

PLANNING GROUP ON DATA NEEDS FOR ASSESSMENTS AND ADVICE (PGDATA)

VOLUME 2 | ISSUE 105

ICES SCIENTIFIC REPORTS

RAPPORTS
SCIENTIFIQUES DU CIEM



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

The material in this report may be reused for non-commercial purposes using the recommended citation. ICES may only grant usage rights of information, data, images, graphs, etc. of which it has ownership. For other third-party material cited in this report, you must contact the original copyright holder for permission. For citation of datasets or use of data to be included in other databases, please refer to the latest ICES data policy on ICES website. All extracts must be acknowledged. For other reproduction requests please contact the General Secretary.

This document is the product of an expert group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the view of the Council.

ISSN number: 2618-1371 | © 2020 International Council for the Exploration of the Sea

ICES Scientific Reports

Volume 2 | Issue 105

PLANNING GROUP ON DATA NEEDS FOR ASSESSMENTS AND ADVICE (PGDATA)

Recommended format for purpose of citation:

ICES. 2020. Planning Group on Data Needs for Assessments and Advice (PGDATA).
ICES Scientific Reports. 2:105. 36 pp. <http://doi.org/10.17895/ices.pub.7571>

Editor

Joël Vigneau

Authors

Rui Catarino • Chun Chen • David Currie • Laurent Dubroca • Edvin Fuggelbak • Jon Helge Vølstad •
Maksims Korvsars • Juka Põnni • Tiit Raid • Perttu Rantanen • Marie Storr-Paulsen • Christoph Stran-
sky • Josefina Teruel • Els Torreele • Sieto Verver • Joël Vigneau



ICES
CIEM

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

Contents

i	Executive summary	ii
ii	Expert group information	iii
1	Introduction.....	1
	1.1 Background	1
	1.2 PGDATA 3-year programme 2018–2020.....	1
	1.3 ToRs for PGDATA 2018–2020.....	2
2	Development of a quality assurance framework on data collection and data processing	3
	2.1 Proposal for an ICES structured approach	3
	2.2 Proposal for inclusion in the ICES QAF.....	3
	2.3 Quality certification or accreditation for institutes participating in data collection for fisheries advice	9
	2.4 Development of SISP.....	10
	2.5 Communication and cooperation with expert groups.....	12
	2.6 Statistical advice and guidance	13
3	Methodological procedures and quality estimates from past ICES technical workshops and working groups.....	15
	3.1 Accessibility to recommendation or good practice	15
	3.2 Identify gaps and needs for statistical and/or tools developments	23
	3.3 Online monitoring of sampling programs (WAO), Ifremer	23
	3.4 Discussion and Summary	24
4	Communication and feedbacks on data issues	25
	4.1 Review and comment on ICES data call	25
	4.2 Publication and communication on findings	25
	4.3 Data Confidentiality	26
5	Future of PGDATA	28
	5.1 Lessons from the 2018–2020 period	28
	5.2 Proposal for the 2021–2023 program.....	29
Annex 1:	List of participants.....	32
Annex 2:	Resolutions	33

i Executive summary

In 2018, PGDATA started a second 3-year programme with renewed terms of reference. After having achieved some practical and concrete objectives in its first 3-year programme, PGDATA entered a round of discussions with ICES on its future and considered some of the weaknesses that appeared in the first years. The new objectives assigned to PGDATA were to focus on the development of the Quality Assurance Framework (QAF) for both fishery-dependent and fishery-independent data, create links between the different expert groups, promote the statistical improvements and good practices for implementation and make them easily accessible to the public.

An ICES structured approach for a Quality Assurance Framework is proposed, taking into account all ICES initiatives in the field of collection, processing and storage of fisheries dependent and independent data. This framework, compliant with the principles developed in the European Statistical Standard, has been presented to ICES in 2019 and is now integrated in the ICES Quality Assurance Framework for advice.

The accessibility to recommendations and good practices has been addressed through a restructuring proposal of the ICES Quality assurance repository. The proposal makes use of the ICES website development and search facilities and will need the involvement of several ICES working groups to help with the implementation. PGDATA is ambitious on this accessibility of past reports and guidelines and would like to set the stage for a longer-term achievement with living documents classified by topics – this would include all recommendations and good practices produced by the wealth of ICES technical workshops and working groups.

The communication and feedback on data issues with assessment working groups was given a special focus, acknowledging the previous difficulties in establishing active interaction and trying to learn from this experience. A large collation of exploratory figures produced by the assessment working groups in their reports has been undertaken and classified by topics in an annex of the PGDATA 2018 Report. The objective was to demonstrate the large creativity undergoing in this field, to propose a catalogue of what is done for every end-user group and set the stage for a forum like WGCATCH and WGISDAA to take over some ideas and develop generic figures capturing the main information needed for the end users. It is the belief of PGDATA that the exploratory figures used on the entry data for assessment models are the link between the data collection and processing world, the QAF and the assessment and advice world.

PGDATA also proposed a theme session on data collection for the ICES ASC 2019 and 2020 sessions, but these proposals did not reach the stage of being eventually included in the ASC programs. The idea was to prepare an edition of a special issue in a journal (ICES J. Mar. Sci.) on the findings.

At the end of this 3-year programme, PGDATA proposed to discontinue its activities and proposed a new working group dedicated to support ICES in implementing a QAF on the data collection field named WGQUALITY.

ii Expert group information

Expert group name	Planning Group on Data Needs for Assessments and Advice (PGDATA)
Expert group cycle	Multiannual fixed term
Year cycle started	2018
Reporting year in cycle	3/3
Chair	Joël Vigneau, France
Meeting venues and dates	13–16 February 2018, Nantes, France (14 participants)
	15–18 January 2019, Copenhagen, Denmark (8 participants)
	21–24 January 2020, Copenhagen, Denmark (10 participants)

1 Introduction

1.1 Background

In 2013, the ICES Planning Group on Commercial Catches, Discards and Biological sampling (PGCCDBS) recommended a shift of the practical work into two separate expert groups, one dealing with collection, interpretation and quality assurance of data on commercial catches (WGCATCH: ICES, 2014) and the other on biological parameters (WGBIOP: ICES, 2015). The remaining work was given to PGDATA, which was tasked, over the period of 2015–2017, to improve the effectiveness of the ICES benchmarking process and the quality of ICES advice, and to ensure the best use of available resources for data collection. The objectives of this first 3-year programme and the lessons learned are summarised in the PGDATA 2018 report (ICES, 2018). These helped ICES to develop the second 3-year programme from 2018 to 2020.

In its first year of the 3-year programme (2018), PGDATA elaborated on each of the ToRs with the idea of agreeing on goals to be achieved within 3 years and the associated roadmap. In 2019 and 2020, PGDATA held its meetings in the ICES Secretariat and, although there is no report for 2019, this document will integrate all findings for the 3-year period and will make recommendations for the future.

1.2 PGDATA 3-year programme 2018–2020

For the new 3-year programme, the focus was put on the development of the Quality Assurance Framework (QAF) for both fishery dependent and fishery independent data, and on creating links between the different expert groups. The statistical improvements and good practices should be put in context, promoted for implementation, and easily accessible to the public.

- i. Design a Quality Assurance Framework to ensure that information on data quality is adequately documented and applied in assessments;
- ii. Ensure consistency of approach for fishery dependent and fishery independent data quality framework, and complementarity with approaches developed in other fora such as STECF, EU-MAP;
- iii. Identifying improvements in data quality, or collections of new data, that have the greatest impacts on the quality of advice;
- iv. Improve or create communication routes between data collectors, data managers and end-users, and advise on new approaches to ease the implementation of the QAF (through publication, RDB-development and, cooperation with other WG including shared workshops).

The terms of reference developed below were meant to focus on methods and their evaluation rather than providing solutions to a specific data issue or recommending a single method to be used in all cases. The reason for this is that many assessments and data collections follow different methodologies and have different assumptions so that a universal answer is unlikely to be appropriate. The new objective is to gather the existing information on data quality in a structured way, develop expertise and tools where gaps are identified, develop communication with end-users, and maintain knowledge of the work done.

1.3 ToRs for PGDATA 2018–2020

a) Implement and maintain Quality Assurance Framework for assessment expert groups to evaluate data quality and its impact on assessments:

- i) Propose a structured approach for agreement within ICES, including the development of the ICES/RDB for detailed fisheries data, and develop a “best practice SISP” for data collection in support of stock assessment;
- ii) Collaborate with EOSG expert groups to identify problems and prioritize actions to progress and improve quality data collection;
- iii) Provide a service to EOSG expert groups for statistical advice and guidance on sampling design to promote good practice seeking to establish effective two-way communication;
- iv) Cooperate with assessment expert groups to show and demonstrate the effects of data collection methodology on the advisory assessments to underline the relevance of good practice to the advisory process.

b) Review the outcomes on methodological procedures and quality estimates from past ICES technical workshops and working groups, and advise on ways forward:

- i) Maintain knowledge of the work done and organize accessibility to any recommendation or good practice provided by the variety of technical workshops and propose changes to SISP as necessary;
- ii) Review the work done in other fora such as STECF and EU-MAP in order to integrate the initiatives and propose complementary work;
- iii) Identify gaps and needs for statistical and/or tools developments, and initiate workshops as needed.

c) Propose ways to improve the communication and feedbacks on data issues:

- i) Review and comment on ICES data call;
- ii) Organize participation to end-user meetings to seek for mutual beneficial improvements;
- iii) Promote publication on findings, likely in the form of peer-reviewed publication (e.g. CRR) that documents the development of methodologies in the field of data collection and the state of scientific knowledge on the topic at the end of the 3-year TOR period.

This document addresses all the terms of references and sums up 3 years of discussion and meetings. It complements the elements already expressed in the PGDATA report 2018 (ICES, 2018). All elements discussed during PGDATA meetings in 2019 and 2020 are developed in this document, given that there was no report for PGDATA 2019.

2 Development of a quality assurance framework on data collection and data processing

2.1 Proposal for an ICES structured approach

PGDATA 2018 was tasked to initiate the development of a Quality Assurance Framework (QAF) on the collection and processing of data needed for assessment and advice. The expected benefits range from quality improvement (reducing errors etc.) to improve transparency and from improved accessibility to standardising tools (PGDATA 2018). These benefits may extend to other fields, such as MSFD work and regional coordination and cooperation.

Within the QAF initiated by PGDATA, ICES is the point of focus and not, e.g. individual countries or other data providers. As part of the QAF however, knowledge of implemented quality control systems at the data provider side is required. To this aim, PGDATA identifies that producing an inventory of quality control systems currently in place at data providers would be needed. As a start, this inventory should focus on certified/accredited systems such as ISO or CoreTrustSeal, while internal (referenceable) quality control systems can be registered as well.

In 2018, from the different possibilities to structure a QAF, PGDATA proposed a general framework following the principles developed in the European Statistical Standard and its Standard for Quality Reports Structure. Building upon this core framework, PGDATA 2019 and 2020 proposed to extend the QAF following *in extenso* the 2017 version of the “European Statistics Code of Practice for the National Statistical Authorities and Eurostat” (ESCoP2017).

2.2 Proposal for inclusion in the ICES QAF

ESCoP2017 is based on three building blocks covering 15 (+1) principles. The building blocks are Institutional Environment (basically the fundamental prerequisites for all processes under the QAF), Statistical processes (focussing on methodologies as well as cost-effectiveness) and Statistical Output (covering comparability, data access and, e.g. timeliness). In 2018, PGDATA included seven principles in the QAF proposal. As topics like professional independence, allocation of resources and data confidentiality form an inseparable part of a QAF, PGDATA further proposed to update the 2018 proposal by taking all 15 principles on board for the QAF. The addition from 2018 are indicated to distinguish the version used by ICES for its QAF (issued from PGDATA report 2018) and the new elements added in 2020.

The following section provides an overview of all 15 principles while addressing all ICES initiatives in the field of data collection, the underlying ICES general principles as well as addressing potential gaps in the QAF that need further elaboration to be covered sufficiently in the QAF.

Institutional environment

o PRINCIPLE 1: Professional Independence

Item added in 2020.

Professional independence is fundamental to ICES as an organisation and covered under the ICES Vision (ICES, 2019a): “To be a world-leading marine science organisation, meeting societal needs for impartial evidence on the state and sustainable use of our seas and oceans”.

o **PRINCIPLE 1bis: Coordination and cooperation**

Item added in 2020.

Communication with ICES Members and their representation in various ICES fora form the basis for coordination and cooperation within the ICES community.

PGDATA plays a pivotal role in the communication with end-users and data providers while interacting with many ICES fora like EOSG, ACOM and SCICOM. However, such a span of attention relies on a functional system for interaction and transfer of ideas, views and decisions. This system is currently not sufficiently set up to release its full potential from PGDATA or other dedicated groups (issue also addressed Section 2.5).

o **PRINCIPLE 2: Mandate for Data Collection and Access to Data**

Item added in 2020.

Within the context of the QAF, the ICES mandate for data collection is to be seen as a mandate to collate data rather than the actual collection of primary data. The technical means to collate the data range from centralised complex data calls covering (almost) all Assessment Working Groups, to standardised routine uploads of data to ICES databases. For EU MS, ICES, in the role of end-user of scientific data, has the legal right to request data collected under the EU Data Collection Framework for advice purposes.

Based on the ICES strategic plan (ICES, 2019b), ICES will “continue to develop services and tools to enable visualization and easy access to these data for a broad range of users”. Thus, ICES is continuously striving to make more data available to the public. While doing this, adequate data access protocols addressing, e.g. GDPR measures as well as documentation of access, shall be in place.

o **PRINCIPLE 3: Adequacy of Resources**

Item added in 2020.

This principle encompasses many different topics. From the “institutional environment” point of view, this principle relates to ICES having adequate resources available to fulfil its tasks, not only in terms of budget but also in terms of qualified personnel and tools supporting the work in an adequate way.

How to prioritise tasks, budget and effort is currently not spelled out, other than the aims specified in the science plan and the strategic plan. The number of Wks, WGs and other meetings in relation to the number of potential qualified participants and their budgets is another issue to consider in the context of adequate use of resources.

From a data providers point of view, statistical tools are needed to optimize sampling e.g. in number of fish to sample in a stratum, number of hauls during a survey etc. Some work is done at groups as BIOPTIM and the scheduled WKUSER.

Tools facilitating efficient access and storage of documents are required as well.

o **PRINCIPLE 4: Commitment to Quality**

Item added in 2020.

Quality, more precisely, continuously improving and maintaining a high level of quality throughout the ICES Advisory processes, is fundamental to the provision of reliable, trustworthy and sound advice. The current implementation of the QAF (and TAF as a specific element) demonstrates the commitment to quality. Response and review loops are embedded in the ICES working procedures, thus highlighting the strengths and weaknesses of the advisory processes.

The 2019 ICES Advisory plan highlights ICES' commitment by stating that "... quality assurance of the advisory processes will underpin our role as an independent and legitimate evidence provider". Having the commitment included in the Advisory plan ensures top-level commitment to quality assurance, rather than a scattered and non-uniform approach to it.

o **PRINCIPLE 5: Statistical Confidentiality and Data Protection**

Item added in 2020.

Statistical confidentiality and data protection are fundamental principles facilitating ICES to store and process huge amounts of data, confidential or not. The enormous amount of data is the treasure of the Advisory process and should be treated with great care not to damage the trust given by the contributors. Various data policies are in place under the governance of relevant ICES groups or ICES itself. These policies shall be reviewed on a regular basis to reflect latest insights and legal adaptations, e.g. the implementation of the GDPR in recent years. When setting policies, these policies need to be workable and transparent, thus avoiding unnecessary restrictions preventing efficient data use in exchange. See also Section 4.3.

o **PRINCIPLE 6: Impartiality and Objectivity**

Item added in 2020.

For ICES, this principle is clearly interlinked with Principle 1 and to the ICES vision and mission and expressed in the Strategic plan. Apart from the Vision, the plan revolves around the statement that impartial evidence is essential for responsible decision making.

In the near future, industry data will be added as a (potential) data source for the advice process. It is of utmost importance that data is collected, processed and provided to ICES in a transparent, traceable and documented way to secure the impartial procedures within ICES as well as to ensure the position as an impartial organisation towards the outside world.

Statistical processes

o **PRINCIPLE 7: Sound Methodology**

Sound methodology underpins quality statistics. This requires adequate tools, procedures and expertise as referred to under principle 3 as well. Guidelines and good practices for data collection have to be developed (See Section 2.4 on SISP). TAF provides an online open resource where all data input, codes and outputs are traceable and versioned.

o **PRINCIPLE 8: Appropriate Statistical Procedures**

Appropriate statistical procedures underpin quality statistics; however these procedures have to be implemented throughout the entire chain of data processing.

Guidelines and good practises for data processing are being developed (further) as well as tools (TAF R Scripts) and software to support the work.

o **PRINCIPLE 9: Non-excessive Burden on Respondents**

Item added in 2020.

As part of its Advisory Plan, ICES tasks itself to ensure that all advice products are based on data that adhere to the FAIR principle. Also, the development of comprehensive quality management systems is listed as one of the future tasks. The development of e.g. RDBES falls directly under this task.

Being a data intensive organisation, ICES should strive to limit the burden on data providers while ensuring the availability of crucial data in a timely manner. By combining multiple data

calls, respondents can provide data efficiently and provide data for multiple use in one response. For a list of recurring advice groups pre-approval for the use of the data can be granted, thus limiting the administrative burden to allow access to the data.

Comprehensive databases shall facilitate the FAIR principle where possible and additional features (e.g. standard reporting, consistency checks) shall aim at reducing the workload for both Working Groups as well as for data providers.

As a basic principle, data is only asked for when based on a specific request and preferably combined within other data calls, thus limiting the burden for data providers.

o PRINCIPLE 10: Cost Effectiveness

Item added in 2020.

Cost-effectiveness is a cross-cutting theme throughout the advisory process, from the basic data collection to data provision and advice. A dedicated theme session was held at the 2018 ASC on cost-effectiveness.

Cost-effectiveness should be a common topic when e.g. requesting additional data collection and data deliveries. As described under Principle 9, combining data requests reduces the burden on respondents.

Cost-effectiveness is not about budgets only but shall be seen in a wider context such as carbon footprint (flights, but mostly surveys) and animal welfare (reduction of the number of hauls, haul duration and e.g. number of biological samples). ICES needs to be at the forefront to address these issues and initiate and support the development of more effective sampling procedures and surveys. Improving the effectiveness of a survey should be a general ToR for survey groups.

In general, when and where possible data collection should be made more effective to reduce all sorts of costs related to sampling. This can be done through strengthening the support for development of regional sampling plans, improved cooperation with the industry and promoting multiple use of survey time and data.

Statistical output

o PRINCIPLE 11: Relevance

The relevance of the Advisory process is a key element in the Advisory plan. The dialogue with requesters of advice as well as the wider society shall ensure maintaining the relevance of the advice provided by ICES. The advice shall meet the requester's demands through improved formulation of requests and the subsequent process of knowledge synthesis and advice production.

This process is strengthened through the constant communication with data providers, continuous improvement of the advisory process (up to date issue lists, benchmark processes including data compilation workshops)

o PRINCIPLE 12: Accuracy and Reliability

Accurate and reliable data is obviously critical to the advisory process. The reliability of data is under constant review. This review is supported by the implementation of data quality indicators, RDBES, reviews of sampling designs and estimation procedures and e.g. traceable data use throughout the process.

As stated under Principle 6, industry data will enter the advisory process in the short term. The reliability of this data has to be ensured through clear documentation available to the wider public.

o **PRINCIPLE 13: Timeliness and Punctuality**

Advice shall reach the requesters in time and in a punctual manner. Updates to the advice products are registered and communicated based on a protocol. Accurate data calls based on a clear request and timely delivery of quality controlled data (for EU MS based on CF obligations) form the basis to timely advice.

o **PRINCIPLE 14: Coherence and Comparability**

Data coherence and comparability is strongly supported through a range of processes, from the initial survey and sampling designs to quality controlled biological sampling and referenced data sets. Biological sampling is quality controlled through calibration exercises for maturity staging, stomach sampling and age reading. Guidelines for biological sampling are available to the scientific community and support the internal coherence of the data collection.

o **PRINCIPLE 15: Accessibility and Clarity**

Robust storage facilities to support the Advisory process. By design, these facilities ensure the provision of quality controlled data to the process (DATRAS, RDB(ES), InterCatch) as well as ensure internal consistency and quality control at the basis of the data collection (e.g. SmartDots).

In addition, various web services and repositories provide guidelines and standardized data sets to a wider audience.

In terms of clarity, TAF provides full insight in the Advisory process.

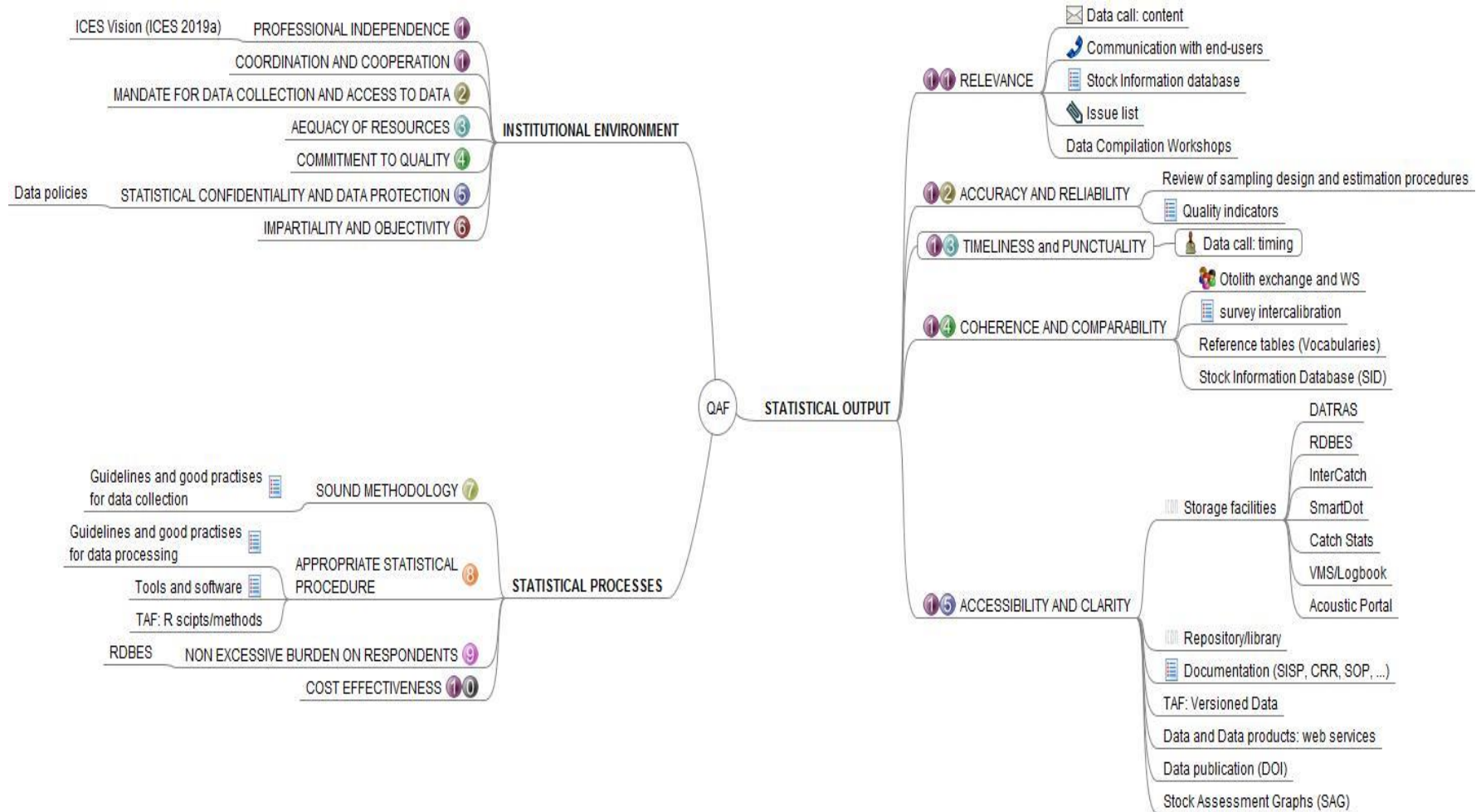


Figure 1: Schematic representation of the proposed ICES QAF

Main outcome: The first proposal (minus the addition made in 2020) was used by ICES/ACOM as part of an internal document related to the roadmap towards Quality Assurance Framework for ICES advice. In 2019, ICES released its Advisory Plan where the QAF was one of the main priorities.

2.3 Quality certification or accreditation for institutes participating in data collection for fisheries advice

PGDATA reflected on the benefits for institutes participating in data collection for fisheries advice to become certified or accredited. The difference between these two concepts is that certification represents a written assurance by a third party of the conformity of a product, process or service to specified requirements. Accreditation, on the other hand, is the formal recognition by an authoritative body of the competence to work to specified standards (<https://www.ukas.com/news/what-is-the-difference-between-accreditation-and-certification/>).

Under the ICES Data Policy and EU Data Collection Framework, data quality is ultimately the responsibility of the data submitters. However, ICES is extremely reliant on the quality of the data it receives – poor quality input data will often result in poor quality output (such as increased uncertainty, more errors, and reduced confidence in advice). A number of different labs/countries have already received certification or accreditation under different schemes and this can give ICES confidence that well-defined, reproducible processes are defined by the data providers.

Benefits of certification or accreditation can include:

- A focus on end-users,
- Processes and procedures are documented and kept up-to-date,
- Remove reliance on ad-hoc processes or specific people,
- Provides training resources for new staff,
- Safe operation and continuous access to archived data in a long-term perspective,
- Disciplinary and ethical standards,
- Metadata to enhance retrieval and reuse of data,
- Ultimately increase the quality of outputs.

The down-sides are:

- Costs, specifically in terms of the resources required to attain and maintain certification and/or accreditation,
- A loss of flexibility to quickly adjust processes,
- The certification/accreditation might only cover a portion of a lab's work,
- There are a number of different standards – which one to use and where to start?

In 2019, PGDATA recommended that ICES and each institute involved in data collection for fisheries advice consider getting accredited / certified from an appropriate body in order to improve the quality of ICES input data.

Since then the ICES Data Centre has decided to aim for obtaining the “CoreTrustSeal” certification (<https://www.coretrustseal.org/>). There is currently a discussion within ICES as to what type of accreditation could be appropriate to the wider ICES advice process.

A number of national institutes that contribute to ICES are also certified/accredited under different schemes. The following is a non-exhaustive list of these.

IODE Accredited National Oceanographic Data Centre

https://www.iode.org/index.php?option=com_content&view=article&id=61&Itemid=100057

- Royal Belgian Institute of Natural Sciences, Operational Directorate Natural Environment, Belgian Marine Data Centre, Belgium
- Flanders Marine Institute/ Vlaams Instituut voor de Zee (VLIZ), Belgium
- Ifremer, France
- Marine Institute, Ireland
- National Oceanography Centre, Natural Environment Research Council, Proudman Oceanographic Laboratory, UK

Data CoreTrustSeal Certification

<https://www.coretrustseal.org/why-certification/certified-repositories/>

- Norwegian Marine Data Centre (NMD), Norway
- Flanders Marine Institute, Belgium

ISO 9001:2015 Certification

<https://www.iso.org/obp/ui/#iso:std:iso:9001:ed-5:v1:en>

- Ifremer are ISO 9001:2015 certified by AFNOR Certification, which is accredited by COFRAC who are the French National Accreditation Body (NAB)
- Wageningen Marine Research are ISO 9001:2015 certified by DNV GL who are accredited by RVA, the Dutch NAB.

ISO/IEC 17025:2017 Accreditation

<https://www.iso.org/obp/ui/#iso:std:iso-iec:17025:ed-3:v1:en>

- The Marine Environment and Food Safety Services (MEFSS) group in the Marine Institute, Ireland obtained ISO/ 17025 accreditation from the Irish National Accreditation Board (INAB).

2.4 Development of SISP

PGDATA agreed with the idea of developing templates for ICES Survey Protocols (SISP) for the national commercial catch sampling programs with the aim of providing good practice for each sampling type regarded as following Statistical Sound Sampling Schemes. According to ICES guidelines the SISP document should be “a manual whereby a person, with some expertise and experience in conducting surveys, but none in these particular surveys, could find all of the information required to competently conduct one of the surveys.”

For fisheries independent data, PGDATA acknowledged the existence of SISPs already developed, and used them as a starting point for a development of SISP for the commercial catch sampling programs. The fisheries independent SISPs reviewed in PGDATA were not all structured similarly, although the SISPs convey the same type of information (description of the survey, survey sampling design, sampling protocols, data analysis, etc.). The expected benefits of having the same structure is that it will be easier to (i) reach a specific information in different SISPs, (ii) develop a new one and (iii) point at missing parts to be completed if any. For this

reason, PGDATA would suggest that the same template is used for the commercial catch sampling SISPs.

It was recognized that not all countries / institutes will be able to fill out the entire template of the suggested commercial catch sampling SISPs (PGDATA 2018) but it was considered important to start the process with describing the sampling design and this process will further ease the transformation from InterCatch to RDBES if countries have a standardized SISP describing the commercial catch sampling programs. As a change / update in the sampling design could potentially have a large effect on the quality of the catch sampling data this should be addressed with the versioning of the SISP. Further, the SISP could also inform on international sampling plans under development in the RCG.

For guidelines on best practice please see: <https://www.ices.dk/community/Pages/PGCCDBS-doc-repository.aspx>

PGDATA presented the proposed SISP for commercial catch sampling programs to WGCATCH and WGISDAA for them to review. WGCATCH informed on the development of R markdown summary reports with the purpose to inform the end-users on the data quality for every data delivery. PGDATA suggested that R markdown summary should not repeat the content of the SISP but refer to the name and version of the SISP used to collect and process the data of interest. In order to progress on the elaboration of the SISP, proposed that content of a SISP should be as followed:

1. Description of the survey
2. Survey sampling design
 - a. Sampled population vs total population
 - b. Description of sampling units
 - c. Stratification
 - d. Allocation procedure
3. Sampling procedure
4. Data storage
 - e. National, International
5. Data quality checks and validation
 - f. National, International, Quality indicators
6. Estimation procedure
 - g. For each parameter, including variance estimators

PGDATA was of the opinion that the two approaches (SISP and dataset specific R markdown summaries) are complementary, it will however be important to refer to the same sampling program naming to get to optimal use of the description of the two initiatives.

However, to more easily be able to use other nations sampling programs in the future RDBES, PGDATA recommends that a simple similar naming structure should be applied to all commercial catch sampling programs;

- On-shore
- At-sea observer
- At-sea self-sampling

Further, due to the development of the RDBES, PGDATA discussed the need to have proper reference and common naming convention between the national sampling description and the

13 classes of sampling types develop in the RDBES. This indicates that a scheme should refer to one of the 13 RDBES classes.

Main outcome: SISP in the world of fisheries dependent data collection should be developed following a structured approach. PGDATA could not propose a final recommendation given that each MS would need to develop their own for each dataset, and a reflection is ongoing in the EU-MAP to document the quality of such datasets. The joint effort between EU-MAP, RCG and ICES must continue to enable such document to become a reality.

2.5 Communication and cooperation with expert groups

Communication and cooperation with end-users on the quality and consistency of the data used as entries in the models has been the object of several initiatives in a recent past

- Data contact person to assessment working groups (PGCCDBS)
- Questionnaires to assessment working groups (WGCATCH)
- Data issues list for benchmarks (PGDATA 2015–2017)
- Data recommendation database (ICES)

Lessons have to be learned from these experiences, and it seems difficult for a stock coordinator/assessor to have a deep understanding on how data has been processed in each country. The question remains if it is essential for an end-user to know how data have been processed when preparing a stock assessment. The previous PGDATA 3-year period focused on communicating to the benchmarks and Data Compilation Workshops with this type of information. Before going further, PGDATA wanted to take time to think of a new approach, and began by investigating the types of data analysis carried out during the assessment working group (AWG), and thus reflecting what type of information they would be interested in. From the AWG reports investigated (see Annex III of the PGDATA 2018 report) the review showed clearly that stock assessors were interested in the stability of the time series and how the last addition of year evolved from the time series. Data submitters rarely perform these types of analyses so far, although these are possible at national level, and would be able to spot unexpected changes to be investigated well in advance of the AWG. There's a lot of benefits to perform these analyses at the time of submission, in order to detect potential errors in the sampling and be able to do the corrections, if possible, before the AWG starts. The issue of the amount of work added to data submitters should not be overlooked in this process. In 2018 and 2019, WGCATCH and WGISDAA were recommended to review these diagnostics plots, with the objective of proposing standard plots and initiating a continuous improvement process.

PGDATA recommended some of the participants to visit AWG in 2018 and 2019 to present the PGDATA ongoing work and interact with the end user. PowerPoint presentations were developed to this aim, and a presentation was made at WGBFAS in 2018.

The Annex III of PGDATA 2018 report illustrates also the creativity of the different AWG with multiple ways of presenting the same information. This raises the question of duplication of work and a need for sharing and developing data analysis functions in a collaborative way. The current GitHub repos used by AWG are mimicking the AWG structure and thus cannot be used for collaborative work across AWG. The development of R functions for the RDBES purpose can be the occasion to set up a GitHub on data exploratory analysis, which would include the types of exploratory graphs seen in AWG reports. AWG experts would be welcomed to participate

collaboratively in the development of such methods which would be useful for their own purpose.

PGDATA recommends ICES, SCRDB and WGCATCH to develop a RDBES exploratory analysis GitHub repo with codes to check quality and consistency of datasets, which would be used for data submitters and during Data Compilation Workshops. The data exploratory graphs and codes should be developed based on already existing codes used in the different AWG GitHub repos. AWG experts are welcomed to participate collaboratively in the development of these codes.

Main outcome: the communication of PGDATA views and expectations to the assessment working groups remained mainly at the status of intention. The progress made by PGDATA is on identifying a region of common interest between the data providers and data users, which is the consistency of the last data year vs its historical series (respectively at a national and stock level). Common approaches and tools would need to be developed between the two worlds.

2.6 Statistical advice and guidance

When exploring the data from InterCatch (IC), stock assessors have to evaluate means to impute data to unsampled strata. Borrowing information from relevant and well informed strata suppose that the later are of sufficient quality to support a widening of their scope. In this situation, the stock assessors would need mapping in the likes of the one below (Figure 2) which could be run from the output of IC, the idea being to detect quickly areas where the informed strata are not suitable to be used for imputation.

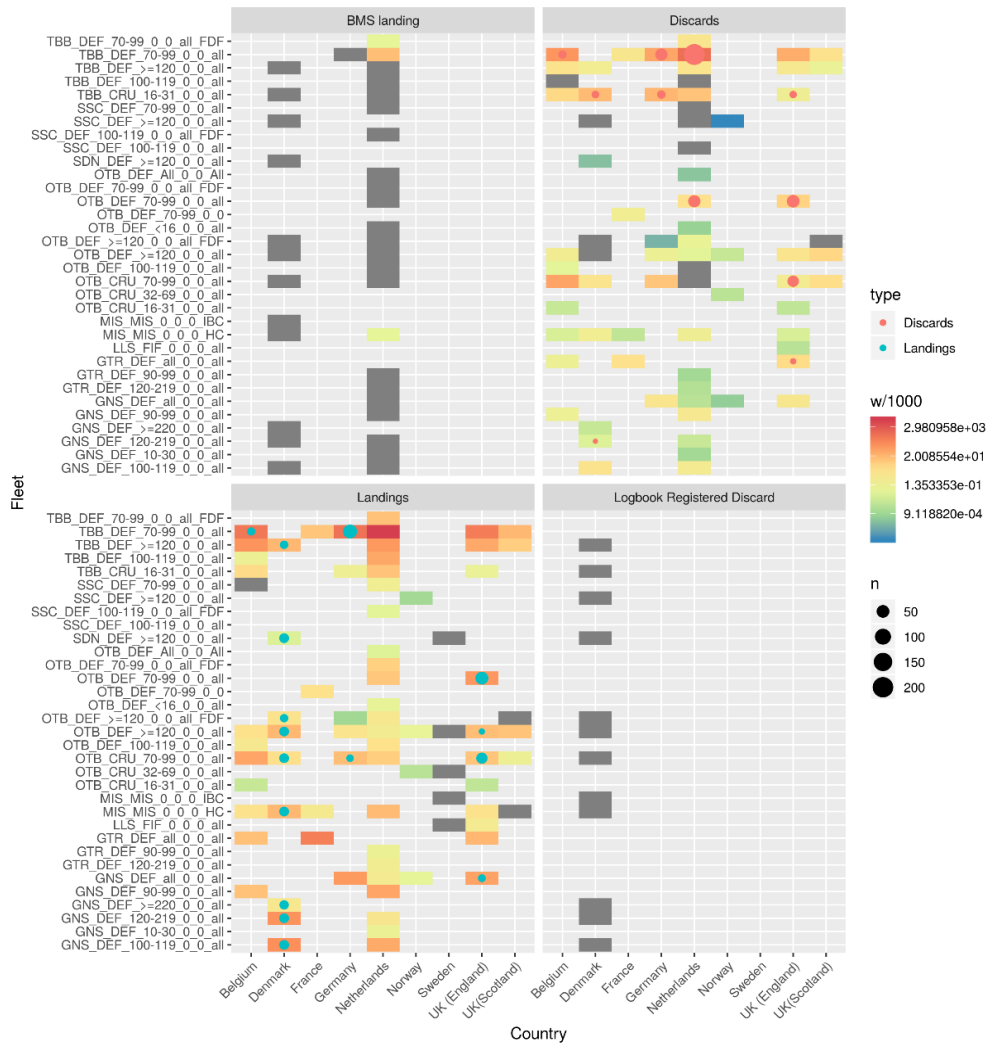


Figure 2: Mapping of catches fractions and related sampling (grey cells relates to Non Available information)

This map, together with a graph on cohort tracking should give a first set of information on the quality of the data used and such tools could be developed in a separate forum (WGCATCH, fishPi-like project...) for the use of ICES stock assessors. Potential analyses can be further explored based on all the graphs scanned in assessment reports (Annex III).

Main outcome: PGDATA recommends this type of graph to be developed as part of sampling overviews by the ISSG on overviews in the RCG GitHub

3 Methodological procedures and quality estimates from past ICES technical workshops and working groups

3.1 Accessibility to recommendation or good practice

This section presents a summary of good practice produced by different groups and organisations.

There are two aspects on the accessibility to any relevant information provided by previous expert groups. The first is about finding easily a desired report, and the second is about organising the information in order to provide directly the excerpts of an expert group related to the domain of interest. The first is about better organising the ICES web page on QAF repository which, year after year, has accumulated more than 200 links to workshops and expert group reports. This means that, a large number of issues have found answers, examples of good practices, recommendations for practical implementation, and as it is, only a few experts in each of these fields are able to dig out this information. The second is about data mining and will necessitate developments that PGDATA agreed to coordinate.

The presentation of more than 200 links to reports could be done manually (as now) or dynamically, and PGDATA recommended exploring this latter option. The first step is to attach agreed tags to each of the reports (past and future), this would enable to present the keywords in a tabulated form. Following the presentation by Ifremer on the QAF for their fisheries Information System based on the life cycle of the data (from the survey design to the transmission of the data to end users), PGDATA proposed a similar approach for the presentation of the links. A first proposal would be as follows:

	Survey design	Control and calibration	Sampling protocols	Recording systems	Quality Guidelines and indicators	Data processing	Data Storage and transmission
Catch	[5], [6]				[3]		
Length	[4], [5], [6]				[3]		
Discards	[5], [6]				[3]		
Age	[5], [6]				[1], [3]		
Maturity					[2], [3]	[2]	
Recreational							
Fish condition							
By-catch							
Stomach contents							
Survey indices							
CPUE/LPUE							
....							

For age and maturity calibration, there is a third dimension with the years where the calibration took place, it would then become:

	Species	Stock(s)	Year	Document
Age	Anchovy	ATL + MED	2016	[7]
Maturity	Herring	NSSH	2010	[9]
Egg staging	Horse mackerel		2015	[8]
Egg staging	Mackerel		2015	[8]

A first attempt to fill the cells was done to illustrate the idea with the following 9 reports

- [1] WGBIOP 2017 Guidelines for Exchanges And Workshops on Age Reading
- [2] WGBIOP 2017 Guidelines for Workshops on Maturity Staging
- [3] Updated guidelines for the ICES benchmark data evaluation process
- [4] Report of the Workshop on implementation studies on concurrent length sampling (WKISCON2)
- [5] Report of the third Workshop on practical implementation of statistical sound catch sampling programmes (WKPICS3)
- [6] Report of the second Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes (WKPICS2)
- [7] Report of the Workshop on Age estimation of European anchovy (*Engraulis encrasicolus*) (WKARA2)
- [8] Report of the Workshop on Egg staging, Fecundity and Atresia in Horse mackerel and Mackerel (WKFATHOM)
- [9] Report of the Workshop on estimation of maturity ogive in Norwegian spring spawning herring (WKHERMAT)

From this exercise, PGDATA demonstrated that a fixed list of tags (headers of lines and columns of the tables above) attached to all documents in the repository would enable presenting the documents dynamically, and this would greatly ease the search for relevant documents.

PGDATA explored the possibility to use the library search facilities of ICES but found that the results of the research were too broad and not accurate enough for the expected usage. The restriction of the search engine to the only documents present in the Quality assurance repository could help and be a supplementary option for the user. Moreover, other information sources exist on the ICES website (e.g. WGCATCH publications: <https://www.ices.dk/community/groups/SiteAssets/WGCATCH-publications.aspx>), but other sources are either not accessible to the public (i.e. held in private sharepoint folders) or accessible only on demand (age-reader forum). For the latter, PGDATA was of the opinion that there was no reason to restrict access (although the restriction is a simple form to fill) and that a sampling design forum would make sense given the many questions from the people in charge of sampling in the field.

ICES data mining tool

The PGDATA 2019 was provided with information about the new data mining platform developed by the ICES to ease the information handling within the ICES publication repository (<http://data.ices.dk/DataMiningICESLibrary/>). The ICES document repository hosts several types of information available (expert group reports, survey protocols, CRR, user handbooks, data outputs, etc. ...) and is available to the public. If the user knows exactly what to look for, then this task is easy to accomplish. In the case where the user wants to make a query or a search to disseminate some information of the library, it can become a little bit more cumbersome.

The screenshot shows the ICES Publications Repository search interface. At the top, there is a search bar and navigation links. Below the search bar, there is a table of search results for 'PGDATA'. The table has columns for Year, Acronym, Publication Type, and Name Document. The results show several documents from 2015 to 2018, all categorized as 'Expert Group Reports'.

Year	Acronym	Publication Type	Name Document
2015	PGDATA	Expert Group Reports	01 Report of the Planning Group on Data Needs for Assessments and Advice (PGDATA)
2015	PGDATA	Expert Group Reports	02 Executive Summary of the PGDATA report
2016	PGDATA	Expert Group Reports	01 PGDATA - Report of the Planning Group on Data Needs for Assessments and Advice (PGDATA)
2016	PGDATA	Expert Group Reports	02 PGDATA - Executive Summary of the Report of the Planning Group on Data Needs for Assessments and Advice (PGDATA)
2017	PGDATA	Expert Group Reports	01 PGDATA - Report of the Planning Group on Data Needs for Assessments and Advice (PGDATA)
2017	PGDATA	Expert Group Reports	02 PGDATA - Executive Summary of the Report of the Planning Group on Data Needs for Assessments and Advice (PGDATA)
2018	PGDATA	Expert Group Reports	file_had_no_title131847078902276142

International Council for the Exploration of the Sea (ICES) - Conseil International pour l'Exploration de la Mer (CIEM)
 ICES Secretariat · H. C. Andersens Boulevard 44-46, DK 1553, Copenhagen Denmark · Tel: +45 3338 6700 · Fax: +45 3393 4215 · info@ices.dk
 Disclaimer · © ICES - All Rights Reserved

The tool is based on discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. The data mining tool is made of three modules: (1) A database that has all the documents of the publication repository, the metadata of the document and for each of the documents the database stores each word in a distinct record. (2) Website that allows querying the database, with some templates for some queries and the challenge here is to be able to facilitate the user to search and extract the information from the database. (3) Web-services that allow the same operations as the website, but the user can use an R script or another interface to interact with the database using the web-services. The overall goal of the data mining process is to extract information from the ICES repository and transform it into an understandable structure for further use.

PGDATA 2019 explored the performance of the data mining tool and found this very helpful for finding the necessary information (documents) from the ICES publications repository. PGDATA notes that in order to achieve good performance of the tool, the certain presuppositions like e.g. knowing the acronym of respective working groups eases the use of the tool.

'Search per Acronym' allows users to find out which words in the reports were used more often. It would be very useful to get a small paragraph where this word was found. The 'Find documents' function allows users to find all documents related to each acronym. The Google advanced search function was tested to better understand the demand. For example, the expression in the search engine: "best practice" sampling age site:ices.dk filetype:pdf could find almost 98 references of ices reports referring to the search with extracts of the paragraphs where the terms were found. PGDATA and the ICES Data Centre agreed on the benefits of developing such a search tool within the ICES website with improved parametrization and reporting.

'Search Expression' allows users searching between all documents in repository interested information. As a result, users are getting a list of acronyms per year with numbers of expression occurrences in each document. The 'Find documents' function allows users to find all documents where expression was found. PGDATA also notes that it would be very useful if the tool could also provide the section where the quested word/expression was found.

Roadmap: The challenge of organising the perpetual flow of working group and workshop reports and finding the right reference to searched expression remains and must be given thorough consideration.

Smartdots

The question of machine learning on age reading appeared recently - a pre-requisite to this would be the sharing of otolith images. This will require a consistency of approach to the images and the application of concepts such as metadata standards - quality indicators should accompany each of the images. The creation of a "smartdots@home" application was discussed but this now seems unlikely to proceed due to the complexity of integrating with each laboratory's proprietary systems.

PGDATA recommends RCG, WGSMArt and WGBIOP to reflect on this and see how standardisation of metadata can be achieved, through e.g. ICES/WGMLEARN.

Marine Institute (Ireland) Data Quality Manual

The Marine Institute, Ireland (MI) is to an IODE accredited data centre - it needed to produce a Data Quality manual to achieve this. The framework for this manual is provided by the "IODE Quality Management Framework for National Oceanographic Data Centres" (https://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=12661).

The purpose of the manual is to define the MI's Data Quality Policy, Data Quality Management System and the management responsibility for quality fulfilment at Marine Institute.

For each directorate and for each team the MI must:

- List the data processes that are within scope
- For each process, describe it using BPMN diagrams (<http://www.bpmn.org/>). An example is shown below (Figure 3).
- Identify where Standard Operating Procedures (SOPs) do or don't exist
- If the process involves personal data there are also GDPR diagrams produced

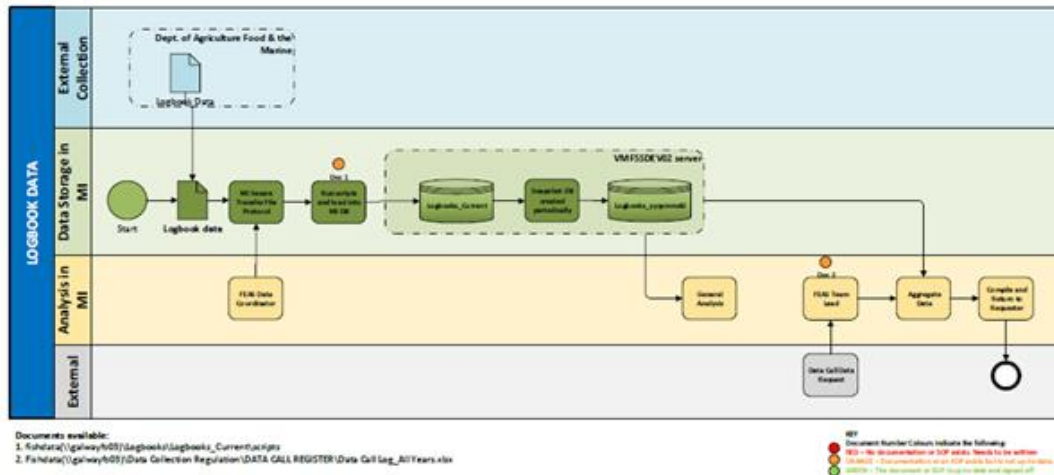


Figure 3: A BPMN diagram as used by MI

Regional Database and Estimation System (RDBES)

The future advantages of using the RDBES are:

- Support the Regional Coordination Groups, RCG with data for coordination
- Ensure approved standardised statistical methods are used for estimating the detailed data for the stock assessment
- Document data used for the stock assessment at detailed level (reproducible, recurrent uploads)
- Higher data quality by using common quality checks across all countries' data
- Reduces the workload for the countries in estimating data because the RDBES contains all needed methods
- One data call for upload of data to the RDBES for the RCGs and estimated for the ICES stock assessment WGs

The RDBES data model should be seen as part of the movements towards:

- Statistically Sound Sampling Schemes (4S),
- Greater regional coordination,
- Transparent Assessment Framework (TAF),
- Improved estimates to ICES assessments.

The data flow of the RDBES is shown Figure 4 below.

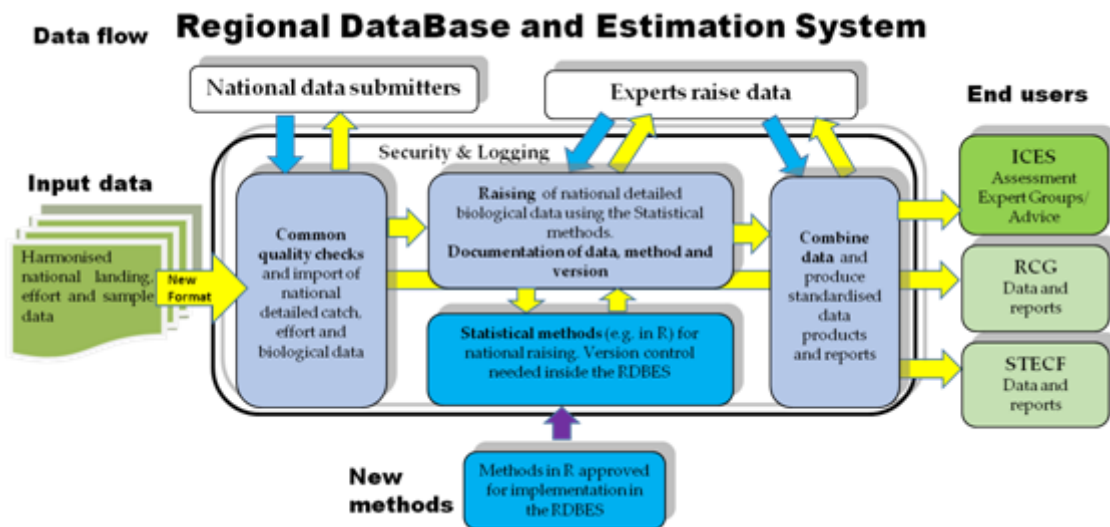


Figure 4: Dataflow of the Regional Data Base and Estimation System

RDBES Status

The RDBES is currently planned to replace both the existing InterCatch and RDB database systems in 2022 and has an important part to play in increasing transparency and improving the quality of stock assessment within ICES. To this end a number of workshops have been planned for 2020 which will both help data submitters with the transition to the new system and encourage more people to be involved in the process. A data call is also planned for 2020 which will give further motivation for people to become involved and provide a robust test of the process.

ICES Council has approved funding for the software development of the RDBES database and web application for the next 4 years. After the first 2 years the progress will be evaluated and if the development requirements are fulfilled the development will continue at the same level for the following 2 years. This should allow the RDBES development to be completed successfully.

The funding requires the delivery of:

- A fully operational ICES Regional Database (RDBES) with a regional estimation system such that statistical estimates for stock assessment can be produced from detailed sample data in a transparent manner by 2022;
- Incorporate detailed data on Bycatch and PETS AND/OR Recreational data (to be determined by SC-RDB) in the RDBES by 2023.

The RDBES database and web application/system is now implemented on a test server. The countries can upload data for all sampling schemas (all 13 specified upper hierarchies). A security module has been implemented which ensures that data submitters can only upload data for their own country. The data checks in the RDBES currently ensure that data is in a valid format and that valid codes have been used. The data can be exported in the same format as the uploaded RDBES format.

The RDBES web application will provide certain functionality (Figure 5) such as data uploading, and managing permissions but stock estimation and imputation will be performed in the Transparent Assessment Framework (TAF)

<https://www.ices.dk/marine-data/assessment-tools/Pages/transparent-assessment-framework.aspx>

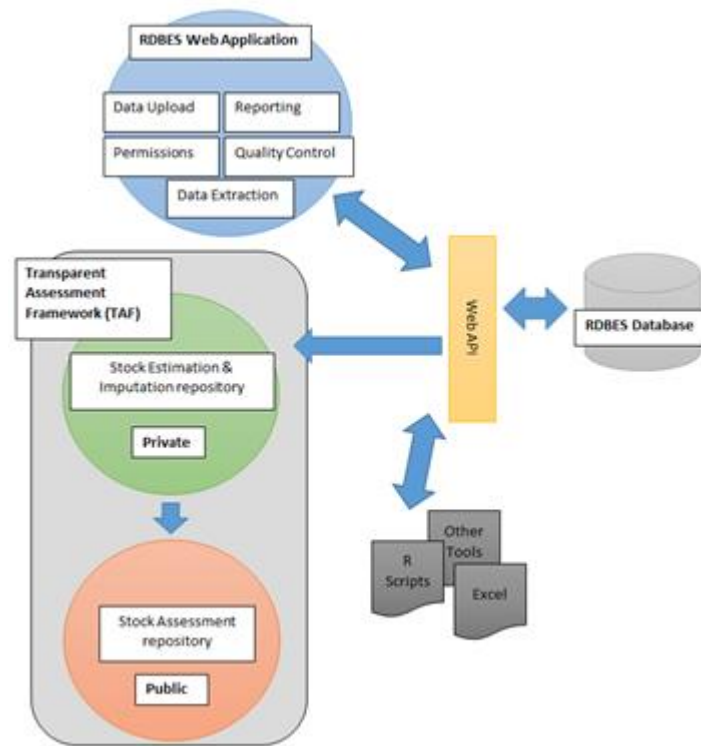


Figure 5: Functionalities of the RDBES web application

There are a number of important RDBES activities planned for 2020.

Date	Name	Chairs/responsible
31 January 2020	Publish draft RDBES Data Call	Henrik Kjems-Nielsen
2 – 5 June 2020	The second workshop on Populating the RDBES data model (WKRDB-POP2)	David Currie and Edvin Fuglebakk
31 July 2020	Release RDBES Test Data Call	Henrik Kjems-Nielsen
14 – 18 September 2020	The second workshop on Estimation with the RDBES data model (WKRDB-EST2)	Nuno Prista and Kirsten Birch Håkansson
30 September 2020	RDBES Test Data Call Deadline	Henrik Kjems-Nielsen
16 – 20 November 2020	The workshop to migrate current national raising procedures to data in the new RDBES format, and using TAF (WKRDB-RAISE&TAF)	Laurent Dubroca and TBC
1 – 3 December 2020	SCRDB Meeting (New name: WGRDBESGOV)	David Currie and Katja Ringdahl
TBC	The workshop to develop ratio estimation scripts using the RDBES format (WKRATIO)	WGCATCH / Laurent Dubroca and Liz Clarke

It is important to remember that the ultimate success of the RDBES will rely on the effort and contributions from a large number of people in the wider ICES/EU data collection community and not just the relatively small groups who attend the SCRDB or Core Group meetings.

Data uploading, data handling and data checks

The Joint Research Centre of the European Commission (JRC) supports the implementation of the EU Data Collection Framework (DCF) managing 5 data calls: 3 socio-economic data calls (Fishing Fleet, Aquaculture, Fish Processing Industry data calls), the Mediterranean & Black Sea data call, and the Fisheries Dependent Information (FDI) data call.

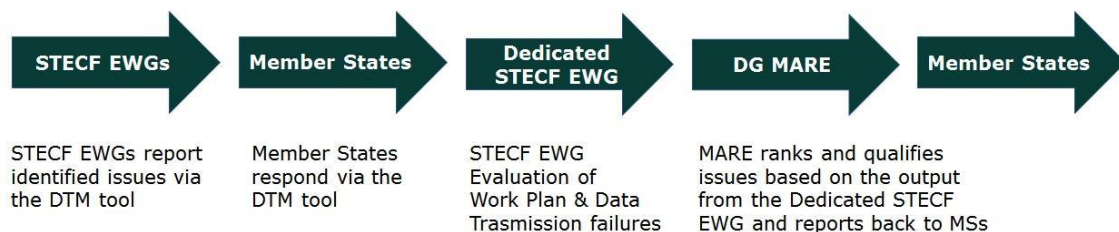
The data handling procedure followed for STECF Expert Working Groups (agreed with DG MARE in 2015) was described (i.e. legal deadline versus operational deadline).

General issues encountered during the management of the data calls were the following: calling data fit for purpose (STECF EWGs, stock assessment, Terms of Reference assigned by DG MARE to EWGs, other stakeholders and the general public); coverage and quality of data provided by Member States; data confidentiality.

Data transmission issues encountered prior to and during the STECF EWG meetings are reported via the Data Transmission Monitoring Tool (DTMT).

The purpose of the DTMT is to make all users aware of issues with the data transmitted by Member States in response to official data calls under the DCF. Any issue entered in the tool requires Member States to provide an explanation as to why the issue has arisen. In this tool each issue is recorded indicating issue type (coverage, timeliness, or quality) and the severity level (high, medium, or low). Recommendation to the STECF experts is to focus on the biggest issues, avoiding to list hundreds of low-severity issues.

The DTMT workflow is shown in the following graph:



Some automated checks are performed prior to and during the data uploading step (syntactic checks: acceptable codes, value types and ranges, coherence among columns; duplication checks). The Data Validation Tool is available online to check data before the uploading step. The same syntactic checks are performed by the uploading facility before storing data in the database; unsatisfactory submissions are relayed back to Member States along with an error report.

The upload facilities and the DTMT have been implemented (in Java) as Liferay portlets; the DCF data are currently stored in a Postgres DBMS.

After the data are stored in the DCF database, JRC assesses the data coverage of the aggregate data transmitted by MSs and implements statistical checks for detecting potential anomalies in the data and reports such anomalies to the data providers and the STECF EWG concerned.

The main checks performed on the data stored in the DCF database are:

- Consistency between the data submitted and the specification of the data call
- Consistency between the data submitted in the different tables of the data call
- Consistency of the data reported at different levels (e.g. segment, national) over time
- Checks for outliers
- Consistency with other data sources

Each STECF EWG assesses coverage and quality of the data in relation to the Term of References assigned to the EWG. In the STECF EWG reports a chapter outlines the coverage and quality of data. During Mediterranean EWGs, only data about species used for stock assessments carried out during that meetings are checked.

3.2 Identify gaps and needs for statistical and/or tools developments

There is a need to improve education of onboard observers, and especially better accommodate the strong staff turn-over. There would also be the need to consider testing the species recognition ability of an observer before giving the green light to embark on fisheries vessels. The idea developed in PGDATA would be to set up a web application on species recognition and test. The web application should be collaborative among institutes and also national museums.

3.3 Online monitoring of sampling programs (WAO), Ifremer

Code ligne	Programme de rattachement	Strate	Type de campagne	Société	Lieu des sorties	Flottille	Art de l'engin	Classe de taille	Obligation	Fin	T.3 2020	T.4 2020
2020_M0001 <small>MAJ récente</small>	DPMA_20_23_Lot1	LOT1_S1_Trainants_AEP_Espèces_profondes	Standard	OCEDEV	De Dunkerque à Cherbourg inolus	AEP_Espèces_profondes	Trainants	> 30 m	Chalut oblatant les espèces profondes	12/2020	0 / 10	0 / 7
2020_M0002	DPMA_20_23_Lot1	LOT1_S2_Dormants	Standard	OCEDEV	De Dunkerque à Cherbourg inolus	Engins dormants	Dormants			12/2020	0 / 18	0 / 17
2020_M0003	DPMA_20_23_Lot1	LOT1_S3_Trainants_<12m	Standard	OCEDEV	De Dunkerque à Cherbourg inolus	Engins trainants <12m	Trainants	<12m		12/2020	0 / 19	0 / 17
2020_M0004	DPMA_20_23_Lot1	LOT1_S4_Trainants_12-18m	Standard	OCEDEV	De Dunkerque à Cherbourg inolus	Engins trainants 12-18m	Trainants	[12-18m]		12/2020	0 / 14	0 / 11

Figure 6: Sampling plan page of the WAO application.

The Web Application for Observers (WAO) was developed by Ifremer to monitor in real-time all observations from at-sea, on-shore and for the Outer most Regions catch assessment surveys. Unlike most of the computer developments carried out in fisheries, which address the management of activity and biological data, WAO is a software for the administrative management of observations. It allows the real-time management of the companies whose mission is to carry out the observations. From a data quality point of view, it allows to disseminate the sampling plan to the different actors, to guarantee its functioning (in particular the question of random draw from a sampling frame) and to adjust the observation effort in real time to optimise the realisation rate (Figure 7).

Before each observation, the observer needs to create a contact (vessel, auction visit, port sampling), which is directly linked to the sampling program and sampling stratum. The observer fills the contact information with the status of the planned observation (observer name, date, sampling location, rendez-vous agreed, refusal and reasons for refusals, ...). Once the observation is done, the contact is updated with its new status (observation realised, date, comments). Once the sample data have been populated in the Ifremer central database, the contact is again updated with the date of synchronising the data in the Ifremer database. The ultimate update of

the contact line is the validation of the data by Ifremer which serves as a proof for the French subcontractors implementing the sampling schemes for receiving their payment.

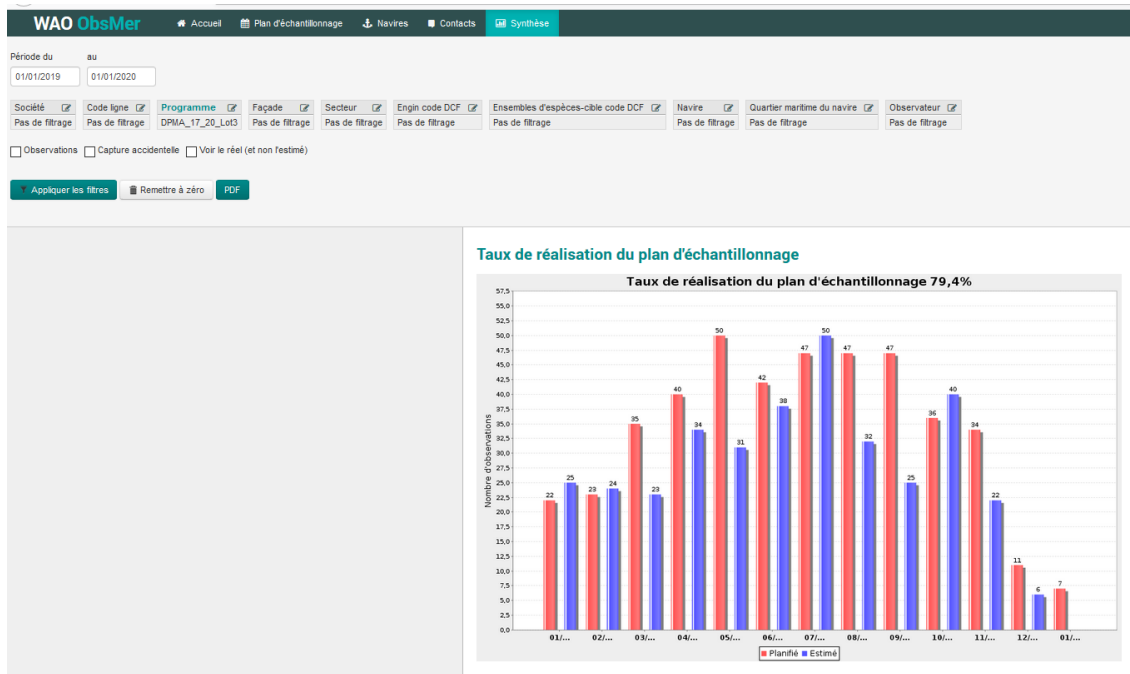


Figure 7: Example of a monthly realisation figure for the observation at-sea (ObsMer) in 2019.

A new version has been developed in 2019 and 2020 to align with the full renovation of the French sampling schemes which started in July 2020. The new sampling scheme applies the 4S principles, backed by PGDATA. One of the new features is the random draw of a vessel when creating a contact for at-sea sampling, and the random draw of species to sample when creating a contact for on-shore sampling. The random draw of species is based on a species frame containing sampling probabilities, which allow Ifremer to adapt any situation on the field (better complementarity with at-sea sampling, high refusal rates in a fishery, bad weather for at-sea observations, species poorly sampled at sea, specific needs, ...).

3.4 Discussion and Summary

There is a large amount of activity in the ICES world focussing on data needs for assessment and advice. One of the major benefits of having a large number of expert groups, organisations, and individuals participating in this process is the high level of innovation displayed. However, the downsides of this can include a lack of knowledge about what other work is being done by other people and a lack of coordination in harnessing this work. PGDATA is one of the few places where a holistic view of this work has been taken and it is important to ensure this continues in the future.

4 Communication and feedbacks on data issues

4.1 Review and comment on ICES data call

PGDATA reviewed and commented both ICES Fisheries and RDBES data calls of 2020. PGDATA has for several years made significant contribution to the ICES data call for joint fisheries. In 2019 and 2020, PGDATA proposed clarification in the text of the call detailing how to handle BMS and registered discards.

The ICES Stock Information Database (SID) <http://ices.dk/data/assessment-tools/Pages/stock-information-database.aspx> has been developed since 2018 and was implemented in 2019, this allows the knowledge in advance on which data will be required in the next data call. PGDATA requested the possibility for stock submitters to select multiple AWG (or all) and multiple stocks (or all), since accessing these one by one is anticipated to be extremely tedious. ICES proposed to test the system.

PGDATA pointed the need to have a feedback on which data has been formally received by ICES. The proposal is for ICES to be able to update the SID with the information transmitted by submitters, e.g. colour code or fixed list of code (S = Data submitted). The current SID does not contain the status of data transmitted as originally planned, so this collaborative platform between end-user and data submitters needs to be developed further (PGDATA future)

4.2 Publication and communication on findings

A theme session was convened in 2016 on data quality (enough is enough), which attracted a lot of participants, however, the presentations made during this session were not summarised in a publication. PGDATA proposed a new session for ICES ASC 2020 to monitor the progress made and summarise the findings in a publication. The idea was to use the same name as during the 2016 ASC to ensure a continuity of work: *“When is enough, enough? Methods for optimising, evaluating, and prioritising of marine data collection.”* The objective was also to edit a special issue in a journal (e.g. ICES JMS).

PGDATA highlighted the potential for editing a CRR on

- ICES QAF

Guidelines and good practices (better living documents) - PGDATA also highlighted the fact that a CRR on age determination proved difficult to finalise after several years in the making.

Other means of publication should also be investigated:

- Publications in a journal
- Presentations in the observer conference (Adelaide, 2020)
- Data doi – good practice? Guidelines from DIG

PGDATA could not progress on these issues during this current 3-year programme. This information is here for the record, to be addressed in the future.

4.3 Data Confidentiality

PGDATA notes that in recent years there has been an increased awareness of the importance of protecting confidentiality whilst still maintaining data quality and ensuring a common approach across countries. To a large extent this was driven by the introduction of the EU General Data Protection Regulation (GDPR) which became enforceable from 25 May 2018.

Confidentiality issues linked to the specificities of the fisheries data and GDPR were discussed during the 2018 RCG meetings in respect to achieve a common approach across EU Member States. A common text summarising the issues was developed and discussed during the 2018 Liaison Meeting in cooperation with DGMARE and was sent to the respective Member States through the national correspondents for them to comment upon. Concurrently the RCGs end-user subgroup started a dialogue with ICES on how to provide data in a uniform way for use within ICES whilst respecting the confidentiality issues. The RCG end-user sub-group provided a report in RCG meetings in 2019, and PGDATA followed this up in its 2020 meeting. The main issues of the discussions were:

Data providers should not suppress any data themselves

1. If the data request defines that data should be pre-aggregated (e.g. VMS data aggregated to c-square level) then the data providers should be requested to supply the number of unique “individuals” in each aggregated unit (e.g. the number of distinct vessels per c-square). If the data request asks for data on a detailed level (e.g. individual fishing trips) then this is not necessary
2. The data will then be sent to the data requester (e.g. ICES) in a secure manner and stored in a secure location with restricted access. The authorised end user (e.g. a working group) will then be given access to the data and can use it for the agreed purpose. The data should be deleted once this purpose is completed.
3. Publication of this data (including maps/charts/tables derived from that data) must use one of the following techniques:
 - a) **Suppression:** Suppress any data that does not include at least 3 different individuals. Suppression can either be done by suppressing the unit or publishing the unit but suppressing the sensitive values (e.g. effort, value). If suppressing sensitive values, then care must be taken to ensure any published totals can't easily be used to infer the suppressed value (e.g. if the value of a single unit is suppressed but the total value is also published then the suppressed value could easily be calculated).
 - b) **Aggregation:** Aggregate the data (spatially, temporally or both) such that each aggregation contains at least 3 different individuals. After aggregation if there are aggregated units that still contain less than 3 individuals then another level of aggregation can be applied, or those aggregated units should be suppressed

When aggregating data, it is not always possible to simply add up the number of distinct individuals in the underlying data to calculate the number of unique individuals in the aggregated unit – this is the case when the same individuals can be present in a number of the original units. An example would be temporally aggregating VMS data to an annual basis when it was originally supplied on a monthly level – the aggregator will not know whether the same vessel was active for all 12 months, or whether there were 12 different vessels active. In this example the data could also be aggregated by country (assuming that individuals can only have a single country per aggregated unit) such that each annual aggregation must contain data from at least 3 vessels from the same country, or data from at least 3 different countries, or both.

Multiple different aggregations of the same data should also not be published since it might inadvertently reveal the confidential data.

The rules which have been agreed to maintain the confidentiality of the RDBES data are specified in its Data Policy (https://www.ices.dk/marine-data/Documents/Data_Policy_RDB.pdf). In line with PGDATA recommendations, at the 2019 SC-RDB meeting a “Conditions for use of detailed RDBES data” document was written that members of ICES expert groups that use RDBES detailed data will be required to sign – this document states that they understand and will follow the rules of the RDBES Data Policy. This document will be reviewed by the RCGs and National Correspondents.

5 Future of PGDATA

5.1 Lessons from the 2018–2020 period

Attendance:

- High quality of the participants (experts with responsibilities on fisheries sampling and fisheries data analysis in their MS, institutes and/or international forum like ICES, STECF, RCG).
- It would have been beneficial if EOSG chair or appointed representative as a participant
- Key chairs or appointed representatives (e.g. WGCATCH, WGBIOP, WG dealing with survey data) would also have needed to participate each year.
- Total absence of fisheries independent survey experts. This is a real issue, although WGSDAA seems to have the same mandate.
- PGDATA has struggled to attract enough members, particularly from the survey world.

Scope, mandate, terms of reference

- Good to have a forum to discuss general issues and approaches to QA, rather than getting bogged down in details or specifics
- QAF based on European Statistical Standard - delivered the structure but no follow-up of its implementation
- Communication with AWG - found a common ground for communicating but not really communicated
- Accessibility - proposed a way forward, ICES will consider these in their developments
- Consistency of approaches in the FD and FI world - Idea of SISPs to document sampling design
- PGDATA had a good connection with the external world (STECF, RCG)
- PGDATA reviewed the data calls, less work now that the data calls have matured, only adjustments for the new variables
- PGDATA was recipient of several recommendations to address and/or reroute
- Did not touch on data collection of new data

Position in the EOSG galaxy, ICES universe

- How visible are we? Are we in a data ghetto? Should we change the name?
- Does PGDATA comprehend all the EOSG WG around?
- Communications between data collection groups and assessment groups has been difficult – do many people know what we do?

New developments of the ICES QAF and TAF

- PGDATA has proposed the structure, now taken by ICES, which needs to continue and start implementing the QAF and TAF (translate the theory to practical) - PGDATA is relevant for accompanying the development of the QAF/TAF in the data collection world.
- PGDATA proposes an ICES QAF but quality frameworks have also been proposed by others within ICES – different groups are siloed.
 - Joint ACOM/SCICOM sub-group in 2019 was a useful first step to integrate these.

New developments of RDBES

- PGDATA does not have a key role here, but may help in reviewing the developments of quality indicators

Risk of discontinuing PGDATA

- Who would point out the gaps?
- The initiatives taken for the last 3-years would be discontinued;
- Even if PGDATA did not coordinate the initiatives of different WG, this is still missing and needs to be tackled;
- Lose the connection with the external world;

5.2 Proposal for the 2021–2023 program

PGDATA spent a large amount of time in its most recent 3-year work cycle discussing and thinking about the quality assurance of data that is used for assessment and advice, however the concept of “quality” is cross-cutting and should be managed throughout a process. The ICES advisory plan highlights the first priority area for development is “Assuring Quality” - it states that quality assurance “encompasses the entire process from data collection to the publication of objective and independent advice”¹. There is a recognition within ICES of the need for an end-to-end **quality management system (QMS)** to encompass best practice in data management, data integration, and translation into advice. A QMS is defined as “...a formalized system that documents processes, procedures, and responsibilities for achieving quality policies and objectives”². Typically, a QMS will follow a particular framework (such as ISO 9001:2015³ or similar) that describes a generic approach to quality management.

Quality management systems should address an organisation’s unique needs however common elements include⁴:

- The organisation’s quality policy and quality objectives,
- A quality manual – this is a high-level document that describes the QMS. It can be used both internally (e.g. as a training resource) and externally (e.g. by advice recipients and stake-holders),
- Documented procedures,
- Data management,
- Measurements of customer satisfaction of output product quality,
- Identification of opportunities to improve.

There has been a large amount of activity dedicated to quality within ICES and whilst it is positive to see so much interest in improving the quality of data and advice it can be hard to manage the application and effectiveness of these different initiatives.

Some examples of these different quality initiatives are:

- the joint ACOM/SCICOM sub-group to develop a road-map towards a quality assurance framework for ICES advice;

¹ https://issuu.com/icesdk/docs/ices_advisory_plan

² <https://asq.org/quality-resources/quality-management-system>

³ <https://asq.org/quality-resources/iso-9001>

⁴ <https://asq.org/quality-resources/quality-management-system>

- PGData proposed a Quality Assurance Framework for ICES which was based on the European Statistical Standard Code of Practice⁵;
- the ICES Data Centre are in the process of applying for CoreTrustSeal accreditation⁶;
- the ICES Data and Information Group (DIG) have developed best practices for data management⁷;
- the quality assurance process for advice;
- PGCCDBS data quality assurance repository⁸ and WGCATCH repository⁹;
- specific workshops have been held to consider specific issues e.g. WKFORBIAS;
- technical solutions have been developed to improve the transparency and reproducibility of the assessment process such as the RDBES and TAF.

PGDATA has previously proposed a number of interesting tools and processes to improve the data feeding the assessment and advice process. With the new RDBES/TAF system becoming fully operational over this next work cycle this is an ideal time to embed these within the workflow. To this end the next 3-year cycle should also ensure that these ideas are operationalised.

PGDATA proposes to change name and scope of its mandate and the creation of the “Governance Group on Quality Management of Data and Advice” (WGQuality). Draft ToRs for the WGQuality 3-year cycle of work are:

- a) Document an operational ICES quality management system for advice that is in line with the scope and direction in the advice plan. This should collate existing quality management tools, resources, and processes related to advice into a single, coherent system. It must also enable ICES to use the best available science in its advice outputs.
- b) Create and implement a communication plan to tell people about the quality management system and ensure there are feedback mechanisms to allow improvements to be identified. Highlight and celebrate good practice.
- c) Use the quality management system to evaluate current activities. Identify gaps and create a plan to fill them. Prioritise issues, identify unnecessary duplication of activities, and propose remedies.
- d) Operationalise the quality tools and processes that were proposed during the previous 3-year cycle of PGData.

Specific tasks within these ToRs could include:

1. Collate existing policies that relate to the quality of ICES advice and identify any gaps (ToR a).
2. Draft an ICES quality manual for the advisory process– the purpose of the manual is to document the overall approach to quality management of advice within ICES (ToR a).
3. Identify the types of generic processes within ICES that contribute to advice outputs (ToR a).
4. Define what documentation is needed for the processes that contribute to ICES advice (such as process flows, standard operating procedures, guidelines, and manuals). Agree who will need to complete the documentation e.g. the benchmark group (ToR a).
- v. Track and review the documentation. Propose tools such as standard templates when required (ToR a).

⁵<https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7>

⁶ <https://www.coretrustseal.org/>

⁷ <http://doi.org/10.17895/ices.pub.4889>

⁸ <https://www.ices.dk/community/Pages/PGCCDBS-doc-repository.aspx>

⁹ <https://www.ices.dk/community/groups/SiteAssets/WGCATCH-publications.aspx>

- vi. Identify new tools or processes that can fill identified gaps (ToR c)
- vii. Define how the proposed “Series of ICES Sampling Protocols” template fits into the RDBES and encourage countries to start populating them (ToR d).
- viii. Create a repository of useful R data quality, scripts, graphs and function that can be used within the RDBES/TAF. Promote the repository. Design processes that will allow people to contribute to this repository. Agree how it will be maintained (ToR d).
- ix. Review draft ICES data calls and give feedback (ToR d).

Proposed date for WGQUALITY: 19–22 January 2021

Proposed chair: David Currie (Marine Institute, Ireland)

Proposed venue: tbd

Annex 1: List of participants

Names	Country/Organisation	Email	Participation in meetings		
			2018	2019	2020
Joël Vigneau (chair)	France/Ifremer	jvigneau@ifremer.fr	x	x	x
Laurent Dubroca	France/Ifremer	ldubroca@ifremer.fr	x		
Josefina Teruel	Spain/IEO	josefina.teruel@ieo.es	x		
Tiit Raid	Estonia/MI	tiit.raid@gmail.com	x	x	x
Christoph Stransky	Germany/Thunen	christoph.stransky@thuenen.de	x	x	x
Rui Catarino	ICES	rui.catarino@ices.dk	x	x	x
Perttu Rantanen	Finland/LUKE	Perttu.Rantanen@luke.fi	x		
Maksims Korvsars	Latvia/BIOR	Maksims.Kovsars@bior.lv	x	x	x
Edvin Fuggelbak	Norway/IMR	edvin.fuglebakk@imr.no	x		
Chun Chen	Netherlands/WUR	chun.chen@wur.nl	x		
David Currie	Ireland/MI	David.Currie@Marine.ie	x	x	x
Juka Poni	Finland/LUKE	jukka.ponni@luke.fi	x		
Jon Helge Vølstadt	Norway/IMR	jon.helge.voelstad@hi.no	x	x	x
Marie Storr-Paulsen	Denmark/DTU AQUA	mSP@dtu.aqua.dk	x		x
Sieto Verver	Netherlands/WUR	Sieto.verver@wur.nl		x	x
Els Torreële	Belgium/ILVO	Els.torreële@ilvo.vlaanderen.be			x

Annex 2: Resolutions

PGDATA – Planning Group on Data Needs for Assessments and Advice

2017/2/EOSG12

A **Planning Group on Data Needs for Assessments and Advice** (PGDATA), chaired by Joël Vigneau, France, will work on ToRs and generate deliverables as listed in the Table below.

	MEETING DATES	VENUE	REPORTING DETAILS	COMMENTS (CHANGE IN CHAIR, ETC.)
Year 2018	13-16 February	Ifremer Nantes, France	Interim report by 2 April 2018 to SCICOM, ACOM, EOSG	
Year 2019	15-18 January	ICES HQ, Copenhagen, Denmark	Interim report by 1 March 2019 to SCICOM, ACOM, EOSG	
Year 2020	21-24 January	ICES HQ, Copenhagen, Denmark	Final report by April to SCICOM, ACOM, EOSG	Change in chair

ToR descriptors

ToR	DESCRIPTION	BACKGROUND	SCIENCE PLAN CODES	DURATION	EXPECTED DELIVERABLES
a	Implement and maintain Quality Assurance Framework for assessment EGs to evaluate data quality and its impact on assessments	The ACOM/SCICOM assessment and advisory process needs to be based on a better understanding of the impacts of data quality. Build on experience in PGCCDBS, WKPICS, SGPIDS and other EGs; Establish close working with case study benchmark workshops; consult with WGCATCH, WGBIOP, WGISDAA, ICES Data Centre, other relevant SSGIEOM EGs & ACOM.	3.1, 3.2, 3.3	Year 1-3	Proposal of a structured approach for agreement within ICES, including the development of the ICES/RDB for detailed fisheries Development of a 'best practice SISP' for data collection in support of stock assessment. Provision of a service to EOSG expert groups for statistical advice and guidance on sampling design to promote good practice and establish effective 2-way communication.

b	Review the outcomes on methodological procedures and quality estimates from past ICES technical workshops and working groups, and advise on ways forward.	Objective procedures are needed to identify where data quality improvements will have greatest impact on quality of advice. Build links with statistical experts within and external to ICES; establish workshops to develop and test methods. Consult with the intergrated assessment working groups. Many ICES groups have provided valid information on best practice and guidelines but it is very time consuming to locate the relevant information on how to improve the data quality for a specific domain. Therefore, PGDATA should locate and host a repository where the information on best practice and guidelines are available	3.1, 3.2, 3.3	Year 1-3	Organisation of a better accessibility to any recommendation or good practice provided by the variety of technical workshops, and including work done by other for a such as STECF and EU-MAP Identification of gaps and needs for statistical and/or tools developments Initiate workshops where needed
c	Propose ways to improve the communication and feedbacks on data issues		3.1, 3.2, 3.3	Year 1 – 3	Direct input in the same years data call in cooperation with ICES sec. Publication on findings (in the form of peer-reviewed publication (e.g. CRR) documenting the development of methodologies in the field of data collection)

Summary of the Work Plan

Year 1-3	Consolidate 3-year workplan; establish membership & skills needed; consultation within SSGIEOM on broader QAF implementation (e.g. surveys); establish links and working procedures with ICES EGs, ICES Data Centre, external bodies, external experts; develop draft QAF guidelines for benchmarks; work with test case benchmark5. Specific ToRs for the plenary meeting will be to:
----------	--

-
- a) Implement and maintain Quality Assurance Framework for assessment EGs to evaluate data quality and its impact on assessments;
 - i) Propose a structured approach for agreement within ICES, including the development of the ICES/RDB for detailed fisheries data, and develop a “best practice SISP” for data collection in support of stock assessment;
 - ii) Collaborate with EOSG expert groups to identify problems and prioritize actions to progress and improve quality data collection.
 - iii) Provide a service to EOSG expert groups for statistical advice and guidance on sampling design to promote good practice seeking to establish effective two-way communication.
 - iv) Cooperate with assessment expert groups to show and demonstrate the effects of data collection methodology on the advisory assessments to underline the relevance of good practice to the advisory process.
 - b) Review the outcomes on methodological procedures and quality estimates from past ICES technical workshops and working groups and advise on ways forward.
 - i) Maintain knowledge of the work done and organize accessibility to any recommendation or good practice provided by the variety of technical workshops and propose changes to SISP as necessary
 - ii) Review the work done in other fora such as STECF and EU-MAP in order to integrate the initiatives and propose complementary work;
 - iii) Identify gaps and needs for statistical and/or tools developments, and initiate workshops as needed;
 - c) Propose ways to improve the communication and feedbacks on data issues
 - i) Review and comment on ICES data calls
 - ii) Organize participation to end-user meetings to seek for mutually beneficial improvements
 - iii) Promote publication on findings, likely in the form of peer-reviewed publication (e.g. CRR) that documents the development of methodologies in the field of data collection and the state of scientific knowledge on the topic at the end of the 3-year TOR period

Year 3	Review of progress / results in implementing QAF; further implementation in benchmarks; Methodological Workshop – developing and testing criteria for evaluating data needs and requests; consultations with end users on data needs; 3rd PG meeting; evaluate future PGDATA workplans.
--------	---

Supporting information

Priority	<p>The focus should be made on the development of the QAF for both fishery dependent and fishery independent data, and on creating links between the different expert groups. The statistical improvements and good practices should be put in context, promoted for implementation, and easily accessible to the public.</p> <ul style="list-style-type: none"> i) Design a Quality Assurance Framework to ensure that information on data quality is adequately documented and applied in assessments; ii) Ensure consistency of approach for fishery dependent and fishery independent data quality framework, and complementarity with approaches developed in other fora such as STECF, EU-MAP, ...; iii) Identify improvements in data quality, or collections of new data, that have the greatest impacts on the quality of advice; iv) Improve or create communication routes between data collectors, data managers and end-users, and advise on new approaches to ease the implementation of the QAF (through publication, RDB-development and cooperation with other WG including shared workshops);. <p>The terms of references should focus on methods and their evaluation rather than providing solutions to a specific data issue or recommending a single method to</p>
----------	--

	<p>be used in all cases. The reason for this is that many assessments and data collections follow different methodologies and have different assumptions so that a universal answer is unlikely to be appropriate. The aim is to gather the existing information on data quality in a structured way, develop expertise and tools where gaps are identified, develop communication with end-users, and maintain knowledge of the work done.</p>
Resource requirements	<p>The national science programmes which provide the main input to this group are already underway, and will need to commit resources to support participation of staff in the PG. Due to relevance of the PG to fishery management under the CFP and to the DC-MAP, use of national EMFF funds to co-finance involvement in the PG should be agreed as eligible.</p>
Participants	<p>The core PG participation required to address annual work plans and plenary meetings will require experts in statistics, sampling design, surveys, modelling, stock assessment, management strategy evaluation methods and other modelling approaches needed, DC-MAP implementation; and RCGs, and efforts are needed to ensure participation of experts directly involved in specific work areas, such as regional benchmark processes, which are being addressed. Other experts, including external experts from USA and elsewhere will be invited when required. EC DG-MARE involvement will be beneficial. A broader pool of experts and other national scientists will be identified for participation in meetings and workshops and to facilitate two-way communication between PGDATA and national institutes.</p>
Secretariat facilities	<p>Support needed from Secretariat involved in setting up benchmarks meetings</p>
Financial	<p>No financial implications.</p>
Linkages to ACOM and groups under ACOM	<p>This is a joint ACOM-SCICOM Expert group. There will be strong and direct linkages with ACOM and with assessment EGs involved in regional benchmarks targeted for case studies.</p>
Linkages to other committees or groups	<p>There will be a very close working relationship with all the groups of EOSG and with ACOM benchmarking groups.</p>
Linkages to other organizations	<p>There will be linkages with STECF, RCMs/RCGs; stakeholder Advisory Committee European Commission and other RFMOs</p>