)

The Study Group on Management of Integrated Data (SGMID) (Co-chairs: Peter Wiebe, USA, and Christopher Zimmermann, Germany), met at IPIMAR, Lisbon, Portugal from 11-13 April 2005 to:

a) Propose an update of the Data Policy;

b) Review the development within ICES towards integrated databases of oceanographic, environmental, and fisheries data;

c) Identify data sources relevant to (b), not yet integrated into the ICES databases;

d) Review existing integrated data systems for fisheries/environmental data and review data integration work in existing projects inside and outside of ICES;

e) Propose strategies and technical solutions for integrating available data including the possibility that data are not physically located in one site;

f) Evaluate and recommend the level of integration and aggregation of data in connection with management issues from an ecosystem perspective including the use of GIS systems;

g) Evaluate problems associated with the accessibility of data.

h) Review the progress of integrating the ICES database system into networks like SeaSearch II (*now SeaDataNet*).

1.2 Participants

Peter H. Wiebe (co-chair)	USA
Christopher Zimmermann (co-chair)	Germany
Larry Atkinson	USA
Richard Ayers	UK/England
Christian von Dorrien	Germany
Steve Flatman	UK/England
Rainer Froese	Germany
Ole Folmer	Denmark
Ibon Galparsoro	Spain/Basque Country
Martina Hennessy	Ireland
Alberto Murta	Portugal
Lesley J. Rickards	UK/England
Reiner Schlitzer	Germany

participating from the ICES Secretariat:

Julie Gillin (part time)	(Data Centre Manager)
Hans M. Jensen	
Henrik Kjems-Nielsen	

1.3 Structure of the report

This report is split into three major and two minor chapters. The introduction deals with standard items and reviews progress on various levels inside and outside ICES since the last SGMID meeting. It also updates the list of information/databases considered to be required for ICES' work (mainly for formulating the advice as agreed with client commissions), but

currently lacking or not accessible to ICES WGs. The second chapter gives a justification for the need of an updated ICES data policy and an evaluation of the problems associated with the accessibility of data. A draft of a new ICES data policy for further consideration by the ICES community, Bureau and Council can be found in Annex 4. Chapter 3 deals with conceptual and technical issues, some of them answering specific questions brought forward to SGMID by the ICES Data Centre. Chapter 4 contains a number of other items, namely a review of the ICES Data Centre's questionnaire planned to be distributed at the Annual Science Conference 2005; report on the preparation of a theme session at the ASC 2006 that is co-sponsored by SGMID, and a brief discussion on SGMID's future. This year's report closes with recommendations (including the proposed ToR for a 2006 meeting) and annexes.

1.4 Review of progress on issues related to SGMID's work during the past year

1.4.1 Developments within ICES (ToR b & h)

A. ICES Data Centre Developments 2004-2005

The ICES Data Centre was conceived as one of three main functional units in the re-organization of the ICES Secretariat that was implemented during 2004. The Data Centre was established by uniting all data and programming staff, and employing a manager.

The *initial* goal of the Data Centre is to fulfil previously commissioned work. The main contractual obligations are:

- HELCOM – COMBINE program
- OSPAR – Workplan 2005
- Development of a new fisheries assessment support system (InterCatch under contract with IMR and in collaboration with DIFRES)
- Development of an integrated Database on Oceanography and Marine Ecosystems (DOME under contract with OSPAR and CEFAS).

Other primary pursuits of the Data Centre are:

- Catch up with the Oceanographic data backlog that has accumulated over some years
- Maintain fishery data, i.e. the new trawl survey system DATRAS and preliminary catch statistics. The preliminary catch database (holding catch data only for the most recent year) is now accessible on the internet with restricted access.

The Data Centre has implemented a cross-disciplinary generic approach to development. Doing so serves a number of purposes:

- facilitates the ecosystem approach via common data structures, references, etc.
- minimizes cost: fewer modules/systems to develop, maintain, support & train
- increases capacity: less specialized knowledge necessary; less for users to learn; less support required
- maximizes system quality: wider testing and use increases robustness; greater familiarity means fewer operator errors and more reliable support.

To implement this approach, new system developments exploit the following systems:

- **Accessions:** registration and documentation of data submissions
- **DATSU:** screening of data prior to insertion in ICES databases
- **RECO**: reference code database

Data submission requirements are being harmonized:

- DATRAS holds data for three coordinated survey groups previously held by separate systems.

- Free formats for environmental data: an experiment has been initiated to determine the feasibility of using free formats for the exchange of environmental data. Free formats have long been used for oceanographic data.

Finally, the data centre actively seeks to extend and enhance capacities by

- Working across disciplines to build-up multi-disciplinary capacity.
- Collaborate with experts and system stakeholders by participating in relevant Expert Group meetings.
- Involve experts in system development tasks such as user requirement design specification, and system testing. These activities will also improve overall system quality and user satisfaction.

B. Bureau Working Group on the Data Development Project (BWGDDP)

An ad-hoc Bureau Working Group on the Data Development Project (BWGDDP) met in early February 2005 to outline the data, databases, and software tools needed to produce advice and support quality assurance procedures. This together with other input determined by the group forms the basis for a synopsis of a data development strategy and business plan for the coming 5 years. Among others, the chairs of the SGMID and the ICES Data Centre Manager participated in the BWGDDP meeting, and elements of SGMID's 2004 report were utilised by the BWG. The BWGDDP's draft report was presented to the Bureau on February 9. It has been amended in accordance with Bureau input. Some proposed activities have been or are being conducted already. The outcome of those activities and final refinements will be incorporated into the report at the BWGDDP's May 12 meeting. A final report is expected shortly thereafter. Although this final report was not yet available at the SGMID meeting, the SGMID was briefed on the report, its background and proposed actions, and, most importantly, the BWGDDP's expectations of the SGMID.

The BWGDDP's vision for data management is threefold:

1. **ICES will remain a focal point for marine data in the North Atlantic.**
2. **ICES will create a portal serving as a hub for distributed data.**
3. **ICES web portal will become more attractive to the science community.**

Implementation of this vision implies greater exploitation of decentralized or distributed databases. Centralized databases still have an important, albeit more specialized, role as a secure harbour for aggregated and/or anonymous data; as an archive for valuable data at risk of extinction; and, last but not least, for the continued maintenance of ICES' traditional (low maintenance) databases. Achieving the BWGDDP vision also requires increasing awareness of and collaboration with other organizations including national laboratories and international initiatives such as SeaDataNet.

The BWGDDP's discussion illuminated a number of issues needing attention via a relevant data policy.

Access:

- Due to new legislation, technical advances, and changing attitudes, rights of the public warrant more open data accessibility.
- Individual security needs of data sources must be respected. These needs may conflict with the rights of the public.
- Data sources must be given due acknowledgement.

Data Quality:

- Speed from observation to availability of data is gaining importance as a quality factor.
- To achieve faster speeds, information about data quality may be an adequate substitute for the actual quality measures.

Functionality:

- Universally applicable and accessible technologies should be exploited whenever possible to optimize the possibilities for data use. Practically speaking, we should aim for platform independence and use open source software whenever feasible.

C. Progress on database development in the ICES data centre

At its last meeting SGMID emphasized the urgent need for a fisheries catch-at-age data handling and collation database. The development of this database started in late 2004: **InterCatch** is a system mainly for use by fish stock assessment working groups. The system can aggregate national commercial fish species catch and biological information (like catch-at-age) to international catches. The output of InterCatch is information on total removal of fish from all stocks assessed by ICES at various levels of disaggregation, including fleets, age/length, sex. This output can be used directly for input to various assessment models currently in use. The reasons for developing such a system are to:

- Provide standard tools for the manipulation and aggregations of data.
- Apply standard checks to assure high quality data.
- Store and log data to make output data reproducible.
- Facilitate the delivery of fleet based advice

The system consists of a web interface and a database. The web interface will be platform independent and developed in VB.NET and C#.NET. The database management system used is Microsoft SQL Server 2000. There will be restricted access to sensitive data, but public access to aggregated and derived data. National data submitters upload data to InterCatch using the web-interface. Data is automatically checked and errors reported. The expert group-appointed stock coordinators can apply advanced data checking and minor value changes. The stock coordinators extract all stock information from the species/area raw data and fill-in unsampled catches following the appropriate rules. The final international aggregated catch and biological information is exported for the use in assessment models. InterCatch is being jointly developed by DIFRES and the ICES Secretariat. It is based on FishFrame (a DIFRES development), and DATSU and RECO (ICES Data Centre standard utility programs). The User Requirement Specification document has been written by a group of assessment experts, development has started, and the system is expected to be tested and functional by the end of 2005.

The **Database on Oceanography and Marine Ecosystems (DOME)** is an integrated database system with the potential for storing data from the 3 data categories currently held within the ICES system:

- Phase 1 includes the import of biological contaminants and effects data (incl. fish disease) together with hydrochemistry data from the oceanographic database.
- Phase 2: Import of biological community data together with CTD and underway data from the oceanographic database.
- Phase 3: Investigate the need and feasibility of including some of the fisheries data.

The first phase is scheduled for 2005 and includes implementation of the DOME database, which uses **Accessions,** an ICES Data Centre standard utility program. The Accessions database stores information on submissions sent to ICES. Also included in Phase 1 is the programming of internal and external interfaces to the database. The internal interface is to be used for importing and exporting data to/from DOME, It is a tool for use within the ICES data centre. The external web interface will allow external users to export data from DOME.

Exported data will be formatted in either its data specific native exchange format or as comma-separated files. Import of data to DOME will require the data to be formatted in either the 3.2 format (for environmental data) or ICES Oceanographic Format (for oceanographic data). Conversion programmes will be created to facilitate the screening and importing of data in former ICES file formats, and eventually, possibly, other non-ICES formats.

Flexibility is a major feature of the DOME database design. DOME is structured with some common top level tables defining the collection event in time and space. Code systems, method- and data restriction information is stored in a set of common tables. These tables are also common for all data categories. Each data category, however, has its own specific data tables.

Integrating the data from the separate databases has given rise to some challenges. A sub-set of the hydrochemistry oceanographic data is also contained in the environmental data; this creates a potential risk of duplicate data. The oceanographic hydrochemistry data have been processed to a higher quality level than its duplicate data in the environmental dataset, however the environmental data have some additional information attached to it. The data sets will have to be integrated without compromising the quality of either data set.

Another challenge is the handling of different units. Inside DOME all parameters will be stored in a standardized parameter specific unit. This necessitates the conversion of values, because different units have been used for the same parameter. A precise conversion may not always be possible and measures need to be taken in order to make it possible to standardize these data.

Because oceanographic data use the BODC data dictionary and environmental data use RECO (ICES reference code) system, it will either be necessary to produce a mapping table (a very complex task) or continue to use both code systems. Using two codes imposes some limitations on the system.

D. ICES Expert Groups

A number of ICES Expert Groups**,** e.g. those co-ordinating international fisheries surveys, are important data contributors to the ICES community. Some of these groups have agreed on their own data policies, covering their specific data sets. SGMID is aware of amendments to data policies of at least two groups in the course of the last year. Participants of the International Bottom Trawl Survey WG (IBTSWG) deliver, at present, only fishery survey held in the central ICES database DATRAS. Its data policy has been rather restrictive, which resulted in serious access problems to IBTS data, not only for the public and external scientists, but also for some groups inside ICES. The IBTSWG has now modified their data policy (see annex 3) aiming at opening aggregated and standardised data to a wider public, and to ease access to more disaggregated data for ICES groups. A number of restrictions for all potential users remain however: data prior to 1982 and the three recent years' data will not be accessible to scientists other than those contributing data into DATRAS.

Another ICES group has recently chosen a different approach: The Planning Group for Herring Surveys (PGHERS) coordinates herring acoustic and larvae surveys in the North Sea, west of Scotland, the Skagerrak/Kattegat, and the Western Baltic. This group has decided to make all aggregated data (by rectangle) from the acoustic surveys available to the public not later than 10 months after the latest survey (see Annex 3). This is the time required to collate and validate the survey data, which includes discussion by PGHERS, the Herring Assessment WG and finally ACFM following HAWG. The group thought that this would give ample time to omit errors, and that it would be unlikely that major errors would have to be corrected regularly after the data has been used in the assessment. The limitation to a rectangle-aggregated dataset reflects the present capabilities and inventory of the HERUR database. PGHERS decided to make this part of its data, thought to be most useful for the scientific

community, publicly available as soon as possible. Routines to handle, aggregate, and extract raw data have to be developed in a later step.

E. Integration of the ICES Data Centre into external programs

ICES is one of the forty-nine partners in the SeaDataNet proposal (see below for a brief outline of the project) that was submitted to the EU on 3 March 2005 under the call FP6-2004-Infrastructures-5. The project aims to provide a pan-European infrastructure for distributed ocean and marine data management. The initial result of the proposal should be known within a few months. If it is unsuccessful, it will be resubmitted in an appropriate form to GMES (Global Monitoring for environment and Security; a joint initiative of the European Commission and the European Space Agency).

1.4.2 Meetings and Developments outside ICES (ToR c & d)

A. Ocean Biodiversity Informatics Conference (Hamburg, 2005)

The Ocean Biodiversity Informatics Conference was a forum for marine biological data managers to discuss the state of the field and to exchange ideas on how to further develop marine biological data systems. Conference topics included marine biological data management - taxon-based, biogeography and also environmental, non-taxon based data management, including information system development.

Specific objectives of the Conference were to:

- bring together biological data managers to discuss the present state and progress, in this field since the meetings in Hamburg (1996) and Brussels (2002)
- provide an opportunity for biological data managers to find out what is happening at international organisations
- discuss potential gaps and overlaps in the taxonomic and geographic scope of existing data systems
- discuss standards and protocols for data exchange
- discuss how to integrate data from separate databases
- learn how these integrated databases have provided new insights

A conference statement was produced encouraging increased availability, sharing and acknowledgement of data – this is available on the OBI web site (http://www.vliz.be/obi).

B. The SeaDataNet Proposal

The SeaDataNet Project aims to develop an efficient distributed Pan-European Marine Data Management Infrastructure for managing large and diverse marine data sets. The objective is to network the existing professional data centres of thirty-six countries bordering the European seas, active in data collection, and to provide integrated databases of standardized quality on-line. In the marine domain, multi-disciplinary data are collected by many organizations using various heterogeneous observing sensors installed on board research vessels, submarines and airplanes, fixed and drifting platforms, and satellites to measure physical and geophysical parameters (i.e., temperature, salinity, current, sea level, optical properties, magnetics, gravity, bottom depth), chemicals, biological species, etc.

The online access to in-situ and remote sensing data, meta-data, and products will be provided through a unique portal interconnecting, in the first phase, 11 interoperable node platforms. The development and adoption of common communication standards and adapted technology will insure the platforms interoperability. This activity will be developed to gradually connect all the other data centres to the interoperable system.

The quality, compatibility and coherence of the data issuing from so many sources, will be insured by adopting standardized methodologies for data checking and by dedicating an important part of the networking activities to training. As raw data may be difficult to interpret by non-specialist users, statistical gridded products will be prepared from the most exhaustive data sets of in-situ and remote sensing observations, which will be made available by the participants.

C. Natural Environment Research Council (NERC) DataGrid

The objective of the NERC DataGrid (ndg.badc.rl.ac.uk) is to create a seamless virtual data centre across the NERC Virtual Organisation. It is a 3-year project that commenced in October 2002. The initial partnership for the project is the British Atmospheric Data Centre (BADC), the British Oceanographic Data Centre (BODC), and the Central Laboratory of the Research Council (CLRC) e-Science Centre. In addition, there is collaboration with the US Earth System Grid community.

The NERC DataGrid Strategy has been as follows:

- Develop a metadatabase based on ISO19115 compliant Directory Interchange Format files (DIFs) (XML)
- Hits returned as browsable XML metadata for user selection
- Design data sets generated by dynamic aggregation from source granules driven by XML data descriptions
- Data delivered to the user documented with full XML metadata

D. Improvements in the integration of oceanographic, fisheries, and biological data at the AZTI Foundation

The “ItxasGis system” is a corporate system to gather all marine data generated in AZTI and other institutions on the Basque Coast and the Bay of Biscay. Its principal aim is to promote access to environmental data to both end users and administrations using high quality tools such as GIS and web pages. It is being financed by the Basque Government under the Basque Country Environmental Programme.

One of the two main steps taken is the implementation of a personalised application made in ArcGis for data management and integration. This customisation allows managing of information available in the system, editing and importing metadata, and the specification of a data accessibility profile for each user. The customisation allows database managers to be informed about the integrated information and who has done it; moreover, researchers in the organization are going to be informed periodically about new information available.

A second module is being developed to allow users to query information available in the database. This part is being developed using open source GIS and web programming technologies. The user is able to query the data available by searching the information graphically as defined in a rectangle on a map or by writing the spatial extent numerically in a geographic or UTM projection. It is also possible to query for information by date or depth range. The system provides the user with the information about the established conditions. At this step user can click and access the metadata of each datum selected. Given the user selected information, the system creates a zip file that the user can download. The system is also able to provide database managers with information about the data that are being downloaded and who has downloaded them, thus providing valuable information about data requirements of end users.

The expected results and future actions include the programming of other applications that are able to display and mapping data using web technologies and open source map servers,

adapting existing data models to specific requirements, metadata development, and migration between different formats (FGDC and ISO19115).

1.4.3 Databases not integrated/accessible for the ICES community, or not developed yet but needed for formulating the advice or creating status reports

SGMID elaborated on this issue at its last meeting. Apart from the development of a centralised fisheries data database (InterCatch, see above), there is very little progress to report on. Some expert groups have undertaken further steps to make their data available to a wider public. There was an attempt in the course of the past year to submit collated zooplankton data to ICES by an experienced external data manager, which proved to be exceedingly difficult. The Data Centre is aware that some data submitters experience difficulty applying the formats, and thus the Data Centre has initiated a free format trial on zoobenthos data in order to evaluate the feasibility of flexing or removing format requirements.

2 Draft of a new ICES data policy (ToR a)

2.1 Why a new ICES data policy is needed

It is widely recognised that ICES has to update its data policy, which is more than a decade old and doesn't reflect the way the scientific community is presently handling data. None of the key groups subject to the data policy – the ICES Data Centre, data contributors, and expert groups/external data users – is content with the current policy and the procedure to access data. SGMID was asked to provide a draft new ICES data policy for further consideration by the ICES community, as one of the main goals for its 2005 meeting.

Any policy must be kept current with changes in legislation, technical advances as well as commonly agreed best practices. Therefore, it follows that the policy and related information should be reviewed at regular intervals and whenever major events warrant such a review.

Implementation of the data policy may require additional resources of the ICES Data Centre, the expert groups, and possibly also of the member states' laboratories. This aspect should not be overlooked in considering the policy proposal itself and its timely implementation.

A key element of SGMID's proposal for a new ICES data policy is open access to data in a timely way. This is perceived to add value to the data, to increase the use and usability of the ICES data portal, and to secure ICES' position as the focal point for marine data in its area. However, it has to be considered that the ICES Data Centre will in most cases not be the data owner (all rights will remain with the originator), but only the collator, distributor, and user. This will require negotiations with the data source during a proprietary period, as well as a potentially complex system of access rights for sensitive data.

ICES recognises its responsibilities for long-term maintenance and access to relevant data. Data are envisioned to have a life cycle that includes initial data collection planning, acquisition, and processing, the submission of data to the ICES Data Centre and their incorporation in or linkage to the appropriate database, the maintenance of the data while being actively used, and the final stewardship of the data. Additional details on this end-to-end data management concept are listed in Annex 3.

The proposal for a new ICES data policy is given in Annex 4 as a self-contained document. This will be considered (and amended if required) by the Bureau Working Group on the data Development Project (BWGDDP) in May, and further by the Bureau and the Council later in 2005.

2.2 Problems associated with the accessibility of data (ToR g) and the benefits of open access

Traditional individual or institutional accumulation of scientific data has caused great losses of knowledge over time, mostly due to a lack of proper long-term storage and archiving, and lack of accessibility by both the scientific community and society at-large. The fact that data can be used for many more and often unforeseen purposes has largely been ignored. This is also true for data produced within the ICES community, where access to such data is difficult even for ICES expert groups and restricted for others. For example, the scientific use of fishery survey data held at the ICES Data Center for scientific purposes outside established expert groups requires approval by all data contributors (up to 19) whose data are affected. Refusal by one member is sufficient to prevent providing any data at all. Also, processing these requests puts a considerable burden on the ICES Data Centre and ICES members and is time consuming.

The current ICES data policy stems from a time when the main goal was to persuade members to share data for use by the working groups and thus a premium was given to confidentiality. Time has come to review this data policy in the light of a general recognition of the need for data sharing and open access to non-sensitive data, as well as legislative mandates. While doing so, remaining concerns of data owners must be resolved. Most of these concerns can be addressed by initially providing sensitive data in an aggregated form for a limited period of time. Also the benefits deriving from sharing of data need to be recognized and most data custodians who have made their data publicly available will agree that in hindsight that move has been overall beneficial to themselves and their institutions. For example, some data owners use public access as a no-cost means of having their data peer-reviewed: feed-back from users helps identifying weaknesses in the data and assists in correcting errors. Online availability of data usually reduces the number of general requests for information, because much of what users need to know is readily available online. In summary, more exposure typically results in more visibility, recognition, invitations, citations and projects. Thus, from SGMID's perspective, there is no alternative for open access to as much data as possible in the ICES community.

3 Technical solutions (ToR f & e)

3.1 Concepts for Data Integration

3.1.1 Integration at ICES

The SGMID discussed present integration activities currently underway at ICES, in particular the Database on Oceanography and Marine Ecosystems (DOME). This will be a centralised database, initially including the oceanography and environmental data, but eventually including other data, for example, fisheries and biological community data. There was much discussion about commercial fisheries data, which are generally available in a highly aggregated form (e.g. covering an area rather than individual points and a time period rather than a particular date and time), and may not easily be included in the same database. However survey data collected by research organisations at a much more detailed level could be included. Intellectual Property Rights issues must be considered for all kinds of data.

It was noted that it is not a simple matter to join together the Environmental and Oceanographic Databases. Some data are duplicated in the databases, however with different quality controls. In addition, within the Oceanographic Data Set itself, some data in the high resolution data set (CTD) are also in the low resolution data set (derived from the CTD data). The SGMID was concerned that there should be a procedure in place to manage this process (if appropriate involving the data originators) and that there should be no compromise on the data quality. An audit trail should be available, so that if problems occurred, it would be possible to go back a step if necessary. When combining the Environmental and

Oceanography databases, it will be necessary to check for duplicate data and consultation with WGMDM would be useful as there is much expertise in this Group for methods of duplicate data checking. This is not a trivial task, in particular when determining near-duplicates. There was also discussion relating to submission of, in particular fisheries data, and the procedures in place to handle the situation where data were submitted immediately after collection prior to quality assurance for immediate use by a Working Group and then resubmitted after the data had been quality assured. It was not clear if the stage of quality control of the data could be easily identified in the ICES databases. The SGMID agreed that the ICES Data Centre should have a documented procedure for replacing data in their databases and that they should keep a record of the status of data in the databases (e.g. raw, non-quality controlled; final quality controlled, etc.). The DATRAS and InterCatch data systems should make this easier.

3.1.2 Integration with data not held physically at ICES

There are many other sources of relevant data, which may be of use to ICES and ICES users, not held at the ICES Data Centre. Conversely, ICES may wish to contribute to other systems or databases. There was some discussion relating to metadata and automatic cataloguing for distributed systems. The SGMID recommends that one system in particular, DiGIR (Distributed Generic Information Retrieval), used by the Ocean Biogeographical Information System (OBIS) and GBIF, should be further investigated and possibly installed at ICES. DiGIR records are made available in a standardised XML format and are transferred via HTTP. Use of DiGIR at the ICES Data Centre would enable users to access data in a standardized way and link into the data available from OBIS, which provides global geo-referenced information on marine species. Members of the WGMDM could be consulted on this as at least one member has already installed the DiGIR software and others are planning to do so.

ICES could also play a part in contributing to other metadata systems, for example, EDMED – a metadata discovery system – and the NSDI (National Spatial Data Infrastructure in the USA). Links could also be made to other on-line data sets, for example, Argo.

The ICES secretariat is a partner in the SeaDataNet proposal submitted to the EU (as described above). This project will implement a distributed data system across 30+ countries in Europe and will allow ICES to link to a range of other marine databases at major national oceanographic data centres and oceanography and fisheries organizations.

3.2 Geographical Information Systems

Given the diverse range of data available from ICES and the growing need to analyse data from an ecosystem perspective, GIS offers many potential benefits to end-users of ICES data. Spatial analysis and the integration of diverse datasets are facilitated by GIS tools and thus offer many benefits for researchers working with scientific data.

One of the main benefits for users would be the provision of a web-GIS interface for display and querying of data held within ICES. This type of interface has the potential to be more user-friendly and intuitive for users than the current system. It would potentially facilitate greater access to and increased use of data. Initially the use of a simple data access interface should be investigated. Over time and with feedback from users, this could be developed into a more sophisticated search interface.

The other consideration in relation to GIS is the possibility of delivering data in a standard GIS format e.g. shape file. This may not be appropriate for all data types, e.g. vertical profile data, but should be investigated. So, for example, a shape file could include the positions at which profile data were collected, or the temperature/salinity values at the surface or at some other depth, but would not be appropriate for the entire profile, as currently the depth dimension is not easily handled by standard GIS files. This approach could facilitate multi-

disciplinary use and analysis of data. In addition, development of software code to output the data to formats compatible with external visualization programs like Ocean Data View should be encouraged.

In order to assess the demand for either of these options, questions on the use of GIS could be added to the questionnaire that the ICES Secretariat plan to circulate to users (see Section 4.1). If demand is demonstrated, this could provide a driver for ICES to move forward with developments in GIS.

The ICES Data Centre is currently evaluating GIS options as part of the OSPAR Work Programme for 2005. Based on this, a technology will be chosen taking into account Open Geospatial Consortium (OGC) compliance, current database formats, software costs, maintenance and development skills required. Once a decision is made, it is suggested that ICES should seek some advice/support for implementation from members of the ICES Community who already have experience with that technology. It was noted that the WGMDM had examination of GIS in marine data systems as one of their Terms of Reference, and input should be requested from them. For ICES users the format of exchanged files should be as widely accepted and applicable as possible, and technical restrictions for accessing data should be minimal. To achieve this, open formats should have a high priority. This could be achieved by the use of both commercial and open source software products.

In the design of any future databases within the ICES Data Centre, it is recommended that consideration be given to providing data access via a GIS interface. A common approach to the management of spatial information, taking into account the three dimensional axes and time, should be considered within all systems.

3.3 Quality Flags

There was agreement among participants that the scientific value of integrated datasets critically depends on the quality and precision of the data. For scientists to perform analysis and provide sound advice they need to be aware of and have the ability to select data based on known levels of quality. The precision level required may vary for different analyses, for example a fisheries biologist using salinity data combined with stock abundance may be prepared to accept a lower level of data quality than a physical oceanographer modelling water masses.

Some data requests may require a simple good data / bad data selection whereas other requests may define a much deeper level, for example precision of instruments used for the data collection or the analysis method undertaken.

There are examples of what were thought to be ‘good’ data sets being marred by the removal of what were initially thought to be bad data points, over time these ‘bad’ data points were recognised as being the first data points in a previously unrecorded event. It is important that data points are flagged as being ‘questionable’ or of a lower quality level rather than being removed from the dataset. This highlights the need for quality flags to be incorporated into integrated data sets.

Data quality flags can be categorised into subjective and objective. Objective quality flags are based on agreed analysis of the data with for example resulting values being within an acceptable range. Subjective flags are based around an expert interpretation of the data (possibly including collection or analysis methods coupled with results of an objective analysis).

A system of quality flags may comprise a combination of subjective and objective flags. Experts familiar with the data as well as its use must define both subjective and objective

quality criteria. Data managers and centres can implement quality flags only if the criteria are unambiguous and derivable from the given data.

For any system of quality flags to be useful it has to be based around agreed descriptions of each quality level in order that each data originator can define a data quality level that is comparable with the other data elements in the integrated data set. The SeaDataNet project will consider standardisation of data quality flags.

The study group realized that implementing quality flags has impacts both on data centres hosting and supporting datasets, but also on the data originators. Good coordination between data originators, data centres, and data users is essential as are long-term resource commitments at all stages of the data flow.

Implementation of data quality flags does not infer an immediate requirement to report quality flags but allows for a phased increase in their use as the data set grows. The lack of a quality flag does not indicate 'bad' data, but indicates that the quality has not yet been assigned.

In the evaluation of data quality, SGMID recognized that there are three levels of data screening. The first is by the originator that has collected the data and/or the data submitter. The second is by the data centre that has a broader view and can spot errors that are difficult to observe by the originator, for example, by cross-organizational comparisons. The third is by the user, who often has an experience level or an application that can spot errors that others have missed. The data centre needs to facilitate the correction of data from all three sources.

3.4 Advantages and Disadvantages of the use of codes in databases

The SGMID discussed the advantages and disadvantages of using codes in databases. Presently ICES uses 3 code systems. An internal integrated code system (RECO) is used for environmental and fisheries data. Besides the integrated code system two external code resources are used in ICES:

- For oceanographic parameters, the BODC data dictionary is used.
- For species codes, the ITIS code system is used.

The Group agreed that unambiguous definitions and code systems are a necessity to maintain efficient and reliable database systems, but that codes visible to inexperienced users may reduce the acceptance of a database. There is an initial overhead to establish codes and to maintain them, but the benefit of using them outweighs this over time. They are useful because they enforce data discipline by preventing multiple or diverging entries for the same parameter/species/…, which could have a negative impact on the quality of the database system. Using a code system does not necessarily imply that the user should be confronted with a complex numbering system when interacting with the database. For example a Taxonomic Serial Number (TSN) number might be the data stored in the database, but the user may only be presented with the scientific name. Ease of use and acceptance of any database system are important aspects in this respect.

There is a need for standardization of code systems and/or mapping within or between databases in an organization, and across different organizations. The mapping between existing systems can be very complex and the group recognizes that a pragmatic use of coding systems is necessary at the present stage.

4 Other Business

4.1 Review of the ICES secretariat's questionnaire to potential data users

SGMID welcomes the ICES secretariat's initiative to explore the needs and requirements of present and future users. It supports seeking professional help for the general layout of the questionnaire to avoid leading questions or questions that lead to ambiguous answers. There are a number of suggestions for this survey raised by SGMID members:

- Users that have previously requested data from the ICES Data Centre should be included in the survey, and the use of ICES data for individual projects (within and outside ICES) should be analysed by the ICES secretariat, by means of an analysis of filed requests.
- A section should be added to the Introduction (first page) on what questions the ICES secretariat would like to answer with the survey, e.g. "who are the users?", "what kind of data do they like to use?", "in which form/format do they require the data?", "what can we do to ease expert group's work when producing advice and status reports?" and "what can we do to increase interest in ICES data?"
- Specific comments: Part 1, Question 3 and 4 need an "other" entry with the possibility to specify this; Part 2 needs an explanation of the type of data (with examples if possible) if users outside the ICES "inner circle" are to be approached; Part 2, Questions 7 and 8 should be omitted (Q 7 will certainly always be answered with "yes", and the difference between distributed and centralised databases and all transitions in between are in most instances not visible to the user any more.
- The questionnaire should include an element asking for the needs of a geographical information system on the ICES website, and if this is perceived to ease or improve access to ICES data.
- One of the potential professional agencies for designing such a questionnaire might be the Danish Institute for Fisheries Management and Coastal Community Development (IfM).

4.2 Preparation of the ASC 2006 theme session on data integration

SGMID proposed last year to co-sponsor a theme session with WGMDM on "Environmental and Fisheries Data Management, Access, and Integration" at an upcoming Annual Science Conference. There is a recognized need for enhanced integration of diverse kinds of fisheries, oceanographic, and other marine environmental data and the development of tools to enable fisheries and environmental assessments. Increased development of a multiplicity of databases and data sources that are distributed throughout the North Atlantic and the world at large requires that the new approaches be developed to enable users to access the data quickly and efficiently, especially if the data is distributed over different databases. Management of the data also requires new approaches enabled by new technologies and especially the internet. The theme session is expected to attract a variety of contributors, from database specialist providing technical solutions for data integration through end users, presenting examples for successful applications and the use and visualization of integrated data.

The Consultative Committee accepted this proposal at its meeting in September 2004 for the ASC 2006, which will be held in the Netherlands. Due to time constraints this year, more detailed planning of that theme session was postponed to the ASC 2005 and SGMID's 2006 meeting.

4.3 The future of SGMID

The group discussed its future perspectives extensively. As no formal meeting could be conducted in the first year after the creation of the group (due to the lack of participants), SGMID to date has met only twice in two consecutive years. The group has now gained

momentum, comprises participants with a reasonable spread of expertise and background, and has not yet completed its work. In addition, experts for seabirds and mammals were not present at the 2005 meeting. With the proposed new draft data policy now to be amended and implemented, and with a theme session on SGMID's key topics scheduled for the Annual Science Conference 2006, the group proposes to reconvene for a third time in spring 2006. This meeting should be held back-to-back with the Marine Data Management WG (if possible with a one-day overlap) to facilitate a greater exchange of information between these two groups.

Beyond 2006, SGMID cannot meet as a Study Group any more, as the maximum lifetime of a study group would have been reached. This leaves a variety of options:

- Dissolve the group without replacement – this would mean that the expertise of the group and its capacity to support and give advice to the ICES Data Centre would be lost.
- Merge the group with WGMDM – this has the drawback that both groups focus on very different aspects: While WGMDM is a rather technical, hands-on group under the Oceanography Committee, SGMID addresses more conceptual and policy related questions under ACE. There might be scope for a change of the name of one of the groups (possibly the newer one, SGMID) to emphasize this difference.
- Transfer SGMID into a working group that could give advice to the ICES secretariat as required. This group could meet regularly, but not necessarily annually, and also by correspondence. For this option, it should be explored if another parent committee (currently ACE) would be more appropriate and would provide a wider dissemination of its work.

After discussion within the group and with the ICES Data Centre manager, the group currently prefers the third option. The chairs and the ICES Data Manager will take this up with the appropriate ICES bodies.

5 Recommendations

SGMID recommends:

(from Sect. 1) … to the ICES secretariat to continue with the exploration of possibilities to acquire data not currently accessible to the ICES community, but needed for the production of advice and status reports

(from Sect. 2) … to ICES to review and implement a new data policy as drafted in Annex 4. One of the key elements of the new policy is open access to as much as possible data, also aiming at minimising technical or financial constraints for access and use of data.

…to the ICES Data Centre and WGMDM that the DiGIR system (Distributed Generic Information Retrieval), used by the Ocean Biogeographical Information System (OBIS) and GBIF, should be further investigated and possibly installed at the ICES secretariat.

…that questions be added to the ICES questionnaire to ascertain the level of demand for access to data via web-GIS (Geographical Information Systems) and delivery of data in GIS formats.

…that the ICES Data Centre carries out an evaluation exercise and makes a selection of a web-GIS technology to enhance and facilitate data access via the internet, taking into account Open Geospatial Consortium compliance, current database formats, software costs, maintenance and development skills required. Consideration should be given to the management of spatial information in the design of all future databases by the ICES Data Centre, in order to facilitate GIS access to data.

(from Sect. 3.2)…to WGMDM to investigate and report back on their examination of the use of GIS in marine data systems and comment on developments using Open Source GIS.

(from Sect. 3.3)…that database developments within ICES DC should include provision for storage of data quality indicator(s) and that the exchange files supporting these developments include the necessary data structure. When data that contains data quality indicators are submitted ICES should retain this information within the ICES database and make it available to the data users.

… to the ICES expert groups and the Data Centre to conduct appropriate quality control procedures that can not be performed by individual data submitters. This includes consistency checks and comparisons between repeat cruises and the analysis of cross-over data. Problems should be reported back to the data submitter. Any correction or calibration change should be recorded in the associated meta data log.

from Sect 3.4) …that ICES maintain (or develop further) their standard definitions and codes. However, in order to promote the ease of use and acceptance of a database also for inexperienced users, codes should have to be used by data submitters and requesters only when unavoidable.

(from Sect. 4.3) …that SGMID meets again next year back to back with WGMDM to

- detail planning of the ASC 2006 theme session (including selection of abstracts if required)
- review the plans for implementing the new ICES data policy
- review status of data integration of external datasets currently not available to ICES but needed for producing the advice relevant to ecosystem based management
- review data policies of individual ICES expert groups and flag differences between those and the new ICES policy, and keeping abreast of new developments

… that SGMID change parent committee (currently ACE) to ConC or the Bureau, as this would better reflect the expected recipients of SGMID's report and recommendations

(other) …that sufficient funding is made available to ICES and national data centres to rescue data currently not available for ICES expert groups and the public. These include data available only non-electronically in historic ICES reports.

6 References

ICES 2004: Report of the Study Group for Management of Integrated Data. ICES CM 2004/ACE:05

ICES 2005: Report of the International Bottom Trawl Survey Group. ICES CM2005/D:05

ICES 2005: Report of the Planning Group for Herring Surveys. ICES CM2005/G:04

ICES 2005: Report of the Bureau Working Group on the Data Development Project. ICES CM2005/June 2005/Bur. Doc. 1419.

Acknowledgement

SGMID is very grateful for the provision of excellent meeting facilities by IPIMAR Lisbon and all the effort undertaken by its staff, especially Dr. Alberto Murta, to host the group.

Annex 1: List of SGMID participants

NAME	ADDRESS	PHONE/FAX	EMAIL
Peter H. Wiebe	Woods Hole Oceanographic Institution Woods Hole, MA 02543 USA	Tel: +1 508 2892313 Fax: +1 508 4572169	pwiebe@whoi.edu
Christopher Zimmermann	Bundesforschungsanstalt für Fischerei, Institut für Seefischerei Palmaille 9 D-22767 Hamburg Germany	Tel: +49 40 38905266 Fax: +49 40 38905263	christopher.zimmermann@ish.bfa-fisch.de or czimmermann@clupea.de
Larry Atkinson	Dept. of Ocean, Earth & Atmospheric Sciences Old Dominion University, Norfolk, VA 23529 USA	Tel: +01 757 6833472 Fax: +01 757 6835303	latkinso@odu.edu
Richard Ayers	CEFAS Fisheries Laboratory Pakefield Road Lowestoft, NR33 OHT United Kingdom	Tel: +44 1502 562244 Fax: +44 1502 513865	r.a.ayers@cefas.co.uk
Christian von Dorrien	Bundesforschungsanstalt für Fischerei, Institut für Ostseefischerei An der Jägerbäk 2 18069 Rostock Germany	Tel: +49 381 810350 Fax:	christian.dorrien@ior.bfa-fisch.de
Steve Flatman	CEFAS Fisheries Laboratory Pakefield Road Lowestoft, NR33 OHT United Kingdom	Tel: +44 1502 524245 Fax: +44 1502 513865	s.flatman@cefas.co.uk
Rainer Froese	IFM-GEOMAR Düsternbrooker Weg 20 24105 Kiel Germany	Tel: +49 431 6004579 Fax: +49 431 6001699	rfroese@ifm-geomar.de
Ole Folmer	Danmarks Fiskeriundersøgelser Charlottenlund Slot DK-2920 Charlottenlund Denmark	Tel: +45 33963469 Fax: +45 33963333	ofo@difres.dk
Ibon Galparsoro	AZTI, Herrera Kaia Portualdea z/g 20110 Pasaia (Gipuzkoa) Spain	Tel: +34 943480000	igalparsoro@pas.azti.es
Julie Gillin	ICES Secretariat H. C. Andersens Blvd. 44-46 1553 Copenhagen Denmark	Tel: +45 3338 6712 Mobile: +45 22444910 Fax: +45 3393 4215	julie@ices.dk
Martina Hennessy	Marine Institute Galway Technology Park Parkmore Galway Ireland	Tel: +353 91 730614 Fax: +353 91 730470	martina.hennessy@marine.ie

Hans M. Jensen	ICES Secretariat H. C. Andersens Blvd. 44-46 1553 Copenhagen Denmark	Tel: +45 33386736	hans.jensen@ices.dk
Henrik Kjems-Nielsen	ICES Secretariat H. C. Andersens Blvd. 44-46 1553 Copenhagen Denmark	Tel: +45 33386736	henrikkn@ices.dk
Alberto Murta	IPIMAR Av. Brasilia Lisbon 1449-006 Portugal	Tel: +351 21 3027120 Fax: +351 21 3015948	amurta@ipimar.pt
Lesley J. Rickards	BODC Joseph Proudman Building 6 Brownlow Street Liverpool, L3 5DA United Kingdom	Tel: +44 151 7954897 Fax: +44 151 7954912	ljr@bodc.ac.uk
Reiner Schlitzer	Alfred-Wegener-Institute for Marine and Polar Research Columbus Center D-27568 Bremerhaven Germany	Tel: +49 471 4831 1559	rschlitzer@awi-bremerhaven.de

Annex 2: Recent modifications of data Policies of ICES expert groups

A International Bottom Trawl Working Group (IBTS WG) – report 2005: ICES Policy on Access to DATRAS (Sec. 7.2)

The protocol on access to the DATRAS database that was presented in last year's report was discussed and some minor alterations were suggested. IBTS agreed that access to various forms of IBTS data should be improved. Though IBTS were not able to agree a change to the current access policy at the meeting, and so that there would be no immediate change, they discussed potential policies and hope to propose a new access policy for 2006. For further information on access policy (e.g. forms for requesting data and "health warnings" for the data), see ICES (2004).

To structure data access, three IBTS survey/area combinations were distinguished, the countries participating in these combinations and whether data were submitted to the database:

Country	North Sea	Western[(1)]	Southern	Data in database
Denmark	X			X
England/Wales	X	X		X
France	X	X	X	X
Germany	X			X
Ireland		X		
Netherlands	X			X
Norway	X			X
Portugal			X	
Scotland	X	X		X
Spain		X	X	
Sweden	X			X

Note: [1] This does not include the porcupine survey due to the Spanish data regulation.

Within each of these survey/area combinations there was agreement on the data access policy. This is shown in the table below, which distinguishes four user categories:

- (1) Institutes that have supplied data to the database.
- (2) ICES Assessment Working Groups
- (3) Other ICES Working Groups
- (4) Public and other parties that request data, typically for research purposes.

and the following three data types:

(1) Standard maps and graphs: Per survey/area combination for all relevant ages of species for which assessments are conducted. Maps will show bubble plots indicating abundance per ICES rectangle or per haul. Time series of the indices and a graph showing the proportion of the age-groups will be generated. These will be available for 8 commercial species (cod, haddock, whiting, herring, sprat, mackerel, saithe and Norway pout).

(2) Aggregated data: A query of the database using pivot tables. Based on these tables, plots and graphs can be made on an interactive basis. The minimum level of aggregation differs between survey/area combinations.

- ICES rectangle: IBTS in the North Sea, Skagerrak, Kattegat and the BTS in the North Sea, Channel and Irish Sea
- Stratum: IBTS western and southern divisions

- Sub-division and stratum: BITS Baltic Sea

(3) Un-aggregated (raw) data. These are catch (numbers at length and/or numbers at age) data on a haul-by-haul basis and SMALK (Sex, Maturity, Age-Length-Keys) data per individual.

Data access for the four "User categories" and "Data types" can be organized according to the following matrix, and includes "Free access", "password protected access" and "access to extracted data after granted request".

Type of data	User Category			
	National Fisheries Institutes (data suppliers)	ICES Assessment WGs	Other ICES WGs	Public and other parties
Standard maps and graphs (9 commercial species)	Free access	Free access	Free access	Free access
Aggregated data (by ICES rectangle) [4,5]	Free access	Free access	Free access	Free access
Aggregated data (by ICES rectangle) for other species and times	Password protected [1]	Password protected [3]	Receive data after request to ICES	If request granted by ICES
Raw data	Password protected [1]	Password protected [3]	Receive data after request to ICES [2,3]	After request granted by national contact person [2]

Notes:
[1] For those survey/area combinations that the laboratory contributes to, otherwise by request
[2] If access to raw data is given, ICES (IBTSWG) and the national fisheries laboratories supplying the data retain "intellectual property", hence if data are to be used for publications, authors must liase with IBTS members to ensure that both data analysis and interpretation are appropriate
[3] ICES Assessment WGs that use IBTS data should provide IBTSWG with feedback regarding the utility of the survey data for the species/stock in question, so that IBTS know which data are performing well
[4] Data would be available for the following dominant species: gadoids (*Gadus morhua, Melanogrammus aeglefinus, Merlangius merlangus, Trisopterus esmarki, Trisopterus minutus, Merluccius merluccius, Molva molva*, *Pollachius virens*), flatfish (*Limanda limanda, Hippoglossoides platessoides, Microstomus kitt, Pleuronectes platessa, Glyptocephalus cynoglossus, Lepidorhombus whiffiagonis*), other demersal species (*Eutrigla gurnardus, Lophius piscatorius, Lophius budegassa, Echiichthys vipera, Amblyraja radiata, Raja clavata, Scyliorhinus canicula*) and certain pelagic fishes (*Clupea harengus, Scomber scombrus, Trachurus trachurus, Sprattus sprattus*).
[5] Data will only be freely available from 1983, and excluding more historical data and the three most recent years data.

B Planning Group for Herring Surveys (PGHERS) – report 2005: Status and future of the HERSUR database (Sec. 7)

(…) In the light of emerging discussions within the ICES community on **data policies and data access**, the group supports the view that open access to data is important for the scientific community. Open access facilitates a wider use, which adds value to data collected by public funding, and increases the quality of the data if a scheme for reporting errors is properly implemented. PGHERS decided that all aggregated data from the acoustic survey should be available to the public without restrictions after a certain period needed for checking. The group considered that data is final after the ACFM spring meeting, so the public should have access 10 months after the survey ended. Users should be encouraged to use the data according to good scientific practice, i.e acknowledge the source of the data, report errors to the source, and to contact the source for additional information to avoid misinterpretations. Raw data stored in the HERSUR database should be kept password protected for the near future.

Annex 3: Elements of end-to-end data management

The concept of a end-to-end data management is fundamental to the whole process of making data more accessible to a wide range of users. Significant resources are invested annually by the ICES community in the planning, acquisition, analysis, and submission of marine data. Resultant datasets are often unique, temporally and spatially, and irreplaceable. For this reason, it is of the utmost importance that best practices of data storage, maintenance, and documentation are implemented to safeguard data for the future and to derive maximum value over the long-term. Such practices can conveniently be combined into a end-to-end data management, and an outline of one possible cycle is given below:

Prior to data submission (and carried out by the data provider):

- Planning – including the purpose, scope, longevity and planned usage. Account would also be taken of relevant legislation, and appropriate risk analysis.
- Capture – definitions of procedures for data capture, including statistical considerations, valid sources, parameters to be measured and accuracy requirements.
- Interim storage – details of procedure for secure storage prior to entering or transmitting to 'permanent' repository.
- Coding – procedure for any data conversion or coding method.
- Transmission – description of security requirements, methods of transmitting data to database or final storage media.
- Cleaning – detailed procedures for error-trapping, consistency checking and validation.
- Metadata description – to be provided as essential additional information as part of the data capture process or, at the latest, prior to release of the data to users. Contains details of the type of data, precision levels, restrictions on its use, and availability.
- Distribution – defines how the data are made available to a wider audience.
- Management, eventual destruction or archiving – based on the original objectives and plans; describes how long the data need to be retained, and the plans for either long-term storage, archival to different media, or destruction of the data (e.g. provisional data in the latter case).

Integration (carried out by data collator e.g. ICES Data Centre):

- Receive – temporarily store received data in secure area.
- Screen for loading factors – including storage requirements, compatibility issues.
- Reformat – data may not be in the final format required for collation/storage, so a reformatting procedure may be required.
- Validate for quality factors – data should comply with agreed quality standards, or should have an associated quality flag so that data users may make decisions on the utility of the data for their own purposes.
- Integrate – add the new data to the database or data collection.

Post-Integration (carried out by data collator e.g. ICES Data Centre):

- Maintain - set procedures in place to promote secure storage and backup of the data
- Update for compatibility with infrastructure incl. operating systems, etc.
- Coordinate quality issues – including dealing appropriately with reported errors, disseminating information to users.
- Ensure availability to users - whether via electronic media such as the web or otherwise. This would probably include the provision of a robust internet service in addition to an on-site library service for non-electronic data sets.

- Ensure confidentiality - as appropriate as determined by any agreed protocols or legislation

Longer-Term (carried out by data collator e.g. ICES Data Centre)

- Release of confidential data after some period – some data may be restricted for a specific period of time, either due to confidentiality, legal, national security or intellectual property issues.
- Data extinction/format migration – the data holder should be responsible for ensuring that data sets are maintained in a usable format; i.e. are not allowed to remain in a format which has become obsolete or unreadable.
- Data destruction or archiving – as with all data collections, procedures should be in place to ensure that data are retained only for as long as necessary. Data no longer required should be considered for disposal or longer-term storage on alternate media if required.

The end-to-end data management outlined above applies regardless of the use of distributed or centralized databases, but data protocols and responsibility issues may be different between the two options. Detailed Codes of Practice (or Standard Operating Procedures) need to be written and used for any such system to be successful.

Annex 4: Draft Policy for ICES Data Management, Distribution and Archiving

A Introduction - Philosophy and Motivation

ICES' purpose is to coordinate and promote marine research in the North Atlantic. It has an international reputation for the provision of scientific advice on the management of fisheries, and the status of fish stocks, environmental quality, and the environment.

ICES has a long history in marine data management. The international community of scientists working within ICES gathers and submits a diverse range of information on the marine ecosystem, with traditional strengths in fisheries, oceanography and the marine environment. As a result some of the most valuable data on these subjects reside at the ICES Data Centre.

The main purpose of these data is to support ICES working groups in carrying out assessments, such as those on fish stocks or environmental status. The management of such large and diverse data within a single data centre has the advantage of providing the ICES community of scientists and managers, and the public at large with timely, open, and platform independent access to marine data covering a large area.

Significant resources are invested annually by the ICES community in the planning, acquisition, analysis, and submission of marine data. Resultant datasets are often unique, temporally and spatially, and irreplaceable. For this reason, it is of utmost importance that best practices of data storage, maintenance, and documentation are implemented to safeguard data for the future and to derive maximum value over the long-term.

By maximizing the availability of all data holdings available to the community at-large, ICES will actively promote the use of these data, thereby ensuring maximum value is derived from them and contributing to an increased understanding of the marine environment.

The purpose of this Data Policy is to set out a framework for the access and use of data held within the ICES Data Centre. It should be read by both data contributors and users.

ICES will promote the adoption and implementation of this new, open data policy by all of its groups, and will support expert groups and the member states' national laboratories in making their data publicly available.

B Intergovernmental Requirements

ICES data policy must be consistent with and in the spirit of national and international policies and laws. The policies and laws may apply to the ICES Secretariat, member states, and/or to the people or organizations that either provide or use data and information managed by ICES. Applicable policies or laws are those related to UN conventions, policies of international bodies often within the UN, policies and laws of the European Union as well as of ICES member states.

Within the ICES member states and the European Union, new legislative mandates to increase public access to information in general and to environmental information specifically are being formulated and, in some cases, enacted. Among these are the "Aarhus Convention on Access to information, public participation in decision-making and access to justice in environmental matters", and the EU Directive 2003/4/EC on "public access to environmental information".

In the US, the Freedom of Information Act provided that any person has the right to request access to federal agency records or information if requested, except for those that are protected by exemptions or exclusions.

In addition to legislative mandates, many new policies are being enacted by various governmental, intra-governmental and other responsible groups. Additionally, the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities", which has been adopted by more then 50 scientific institutions and organisations form all over the world, sets the goal to disseminate knowledge by establishing open access contributions on the web. Open access contributions include original scientific research results, raw data, and metadata.

In Appendix 1 some of the relevant laws and policies are more fully referenced.

C Objectives of the ICES Data Policy

The objectives of this policy are to lay forth the guidelines for data submission/acquisition, distribution to users, and the structure of the data management system itself. Implicit in this policy is that ICES will remain a focal point for marine data in the North Atlantic and that ICES will create a portal serving as a hub for distributed data, which will be attractive to the science community and useful for ecosystem management. The updated ICES data policy will provide the basis to serve the scientific community as well as the needs for the production of the advice and status reports.

D Policy Elements

The ICES Data Centre will make reliable data available in an open, timely, and easy fashion to the marine research community. Users must be aware that ICES Data Centre only provides accessto the data; the data centre remains totally dependent on data contributions ICES urges data sources to strengthen their commitment to free and unrestricted use of data.

D.1 ICES Data Sources, Quality, and Requirements

Data and information are provided to ICES from a number of sources with variable data quality and obtained with various methodologies (Appendix 2). Data sources may be the originators of the data/information (for example, the persons responsible for the scientific work that produce the data/information) or some intermediary (such as the institute(s) to which the scientist(s) belong, the agency that commissioned or funded the work, or even an information technology group that is responsible for preparing the data for its submission to ICES). The data source is assumed to have the right to grant or deny access to the data. Data submitted to ICES is considered to be in the public domain unless otherwise explicitly specified. Some cases that call for restrictions include data submitted during a prescribed period of exclusive use, or legally or economically sensitive data (examples are given in Appendix 3). Restricted access is particularly important for the fisheries data. The period for data sets from an exclusive scientific origin is normally not more than two years. ICES remains the repository and/or the distributor of submitted data.

At the time of data submission, documentation about the data origins should be included if possible, including the contact person for questions that arise by users and cannot be addressed by the metadata. Documentation of measurement and analysis techniques used to produce the data set (the metadata) must be submitted with the data to the ICES Data Centre. This should include information about the accuracy and precision of each measurement. All data (oceanographic, environmental, and fisheries) that reside in the ICES Data Centre repositories should have a common set of quality flags. ICES guarantees neither the quality nor the usage of the data. Data sources retain the primary responsibility for data quality control and assurance.

D.2 Data Exchange with other bodies and organisations

Data sharing and exchange of data are an essential component of ICES activities. ICES therefore will become an active member of relevant networks for data contributors and users.

D.3 Archival / Data Rescue

ICES recognises its responsibilities for long-term maintenance and access to data relevant to its mission. Data are envisioned to have a life cycle that includes initial data collection planning, acquisition, and processing, the submission of data to the ICES Data Centre and their installation into the integrated database, the maintenance of the data while being actively used, and the final archiving of the data.

Data can be unintentionally destroyed or lost. ICES strives to rescue and archive valuable data sets that are in danger of being lost, including those that reside in reports and documents. In some cases, the ICES Data Centre may serve as the archive for such data. The data source, however, is responsible for migrating the data to the ICES database with sufficient documentation.

D.4 Access to data

All data held by the ICES Data Centre should eventually become publicly available, with due regard to relevant legislation. However, access to sensitive data will be restricted or data may be aggregated for a limited period of time if specifically stipulated by the data contributor. The ICES web site will be a key focus point in disseminating information to the ICES community.

D.5 Responsibilities

Data Centre Responsibilities

ICES Data Centre is responsible for implementing the data policy. This encompasses simultaneously supporting ICES data users in accordance with the data policy and the secretariat's work plan while protecting the rights of data sources. While the ICES Data Centre and expert groups may perform some data quality control, the responsibility for overall data quality ultimately resides with the data source. The ICES Data Centre is not responsible for how the data are used. The Data Centre will provide a user friendly method for reporting errors in the database by external users. It will forward the error reports to the data source for correction. ICES will promote acknowledgement of data sources. To support this, ICES will provide as much data source information as possible.

Data Users/Contributors Responsibilities

Data contribution: Speed is becoming a primary quality factor, thus data should be made accessible as soon as possible and to the broadest user group possible. This implies both technical and policy considerations and coordination on the part of data sources, users, and the ICES Data Centre. In order to maximize the usability of data and thereby their value, data contributors must supply meta-data and, if available, data quality indicators. All data including meta-data and quality indicators should be submitted using standard coding formats, and protocols to the extent possible.

Data Use: Correct and appropriate data interpretation is solely the responsibility of data users. Users must acknowledge data sources. It is not ethical to publish data without proper attribution or co-authorship. Any person making substantial use of data must communicate with the data source prior to publication and anticipate that the data originator(s) should

possibly be considered for co-authorship of published results. Use or reproduction of any material herein for any commercial purpose is prohibited without prior written permission from ICES and/or the data source. Users must not imply ICES' substantiation of their work, results, conclusions and/or recommendations.

Data users are obliged to inform the Data Centre of any suspected problems in the data and encouraged to inform of possible sources of other relevant data.

E The ICES Data Management System

Data either reside physically at ICES in centralised systems or are dispersed in various databases elsewhere (distributed system). ICES has chosen to continue support of the following centralized integrated databases:

- Databases with a long history of residing at ICES that require little maintenance
- Aggregated data used as the basis for ICES advice
- Sensitive data.
- Valuable data at risk of extinction.

All other data are or will eventually be stored in distributed systems. ICES promotes development of distributed databases and web interfaces for them.

F Modification of the Policy

The ICES Council ensures that the data policy continuously adapts to evolving needs of working groups and advisory committees, and in response to changing regulatory environments, as well as to technological advances. This is effected a working group whose terms of reference include keeping abreast of perceived need for change and recommending modifications of the policy.

ICES Data Policy Appendix 1: Relevant Documents for a new ICES Data Policy

A closer look at the Aarhus Convention:

Many of the new legislative mandates in Europe have been set up to implement the provisions set out in the CONVENTION ON ACCESS TO INFORMATION, PUBLIC PARTICIPATION IN DECISION-MAKING AND ACCESS TO JUSTICE IN ENVIRONMENTAL MATTERS (the so-called Aarhus-Convention, Denmark, 25 June 1998). The Aarhus Convention entered into force in 2001, and up to early 2005 has 40 signatories and 35 parties, including the European Union. To implement the Aarhus Convention, the European Parliament and Council has adopted the Directive 2003/4/EC on "public access to environmental information" that has to be complied with by laws, regulations and administrative provisions of EU member states by 14 February 2005.

According to the Aarhus Convention,

> "3. 'Environmental information' means any information in written, visual, aural, electronic or any other material form on:
>
> (a) The state of elements of the environment, such as air and atmosphere, water, (…) natural sites, biological diversity and its components, including genetically modified organisms, and the interaction among these elements;
>
> (b) Factors, such as substances, energy, noise and radiation, and activities or measures, including administrative measures, environmental agreements, policies, legislation, plans and programmes, affecting or likely to affect the elements of the environment within the scope of subparagraph (a) above, and cost-benefit and other economic analyses and assumptions used in environmental decision-making; (…)
>
> 4. "The public" means one or more natural or legal persons, and, in accordance with national legislation or practice, their associations, organizations or groups;
>
> 5. "The public concerned" means the public affected or likely to be affected by, or having an interest in, the environmental decision-making; for the purposes of this definition, non-governmental organizations promoting environmental protection and meeting any requirements under national law shall be deemed to have an interest."

Under the Aarhus Conventions (Article 4), "each party shall ensure that public authorities (…) in response to a request for environmental information, make such information available to the public, within the framework of national legislation, (…) copies of the actual documentation containing or comprising such information

> (a) Without an interest having to be stated;
>
> (b) In the form requested unless:
>
> (i) It is reasonable for the public authority to make it available in another form, in which case reasons shall be given for making it available in that form; or
>
> (ii) The information is already publicly available in another form. "

This information "shall be made available as soon as possible and at the latest within one month after the request has been submitted, unless the volume and the complexity of the information justify an extension of this period up to two months after the request. The applicant shall be informed of any extension and of the reasons justifying it."

However, under certain circumstances, the Aarhus Convention allows to refuse a request for environmental information, if for example

> "(a) The public authority to which the request is addressed does not hold the environmental information requested;
>
> (b) The request is manifestly unreasonable or formulated in too general a manner; or
>
> (c) The request concerns material in the course of completion or concerns internal communications of public authorities where such an exemption is provided for in national law or customary practice, taking into account the public interest served by disclosure.
>
> A request for environmental information may be refused if the disclosure would adversely affect: (…)
>
> (c) The course of justice, the ability of a person to receive a fair trial or the ability of a public authority to conduct an enquiry of a criminal or disciplinary nature;
>
> (d) The confidentiality of commercial and industrial information, where such confidentiality is protected by law in order to protect a legitimate economic interest. Within this framework, information on emissions which is relevant for the protection of the environment shall be disclosed;
>
> (e) Intellectual property rights; (…)
>
> (g) The interests of a third party which has supplied the information requested without that party being under or capable of being put under a legal obligation to do so, and where that party does not consent to the release of the material; (…). "

However, "the aforementioned grounds for refusal shall be interpreted in a restrictive way, taking into account the public interest served by disclosure and taking into account whether the information requested relates to emissions into the environment." It is also required that a refusal shall state the reasons for the refusal and give information on access to the review procedure provided for.

The Aarhus Convention includes provisions that parties may allow its public authorities to make a charge for supplying information, but such charge shall not exceed a reasonable amount.

In Article 6, dealing with collection and dissemination of environmental information, it is stated that "each Party shall ensure that environmental information progressively becomes available in electronic databases which are easily accessible to the public through public telecommunications networks. Information accessible in this form should include:

> (a) Reports on the state of the environment, as referred to in paragraph 4 below; (b) Texts of legislation on or relating to the environment;
>
> (c) As appropriate, policies, plans and programmes on or relating to the environment, and environmental agreements; and
>
> (d) Other information, to the extent that the availability of such information in this form would facilitate the application of national law implementing this Convention, provided that such information is already available in electronic form.

The CLIVAR Data Policy

February 2005; CLIVAR is an International Research Programme on **Cli**mate **Var**iability and Predictability; http://www.clivar.org

Introduction

CLIVAR, a global multidisciplinary climate research project, requires a wide range of data and needs data centres to ingest, quality control, archive and distribute these data. The CLIVAR data policy provides guidelines for how these data should be handled in a consistent manner so as to achieve the project's scientific objectives. The policy aims to strike a balance between the rights of investigators and the need for widespread access through the free and unrestricted sharing and exchange of CLIVAR data and metadata. CLIVAR data policy is intended to be fully compatible with IOC [1], WMO [2], GCOS [3] and GEOSS [4] data principles.

The multidisciplinary nature of CLIVAR and its subprogrammes means that the principles enshrined in the Data Policy must be applied to data in each subprogramme's implementation plan.

Definitions used in the CLIVAR Data Policy

1. CLIVAR data

"CLIVAR Data" consists of directly observed data, derived data, gridded fields, and other data products generated and/or used within CLIVAR to further its scientific goals. CLIVAR data and related products can be categorized in terms of those that are specifically sponsored or endorsed by the international CLIVAR programme, those generated by other related bodies and programmes such as the World Weather Watch of the WMO, GCOS, JCOMM, and other projects of the WCRP and those generated by relevant national and institutional CLIVAR-related projects and programmes. CLIVAR should strive to ensure that all data relevant to CLIVAR are accessible freely and without restriction, including those collected by other projects and programmes.

2. Metadata

Metadata is defined as the descriptive information such as content, quality, condition that characterises a set of measurements.

CLIVAR Data Policy and Principles

For CLIVAR to succeed, high-quality data and metadata need to be collected, processed and exchanged without significant delay in a free and unrestricted manner. This was recognized in the CLIVAR Implementation Plan [5] in discussing 'The Principles for CLIVAR Data'. CLIVAR data policy is enshrined in the CLIVAR data principles below:

1. Free and unrestricted exchange

All CLIVAR data should be made available freely and without restriction. "Freely" means at no more than the cost of reproduction and delivery, without charge for the data itself. "Without restriction" means without discrimination against, for example, individuals, research groups, or nationality. In exceptional circumstances involving highly specialized or experimental data, principal investigators may temporarily limit access until such time as the data can be adequately validated.

2. Timely exchange

CLIVAR investigators should make data available voluntarily and with minimal delay, preferably also in real-time, to maximize their value to CLIVAR. In cases where extensive

post-processing of delayed mode data is needed before a final research quality data set can be generated, early release of a preliminary version of the data is required.

3. Quality control

CLIVAR investigators retain the primary responsibility for the quality of the data they produce and distribute. Data originators and those generating climate data products are required to ensure that their data meet international quality standards wherever possible.

4. Metadata

Metadata are required to enable the use of data without ambiguity or uncertainty. Metadata for CLIVAR data sets will be developed and managed in accordance with international standards.

5. Preservation of data

Long-term survival, integrity, and access to CLIVAR data must be preserved for future generations. Internationally agreed standards should be used for the acquisition, processing, and final archival of data and metadata. Data distributed in real and near-real time should, wherever possible, be replaced in a delayed mode after it has undergone quality control and full documentation.

6. Plan for reuse in reanalysis

While datasets will be used individually and in combination for research purposes, the sum total of all CLIVAR and CLIVAR-relevant data will have great value in reanalysis activities. To aid this, uniformity of data format and quality should be a high priority.

7. Easy access

CLIVAR encourages the use of the most recent advances in communication to ensure widespread access to data collected under auspices of the programme.

8. Use of existing national and international mechanisms and centres

Where feasible, CLIVAR will use existing national and international mechanisms for the exchange and storage of oceanic and atmospheric data, and build on the data management structure of existing programmes. In this way, the effectiveness of the data system will be improved by reducing redundancy and duplication and identifying opportunities and system economies, with financial costs minimized.

9. Reporting Requirements

Data and metadata should be submitted to recognized data assembly centers as well as to appropriate national and international archival institutions so that the collected information may be safeguarded for future analysis. Inventories of data and related information should be readily accessible and updated as needed on a routine basis.

References

[1] IOC Data Policy (http://ioc3.unesco.org/iode/contents.php?id=200)

[2] WMO Resolution 40 (Cg-XII; see http://www.nws.noaa.gov/im/wmor40.htm)

[3] Implementation plan for the Global Observing System for Climate in support of the UNFCCC, 2004; GCOS – 92, WMO/TD No.1219.

[4] Global Earth Observation System of Systems GEOSS. 10-Year Implementation Plan Reference Document (Final Draft) 2005. GEO 204. February 2005.

[5] CLIVAR Initial Implementation Plan, 1998; WCRP No. 103, WMO/TS No. 869, ICPO No. 14. June 1998.

International Council for Science - World Data Center System

A. Principles and Responsibilities of ICSU World Data Centers

The basic principles and responsibilities of the international exchange of solar, geophysical and environmental data through the World Data Centers have carried forward under ICSU rules, essentially unchanged since the establishment of the WDC system for the IGY. The following text replaces the sections on "Principles and Responsibilities of the World Data Centers" in Part I of the Guide to the World Data Center System, dated November 1987.

1. World Data Centers are operated for the benefit of the international scientific community. WDCs in the United States are designated as WDCA, in Russia as WDC-B, in other European countries as WDC-C or WDC-C1, in Japan or India as WDC-C2, and in China as WDC-D. They are supported by national organizations according to these Principles laid down by the ICSU Panel on World Data Centers.

2. The resources required to operate WDCs are the responsibility of the host country or institution, which is expected to provide these resources on a long-term basis. If for any reason a WDC is closed, the data holdings shall be transferred to another WDC.

3. WDCs will, subject to their financial resources, accept data according to the data management plans of appropriate ICSU scientific programs or monitoring activities, and store these data safely and in good condition. WDCs may enhance their holdings by seeking and collecting related data sets. They may prepare higher-order data products such as indices of activity and collated or condensed data sets.

4. WDCs will prepare and publish catalogs of their data holdings, or otherwise make freely available information on their holdings, e.g., by electronic access.

5. WDCs will exchange data among themselves, as mutually agreed and when-ever possible without charge, to facilitate data availability, to provide back-up copies, and to aid the preparation of higher order data products.

6. No confidential or security-classified data are to be held in a WDC.

7. Data may be subject to privileged use by their originators, for a period to be agreed beforehand, and not to exceed two years from the date of acquisition by the WDC.

8. WDCs will provide data to scientists in any country free of charge, on an exchange basis or at a cost not to exceed the cost of copying and sending the requested data. Additional charges may be made for special services, or for acquiring data from outside the WDC system.

9. WDCs will accept any scientist as a visitor to work on site with data holdings held under WDC auspices.

10. WDCs will report to the ICSU Panel as requested.

Ocean Observing System Data Policy

International efforts within the UN to develop ocean observing systems have lead to many policy statements. For example the International GOOS Design Principle 7 .states:

Principle D7. The management, processing and distribution of data will follow a specified data policy.

In concert with the policies of IODE, IGOSS and GCOS, and following the data management plan for the World Weather Watch of the WMO, commitment is required by GOOS

participants to establishing, maintaining, validating, making accessible, and distributing high quality, long term data meeting internationally agreed standards. Preservation of GOOS data is required in suitable archives following appropriate procedures and criteria for data acquisition and retention, and should include information about data holdings. Data should be processed to a level which is generally suitable for the generation of operational products and for research, and described in internationally accessible on-line computerised directories that can also be made available by other means. GOOS contributors are responsible for full, open and timely sharing and exchange of GOOS-relevant data and products for non-commercial activities. Exchange implies that donation by individual nations gains access to data from others as well as to products derived using all available data, such that the benefit of cooperation exceeds the cost.

EuroGOOS

The EuroGOOS initiative has recently started to develop its data polices and no doubt must adhere to the same policies and laws that ICES will. They specifically mentioned the following:

Policy and practice for EuroGOOS for the exchange of oceanographic and related data and products including guidelines on relationships in commercial oceanographic activities

NOTING:

(1) WMO Resolution 40 (Cg-XII) – WMO policy and practice for the exchange of meteorological and related data and products including guidelines on relationships in commercial meteorological activities,

(2) The Decision of the Commission for the European Union dated 21.10.1999 regarding ECOMET Economic Interest Grouping according to Belgium Law,

(3) The UN convention of 1998 on the access to information, public participation in decisionmaking and access to justice in environmental matters (the Aarhus-Convention)

(4) Council Directive 90/313/EEC of June 1990 on the freedom of access to information on the environment,

(5) Directive 96/9/EC of The European Parliament and of the Council of 11 March 1996 on the legal protection of databases,

RECALLING:

(1) The present GOOS Policy for exchange of data and products

CONSIDERING:

(1) The continuing fundamental importance, for the provision of oceanographic services in all countries, of the exchange of oceanographic and related data and products between EuroGOOS Members,

(2) Other programmes of world importance and the ocean observing and Marine Meteorological programmes under JCOMM,

(3) The basic role of EuroGOOS Members to furthering applications of oceanography to all human activities,

(4) The call by world leaders at UNCED (Brazil, 1992) for increasing global commitment to exchange scientific data and analysis and for promoting access to strengthening systematic observations,

(5) The provisions in the UN/FCCC committing all Parties to the Convention to promote and cooperate in the full, open and prompt exchange of information related to the climate system and climate change,

(6) UN Convention on the Law of the Sea, especially those articles on international data exchange: 200, 244.2, 249.1(c), 249.1(d) and 277(e).

RECOGNISING:

(1) The increasing requirement for the global exchange of all types of environmental data and in particular ocean data,

(2) The basic responsibility for the Members to provide universal services in support of safety, security and economic benefits for the peoples in their countries,

(3) The dependence of Members on the stable, co-operative international exchange of data and products for the discharge of their responsibilities,

(4) The continuing requirement for Governments to provide for the oceanographic infrastructure of their countries,

(5) The continuing need for, and benefits from, strengthening the capabilities of Members, in particular in developing countries, to improve the provision of services,

(6) The dependence of the research and education communities on access to oceanographic and related data and products,

(7) The right of Governments to choose the manner by, and the extent to, which they make data and products available domestically or for international exchange.

RECOGNISING FURTHER:

(1) The existence of a trend towards the commercialisation of many oceanographic and related services and products,

(2) The requirement by some Governments that the Members initiate or increase their commercial activities,

(3) The risk arising from the commercialisation to the established system of free and unrestricted exchange of data and products, which forms the basis for the European co-operation in oceanography,

(4) The EU-principle that all data and products, financed with public means and used by public entities for commercial activities shall also be available to other Service Providers.

ADOPTS the basic principles for the exchange of oceanographic and related data and products:

(1) Exchange on a free and unrestricted basis of essential, additional and other data and products between the Members of EuroGOOS,

(2) The right for the originator of data and products to place conditions on additional and other data and product for re-distribution for commercial purposes,

(3) Free and unrestricted access to data and products for non-commercial research and education,

(4) All data and products that is financed with public means and used for commercial purposes must be available for other Service Providers,

(5) Transparency regarding availability, prices and conditions for re-distribution regarding oceanographic and related data and products through the maintenance of a EuroGOOS Product Catalogue. These principles are in harmony with WMO Res. 40 and the EU Commissions decision regarding ECOMET.

FURTHER ADOPTS the following practice on the international exchange of oceanographic and related data and products:

(1) Members shall provide on a free and unrestricted basis essential data and products which are necessary for the provision of services in support of the protection of life and property and the well-being of all nations, particularly those data and products, as, at a minimum described in Annex 1 to this document, required to support WMO or GOOS Programmes;

(2) Members should also provide additional data and products which are required to sustain programmes at the global, regional and national levels and, further as agreed, to assist other Members in the provision of oceanographic services in their countries. While increasing the volume of data and products available to all Members by providing the additional data and products, it is understood that EuroGOOS Members may be justified in placing conditions on their re-delivery for commercial purposes;

(3) Members shall also provide to anybody all other data and products which are used in commercial activities and which have been funded through core/infrastructure activity or government grants. It is hereby understood that Members may be justified in placing conditions on their re-distribution for commercial purposes as well as to charge for the information and the delivery;

(4) Members should provide to the research and education communities, for their non-commercial activities, free and unrestricted access to all data and products exchanged under the auspices of this document with the understanding that their commercial activities are subject to the same conditions identified in FURTHER ADOPTS (2) and (3) above;

URGES Members to:

(1) Strengthen their commitment to the free and unrestricted exchange of oceanographic and related data and products;

(2) Increase the volume of data and products exchanged under the auspices of this document;

(3) Assist other Members, to the extent possible, and as agreed, by providing additional data and products in support of time-sensitive operations regarding oceanographic and environmental warnings, rescue operations and safety of life at sea;

(4) Strengthen their commitment in their collection and supply of oceanographic and related data and products;

(5) Implement the practice on the international exchange of oceanographic and related data end products, as described in ADOPTS (1) to (4) above;

(6) Make known to all Members those oceanographic and related data and products which have conditions related to their re-distribution for commercial purposes;

(7) Make their best efforts to ensure that the conditions which have been applied by the originator of additional and other data and products are complied with and made known to initial and subsequent recipients;

DECIDES to review the implementation of this document at the EuroGOOS Annual Meeting.

United States

The US has many relevant laws and policies. Within the GOOS context the effort is lead by Ocean.US. Data policy developed by this group will apply to all ocean data collected for GOOS. The data policy are now being developed.

Food and Agricultural Organization

The FAO published a document on the legal aspects of collection of fisheries data (Edeson, 1999). The purpose of the paper was to outline the provisions of the principal international fisheries instruments concerning the collection and exchange of fisheries data. They noted the relevant instruments are:

- UN Convention on the Law of the Sea, 1982
- Agreement for the Implementation of the Provisions of the United Nations Convention on the Law of the Sea of 10 December 1982 relating to the Conservation and Management of Straddling Fish Stocks and Highly Migratory Fish Stocks
- FAO Agreement to Promote Compliance with International Conservation and Management Measures by Fishing Vessels on the High Seas, and
- FAO Code of Conduct for Responsible Fisheries.

References

Edeson, W.R. (1999) Legal aspects of the collection of fisheries data. *FAO Fisheries Circular.* No. 953. Rome, FAO. 18p.

The Berlin Declaration – Open Access to Knowledge in the Sciences and Humanities" http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html

ICES Data Policy Appendix 2: Data Types

Three types of data are distinguished. Detail data are the results of individual sample measurements or observations and related attribute data that provide information about where and when a sample was collected, what it is, etc. Aggregate data are those available in a summarized form; they frequently lack information about the procedures used to do the aggregation. Metadata provide information about the data without providing actual results. These include units, accuracy, precision, method of measurement or sampling, investigator/originator, reference to publications describing the data set, a description of the processing of data, etc.

For additional information about metadata and metadata standards the following sources provide an introduction. At a recent meeting, a talk was given by Lillian W. Gassie on " Metadata Tools, Practices, and Ontology". A PDF of her talk can be obtained by going to: http://www.mbari.org/iag/workshops/dmv/Presentations.html and paging down to the Section II link called "II. Metadata Tools, Practices, and Ontology", which takes you to the following URL: http://www.mbari.org/iag/workshops/dmv/Presentations.html#metadatatools. Clicking on this brings up the PDF file. Slide 8 provides links to sites where Metadata Standards are provided i..e. http://www.fgdc.gov/metadata and http://www.csc.noaa.gov/metadata/text/metstandard.htm. Other links are also provided for more information about the standards.

ICES Data Policy Appendix 3: Restrictions on release of fisheries based data from the ICES Data Centre

Scientific Survey Data

TYPE	TYPICAL LIFE / RESTRICTION
Surveys - coordinated – Raw Length Distribution, Haul, Acoustics	Restricted 2yrs
Surveys - coordinated – Aggregated / Indices	Public after final quality control <1yr
Surveys – Environmental Data	Public after final quality control <1yr
Biological Parameters – Aggregated	Public after final quality control <1yr
Biological Parameters – Raw	Restricted 5 yrs

Commercial Fisheries Data

TYPE	TYPICAL LIFE / RESTRICTION
Discards – Aggregated (level to be defined)	Public
Discards – Raw	Restriction stipulated by contributor and dependent data subject (fisher).
Landings – Reported – Aggregated	Public
Landings – Reported – Raw	Restriction stipulated by contributor and dependent data subject (fisher).
Landings – In Confidence	Restriction stipulated by contributor and dependent data subject (fisher).
Effort – Reported – Aggregated	Public
Effort – Reported – Raw	Restriction stipulated by contributor and dependent data subject (fisher).
Effort – In confidence	Restriction stipulated by contributor and dependent data subject (fisher).

Processed and Consolidated Data

TYPE	TYPICAL LIFE / RESTRICTION
Working Group Report – In prep	Restricted to members of Working Group
Working Group Report – Final – Not Approved by ACFM	Public after maximum 3 months
Working Group Report – Final – Approved by ACFM	Public after maximum 6 months

Note: As disaggregated (raw) fisheries data can be linked to a person or persons, or vessel, or enterprise, its use and dissemination will be governed by the various data protection acts in force in member countries and that of the European Union.

Scientific Data Access (Examples)

ORGANISATION	TYPICAL LIFE/RESTRICTION*
GLOBEC	6 months with allowance for manual analysis
US/IOOS	Realtime or as soon as possible. Must meet metadata and FGDC standards
GOOS	Realtime and ????
EuroGOOS	Realtime and ????
MARGINS	Basic metadata within 60 days. Data as soon as practical.
Ocean Drilling Program	Submission plan and submission required.
JGOFS	1 year with provision for long analytical procedures
CLIVAR	2 years but some data required earlier
RIDGE	Rapid metadata and data dissemination.

*Realtime data is served to national and international servers with minimal lag.