Proposal for an update to the ICES Data Policy

This document contains the proposed new Data Policy.

The current (2012) Data Policy was evaluated by DIG 2016 and updated, in summary these were:

- Exceptions on data permissions
- Incorporating the eight principles on Data Citation (paragraph 7)

As external developments (eg. EU INSPIRE Directive, adoption of CC by institutes) might influence the way institutes or MS deal with data, it is important to re-investigate the ICES Data Policy if necessary, even before the official update is scheduled.

Council is invited to approve this update to the Data Policy.

1   Scope

ICES will be a leader in marine data and information management, providing best practices, data mobilisation and services for its advisory and science groups and the wider marine and maritime communities

This policy states the conditions for data submission, access and use in order to facilitate the production of science based advice and status reports, and serve the scientific community.

This policy applies to data managed by ICES, and to ICES activities for providing access to data managed elsewhere.

Regarding public access to these data; this policy may not apply to underlying data where the data provider has exerted their right to restrict public access. All data products are by default publicly available, including those derived from restricted data. The 'Annex' available on this link denotes the specific dataset collections where this applies.

2   Definitions

Data Products: data outputs resulting from aggregation of or calculated from underlying data

Underlying data: data delivered by the Data provider

Data user: entity (eg. person, organization, group, including ICES expert groups) using data

Data provider: entity providing data to ICES

Data source: ICES, or ICES and data provider(s)

Publicly available: online open access

Meta-data: descriptive information about detail or aggregate data sets, necessary to interpret, use and disclose data

3   Use of Data

a.   Data users can obtain publicly available data as soon as is feasible
b.   Correct and appropriate data interpretation is solely the responsibility of data user
c.   Results, conclusions, and/or recommendations derived from the data do not imply endorsement from ICES
d.   Data sources must be acknowledged, preferably using a formal citation
e.   Data users must respect any restrictions on the use of data
f.   Data users are requested to inform ICES of any suspected problems in the data
g.   Data users are encouraged to inform ICES of possible sources of relevant information

4   Contribution of Data

a.   All data provided to ICES are considered to be publicly available unless otherwise explicitly specified and agreed

b. The data provider must be authorized to provide the data
c. Data contributions should be made as soon as possible after the data collection event
d. All data including meta-data, supplemental information and quality indicators should be provided using standard codes, formats, and protocols to the extent possible. Further guidance can be found on the ICES website http://www.ices.dk/datacentre/Submissions/
e. The quality assurance of data is the responsibility of the data provider
f. Data providers are requested to inform ICES of any national policies that may place special conditions on the redistribution of data

5 Redistribution of data

a. Data may only be redistributed, i.e., made available in other data collections or data portals, with the prior written consent of ICES
b. Redistribution of meta-data is always allowed

6 Data Quality

a. ICES develops and applies quality assurance procedures as appropriate and feasible, and in cooperation with data providers, ICES Expert groups and other organizations
b. ICES may be informed of potentially erroneous data. ICES will ensure that data providers are informed of quality issues
c. The ICES Data Centre will never change the original data record from a data provider, but may undertake conversions or transformations of that data to allow its inclusion in ICES databases
d. Although the ICES Data Centre may perform some data quality control, the data provider always retains complete responsibility for data quality
e. Data users are responsible for proper use of the data, including regard to data quality

7 Data Citation

Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data.

a. ICES stores Persistent IDs (PIDs) when supplied with a dataset – at what level the DOIs are created at will then not be in ICES' control. The provision of PIDs should be best practice not a requirement. ICES is able to mint DOIs to datasets created within ICES.
b. A data citation should include a persistent method for identification that is machine actionable, globally unique, and widely used by a community. Where DOIs exist these should be used but otherwise the existing citation guidelines should be used. ICES provides specific examples of citing using the DOIs should be given - specifically how different levels in the hierarchy should be cited.
c. PIDs are available even for datasets with restricted access

Supplemental information to the ICES data policy

Motivation Objective and Framework for the Data Policy

This policy sets the framework for ICES' work involving data collected, evaluated and/or used. It gives rules needed as a prerequisite to make ICES data and ICES work attractive to a wider public and to clarify rules and procedures with regard to data used by ICES expert groups. It will therefore improve the capacity of the ICES to provide quality advice in an ecosystem context.

Elements on how to implement the policy within ICES can be found in the ICES Data Strategy and the ICES Data Centre Business plan.

Data used might be stored in centralised or distributed systems. In a centralised database all data are physically located at, and served from, a single location. A distributed database is where data can be located at various geographically distributed nodes (but still be accessible through one central node or hub). The ICES data policy applies to data managed by ICES and to ICES activities for providing access to data managed elsewhere. A full overview of databases (centralised and distributed) available at ICES can be found at http://www.ices.dk/datacentre/Submissions/index.aspx

By maximizing the availability of data to the community at-large, ICES promotes the use of these data, thereby ensuring that their maximum value can be realized and thus contribute to an increased understanding of the marine environment.

The ICES data policy is 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. A review of data policies relevant to ICES' work is given in annex 4 of the Report of the ICES Study Group on Management of Integrated Data (ICES CM2005/ACE:03, Appendix 1).

Data security and storage

a. ICES makes every effort to ensure that data received are handled and stored in a way that preserves the integrity of the data as it was provided to them
b. ICES maintains an accession system that ensures that all data can be identified in the system, and any resubmissions of data are recorded as such
c. All data, meta-data and supporting information are stored as original files and also as part of the database systems backups

Use of Data

The ICES website is a key focal point in disseminating information to the ICES community and beyond. Data might be quality controlled (see below): regardless of whether the data is quality controlled or not, ICES and the data provider do not accept any liability for the correctness and/or appropriate

interpretation of the data. Interpretation should follow scientific rules and is always the user's responsibility.

Users must acknowledge data sources, as 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 provider prior to publication, and should possibly consider the data provider(s) for co-authorship of published results.

All data held by ICES should eventually become publicly available, with due regard to relevant legislation. However, access to sensitive data may be restricted or data may be aggregated for a limited period of time if specifically stipulated by the data source (see below). Also, the use or reproduction of data for commercial purposes might require prior written permission from ICES and/or the data source.

Users are requested to inform ICES of any problems encountered with ICES-provided data. A timely and easy-to-use feedback procedure will be available, aimed at correcting data at the data source. This feedback will increase the quality of the data and therefore cover one aspect of added value through open access to data.

To become the focal point for marine data in the North Atlantic, ICES will continuously expand its data repositories as well as links to external data. Users are therefore encouraged to contribute information on data sources currently not available through ICES, but possibly important for ICES' work.

Citation of Data

Data Sources should be acknowledged by a citation. The citation must include as a minimum a reference to the ICES database where the data extraction was made and the year in which the database was referenced. Preferably, data is cited by using the dataset's PID. When no PID is available, one can cite the dataset using one of options below can be used as examples:

Examples of citation are given below:

standard citations

"ICES Historical plankton dataset 2011. ICES, Copenhagen"

"ICES EcoSystemData data portal, 2012. ICES, Copenhagen"

Extended citations

"ICES Database of Trawl Surveys (DATRAS), Extraction 3 JUNE 2012 of International Bottom Trawl Survey (IBTS). ICES, Copenhagen"

"ICES Environmental database (DOME), Extractions 3-10 JUNE 2012; Chemical data for the OSPAR CEMP, Reporting laboratory(s) via British Oceanographic Data Centre (UK). ICES, Copenhagen"

A Data Citation may also include a URL to the database, and/or a URL to the meta-data record for the ICES dataset in the ICES Spatial facility (http://geo.ices.dk). Additional citation information is made available in the

Disclaimer file that accompanies the data download under the section 'Data Acknowledgment'.

Data citation should follow community best practices, please refer to the 8 principles of data citation available here[1]

Contribution of Data

Data providers may be the originators of the data/information, for example, persons responsible for the scientific work that produce the data/information; or an intermediary such as the data providers' associated institute(s), the agency that commissioned or funded the work, or even the information technology group responsible for preparing the data for submission to ICES. The data provider must precisely specify any access restrictions that it wishes ICES to uphold. Some cases that call for restrictions include data which is protected by law and data submitted during a prescribed period of exclusive use (which is normally not more than two years for data from scientific origin - the time needed for initial collation and quality control). Restricted access will be considered on a case-by-case basis. ICES urges data providers to re-enforce their commitment to free-of-charge and unrestricted use of their data.

Data and information are provided to ICES from many data sources. They are of variable quality and can be obtained using a variety of methodologies. Three types of data are distinguished:

1. Detail data are individual measurements or observations. In order to interpret detail data, expertise as well as related attribute data such as type of date, location, time and unit of measurement are also required
2. Aggregate data are summarized detail data
3. Meta-data are data about data. That is, they provide information about detail or aggregate data sets. Examples of meta-data include accuracy, precision or method of measurement, and location, structure or ownership of the data.

In order to maximize the usability of data and thereby their value, data providers 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.

Speed is a primary factor determining the usefulness of data, 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 ICES. For example, it will be possible for data sources to submit multiple versions of the same data set during the process of quality control.

End-to-end data management (data life cycle) is encouraged (see Annex 3 of ICES CM 2005/ACE:03; BCO-DMO Data Management Best Practices Guide, www.bco-dmo.org/resources).

---

[1] Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014

Some -often older- data may be unintentionally destroyed or lost. ICES strives to rescue and archive valuable data relevant to the ICES mission that are at risk, including those residing in reports and documents. The data provider, however, is responsible for the providing sufficient documentation with the data.

Data Quality

To indicate the quality controls that have been applied to a specific data set, ICES' systems will accommodate quality flags. The system will allow re-submission of data throughout the quality control process, and thus also allow for accelerated submission of data. For example, preliminary data can be submitted immediately after collection and replaced later by cleaned data.

Finally, the reporting of suspected errors in the data will be facilitated, and that information relayed to the respective data source so corrections can be made.