

Report of the Data and Information Group (DIG)

21-23 May 2019
ICES Headquarters, Copenhagen, Denmark



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.

<https://doi.org/10.17895/ices.pub.8391>

© 2019 International Council for the Exploration of the Sea

Contents

Executive summary	1
1 Opening of the meeting	2
2 Adoption of the agenda.....	3
2.1 Review of actions and previous recommendations.....	3
2.1.1 Recommendations to DIG.....	3
2.1.2 Recommendations made by DIG	4
2.1.3 Actions	5
3 Data Governance.....	8
3.1 Governance evaluation	8
3.2 Feedback from governance groups	8
3.2.1 Benefits	9
3.2.2 Points of attention	9
3.3 New governance structures identified.....	9
3.4 New governance evaluations.....	10
3.5 Governance summary	11
4 Data Policy Review	12
4.1 ICES Data Policy.....	12
4.2 Conditions for VMS Data Use.....	12
4.3 RDB and RDBES Data Policy	12
5 Activities relevant to DIG.....	13
5.1 ICES Annual Science Conference 2018 – Data’s Den	13
5.2 WKSEATEC 2.....	13
5.3 OpenSeaLab 19	13
5.4 OSPAR ICG-DATA.....	14
5.5 IODE/UN Decade of Ocean Science	14
5.6 Dark Data work in WGZE.....	14
6 Data Centre Activity Update.....	15
6.1 New ICES Data portal.....	15
6.2 Transparent Assessment Framework (TAF)	15
6.3 Acoustic data portal.....	16
6.4 VISA: Web based Single stock advice.....	16
6.5 Demonstration of ICES use of GitHub	17
7 ICES Data Centre Accreditation	18
7.1 Perceived quality of accreditation	18
7.2 Accreditation response time.....	19
7.3 Accreditation transparency.....	19
7.4 Accreditation fees	19
7.5 DIG decision	19
8 Tracking challenges and opportunities.....	21
8.1 Machine learning and automation	22
8.2 Cloud and remote services	22
8.3 Open data and code.....	22
8.4 Quality assurance.....	22
9 Data guidelines and standards	23
10 New guidance on reference data	25
11 Best practise for Data Management Handbook.....	26
12 ICES Strategy, DIG, and alignment of activities	27
12.1 Strategic plans.....	27
12.2 DIG role and function.....	27
12.3 Engagement with ICES community.....	28
12.4 Review of terms of references.....	28

12.5	DIG Chair	28
Annex 1:	List of participants.....	29
Annex 2:	Agenda	31
Annex 3:	DIG terms of reference for the next meeting	34
Annex 4:	Recommendations	36
Annex 5:	Actions	37
Annex 6:	A semantically-enabled summary of the Marine Institute’s commercial sampling data	39
Annex 7:	DIG Data Governance Questions	41
Annex 8:	Future Challenges and Opportunities	44
Annex 9:	Connectivity of DIG	48

Executive summary

The Data and Information Group (DIG) met in ICES Headquarters, Copenhagen, Denmark 21–23 May 2019. 16 members from 9 countries took part along with 10 members of the ICES Data Centre.

DIG has progressed evaluation and facilitation of data governance work over the last few years. There are now functioning governance groups for the DATRAS data portal and the Smartdots application and database package. In the past year, DIG evaluated data governance for the Transparent Assessment Framework (TAF), which has helped identify and prioritise tasks in a rapidly expanding application framework. Two new candidate governance groups are identified (TAF and Acoustic). For other data systems, existing expert groups will collaborate with DIG to perform at least 3 further evaluations. Overall, this will bring a broader suite of data into a common approach to address strengths, weaknesses, priorities and routes for improvements.

ICES Data Centre presented on a range of current activities, which are all realising the benefits from the enhanced interoperability of data in ICES via web services. The work demonstrated are setting standards in their fields (e.g. acoustic portal), adopting new techniques to benefit from services (new data portal), streamlining core advice (TAF), and bringing the full value chain of data and advice to the stakeholders (VISA).

DIG established a method for providing better information about the status of various data guidance in ICES. It is starting with Data Type Guidelines, WGFAST data formats, and new guidance being developed by DIG on reference data management. The approach will include signposting of the maintenance status, links to the formally published documents by ICES, as well as more dynamic development versions of guidance. DIG will transition existing data type guidelines to other expert groups where possible. The UK Marine Environment Data and Information Network volunteered to collaborate with ICES on maintaining a number of the guidelines.

Activity updates from members highlighted the widening of the contact network in DIG, and dialogue with the chairs of SCICOM and ACOM about the new ICES strategy, Science Plan and Advice Plan indicated that DIG is working on improvements for managing data in the right areas of priority. The overarching activities of DIG are being brought together in the tracking of broad topics of challenges and opportunities related to data and technology for ICES. DIG identified 15 broad areas and evaluated their potential to disrupt (for good or bad).

With the decision to progress accreditation for the ICES Data Centre, DIG evaluated different accreditation options. DIG decided that the Core Trust Seal will be the most approachable accreditation model for ICES – but that IODE accreditation would likely follow on in light of ICES status as an Associated Data Unit. The core trust seal was selected primarily for its multidisciplinary and transparent structure.

The steps towards accreditation is a positive, albeit resource-demanding exercise, and both ICES Data Centre and DIG recognised that the work done on data governance in previous and coming years will assist in providing relevant evidence for this process.

1 Opening of the meeting

The Data and Information Group (DIG) met in ICES headquarters, Copenhagen, Denmark 21–23 May 2019.

16 members, representing 9 different countries took part in the meeting. A further 10 members of the ICES Data Centre attended (parts of) the meeting. DIG was also joined by one member of WGSMA on the first day of the meeting to discuss governance.

2 Adoption of the agenda

The agenda for the DIG meeting was adopted without major revisions.

The terms of reference for DIG were:

- a) Review priorities in the ICES Data Centre
- b) Provide guidance and feedback to the ICES Data Centre
- c) Advise on data regulations and their impact on ICES Data Strategy and ICES Data Policy
- d) Propose ad-hoc groups (governance, workshops, training etc.) related to specific topics and/or datasets, to facilitate improvements related to data issues, to SCICOM, ACOM, SCICOM SSG and/or EGs, and review the outcome of those ad-hoc groups.
- e) Evaluate and monitor future challenges and opportunities in data management and new technologies for ICES

DIG will report to SCICOM at the next SCICOM meeting during the 2019 Annual Science Conference, and the SCICOM mid-term meeting in March 2020.

The full agenda is available in Annex 2.

2.1 Review of actions and previous recommendations

2.1.1 Recommendations to DIG

On the whole, DIG does not receive many recommendations for data-specific work. Given the more strategic remit of the group, this makes sense. Most/all recommendations relating to data related activities tend to be directed to the ICES Data Centre. When questions arise relating to recommendations, the ICES Data Centre will raise these with DIG rather than delaying the process by first sending a recommendation to DIG. However, four recommendations were received between the DIG meetings in 2018 and 2019.

Year	Expert group	Recommendation	Recipient(s)
2018	WGRFS	A database that brings together estimates of marine recreational fisheries catches for end users is needed as a matter of urgency. A paper that summarises the key issues and proposed solution to include recreational catches in the RDBES is provided in Annex 7 of the WGRFS 2018 report. Support is needed from ICES to resolve this issue, agree timescales, and put a solution in place for 2019.	DIG, Data Centre, SCRDB
2018	WGPME	Establishment of an online reference repository for Lugol-fixed plankton imagery. Advice is sought on how best to implement this in order to make it available to the ICES community and other stakeholders.	DIG
2018	WKDATR-NSCS	Inclusion of DATRAS surveys in official data calls	DIG, WGDG
2018	WKFATHOM2	T-S data from the MEGS should be included into the egg and larvae database	Data Centre, DIG

All recommendations were discussed during the DIG meeting, either as part of the main plenary session on actions and recommendations, or in more detail during breakout and discussion sessions.

The recommendation on the creation of a database for recreational fisheries, or inclusion in RDBES from WGRFS has been handled by SCRDB, who have included consideration of this in the RDBES plan.

The request for advice on establishing a reference image database for plankton imagery by WGPME will be followed up by the DIG chair to ensure the steps listed in the best practise in data management handbook can be followed (Annex 5, Action #2).

The request of the inclusion of DATRAS surveys in official data calls came from a workshop – WKDATR-NCNS. The chair of the DATRAS governance group (WGDG) also is a member of DIG and was aware of the situation and will bring this to the attention of WGDG.

Finally, the recommendation to include temperature and salinity fields in the eggs and larvae database has already been implemented by the ICES Data Centre at the time of the DIG meeting.

2.1.2 Recommendations made by DIG

DIG also followed up on recommendations made in the previous year:

Recommendation	To
Agree with user community to define a clear transition period where the RDB, InterCatch and the new RDBES system will be operational, and a clear date for when only RDBES system will be utilised.	SCRDB, SCRDB ICES, ICES Data Centre – RDBES Development
DIG recommends that the data requirements and necessary data extractions for WGBYC are included in the functional requirements for RDBES development for consideration. It is recognised that this may be a longer term solution than the initial development, but should still feature as a functional requirement	ICES Data Centre, WGBYC, SCRDB, SCRDB ICES,
DIG recommends that WGZE records all dark data or historical data sources identified as part of their Tor C. Recognising that work has already been done to outline metadata requirements, this should be a relatively straight forward task that will allow identification of the data for future recovery projects. Dialogue with the ICES Data Centre would enable tagging of data identified by WGZE to be easily locatable.	WGZE, ICES Data Centre
DIG recommends the development of a FAQ page concerning the processing of any potential personal data associated with ICES data calls. This will make the position clear both for ICES and national data submitters	ICES Secretariat
DIG recommends that all new or revised data submission formats incorporate the optional ability for national data submitters to include persistent identifiers	ICES Data Centre

The first two recommendations were relating to the development project for RDBES and have been incorporated into the core development team and SCRDBS. Final decisions on whether to implement the recommendations is largely in the hands of RDBES development and WGBYC, but DIG will follow up and correspond as necessary.

DIG also recommended that WGZE's work on dark data could utilise existing facilities in ICES to record metadata on data collections held in various institutes – along the same lines as what has been done by WGHIST. While the records have not yet been created, the update received from WGZE clearly indicated that work on this is progressing.

The recommendation to the secretariat to implement a Frequently Asked Questions page on aspects of personal data has not yet been implemented. However, there are pages that highlight how data are handled in terms of applications or registrations within ICES.

The final recommendation to the ICES Data Centre to incorporate persistent identifiers into data submissions for new or substantially revised systems has been acknowledged, and work on revisions and updates to the accession system will implement this recommendation.

2.1.3 Actions

As an operational group, DIG also utilises actions to keep track of inter-sessional activity. The majority of actions stem from the previous DIG meeting where work ahead is planned. However, actions may also be added through the year as a result of requests from other parts of ICES.

Number	Action	Addressed To	TimeLine	Status
1	Initiate dialogue with WGBIOP to clarify recommendation and understand potential scale of work	Ingeborg de Boois, Neil Holdsworth, Carlos Pinto	Update May 2019	Completed. Follow up actions for DIG and WGDG this year to write more detailed guidance on the effect of changing reference data
2	Continue dialogue with WKSEATEC, ensuring alignment and communication with ICES Data Centre	Marcellus Rödiger, David Currie, Jens Rasmussen	Next WKSEATEC 2018 + update May 2019	Completed, DIG members attended WKSEATEC II. Presentation of workshop at DIG
3	Discuss and decide if a joint hackathon event between EMODNet and ICES (and potentially other partners also) can go ahead	Simon Claus, Neil Holdsworth	End August 2018	Completed. ICES will Participate in joint 2019 Hackathon. OpenSeaLab 2 presented at DIG.
4	Draft a resolution for a new workshop format ICES hackathon, to be submitted if the joint event in Action 3 does not progress	David Currie, Sjur Ringheim Lid	End August 2018	Completed but not put forward due to Action 3 status
5	Include review of VMS/Logbook Data policy in the DIG 2019 agenda/programme of work	Jens Rasmussen	Mar-19	Completed as part of data policy review during DIG meeting.
6	Draft document on new approach to data guides (collection/format/process guides) with emphasis on a flexible structure and the process for revision and review	Peter Wiebe, Taco de Bruin, Hjalte Parner, Neil Holdsworth, Graham Allen, Ingeborg de Boois, Colin Millar, Susanne Tamm	Draft by September 2018, Final Document by Feb 2019	Update was produced for SCICOM – follow up action plan during DIG 2019 (See section 9)
7	Complete a governance framework profiling of TAF	Jens Rasmussen, Christian von Dörrien, Colin Millar, Arni Magnusson	Dec-18	Completed. Presented to TAF team and subsequently to DIG. Follow up activities to maintain and discuss updates of governance evaluations.

Number	Action	Addressed To	TimeLine	Status
8	Develop 2-3 small use cases demonstrating linkage and usability of the new semantically enabled ICES vocabularies to help demonstrate the significance of this change	Simon Claus, Sjur Ringheim Lid, David Currie, Hans Mose Jensen, Graham Allen	Draft cases by September 2018, Final cases by Feb 2019	One use case developed by the Marine Institute, Ireland. Included in Annex 6. This example provides clear information on how the development of the ICES vocabulary server is now enabling the wider community.
9	Review the initial ICES Data Portal Use Case and collate comments	Sjur Ringheim Lid, Simon Claus, Ingeborg de Boois, Taco de Bruin, Graham Allen, Wim Allegaert	End August 2018	Initial comments provided around the time of the ICES ASC. Development has since progressed and was presented to DIG during the meeting.
10	Complete a governance framework profiling of ESAS	Ingeborg de Boois, Neil Holdsworth	Dec-18	This activity requires some clarification of terms of reference for existing group, so have been actioned for the coming year.
11	Set up a ICES SharePoint template for tracking future data challenges and opportunities using a risk matrix style approach	Jens Rasmussen, Neil Holdsworth, Vivian Piil	Sep-18	Completed and reviewed by DIG during meeting.
12	Advertise and encourage submission of pitches to the Data's Den open session during ASC 2018. Specifically contact ICES Comms team, WKIN-VITED, WKMLEARN to encourage participation	Mail Werner, Christian von Dorrien, David Currie	Jul-18	Completed - a total of 6 pitches went ahead
13	Identify judging panel for the Data's Den open session during ASC 2018	Malin Werner, Neil Holdsworth, Christian von Dorrien	Aug-18	Completed. Many thanks to DIG member Peter Wiebe for stepping in last minute as one panel member was unable to attend on short notice.
14	ICES Data Centre to provide a lists of systems with data submissions, detailing the capabilities for traceability/persistent identifiers of national submissions	Neil Holdsworth	Feb-19	Completed. Somewhat overtaken by the events on reworking accession system within ICES. No need for any further activity specifically for this action.
15	Draft a single line reporting format for the outcome of governance framework evaluations	Jens Rasmussen	Dec-18	Completed – graphic visualisation presented at DIG meeting.
16	Compare and map out overlaps between DataCite and INSPIRE metadata schemas	Ruth Lagring	Oct-18	Completed

Number	Action	Addressed To	TimeLine	Status
17	DIG Chair to write to chair of WGML to highlight potential other sources of macro litter data than trawl data	Jens Rasmussen	Sep-18	Completed. Dialogue with WGML indicates this is on their radar, and that they are considering alternative data sources as well

3 Data Governance

Work on implementing structured governance of systems and data collections in ICES has progressed rapidly over the past year.

Governance groups group DATRAS and the Smartdots applications have established their work in identifying priority tasks, evaluating the completeness and maturity of their respective systems, while resolutions for new governance groups have been set up for the Transparent Assessment Framework, and the Acoustic Data Portal.

The chairs of WGDG and WGSMART both attended the section on governance, and provided feedback for the report (see Section 3.2).

3.1 Governance evaluation

DIG has produced a structured governance evaluation template that consists of a number of questions in each of the governance framework categories presented in last year's report (See Annex 7 for full list of questions and categories). This template was used on the Transparent Assessment Framework as a pilot in January 2019. The pilot consisted of an interview around the questions and categories, with a subsequent assessment of each category's maturity along with a number of identified improvements.

The questionnaire template was subsequently shared with the TAF development project team to help prioritise tasks and balance out some of the improvements identified. A scoring system from 1–5 is used to give indicative maturity levels ranging from 1 being completely reactive with no established process or documentation to 5 being fully documented, managed, planned, and where the information is used to optimise performance. The indicative scores are useful to examine the internal consistency across categories to help identify priority areas of work. However, the direct numerical values of scores should be interpreted with caution as consistency between different evaluations still needs to be examined once more have been completed.

DIG was presented with the outcome of the TAF evaluation, and the ICES Data Centre staff working on TAF expressed satisfaction that the exercise had provided them with useful information that have helped prioritise improvements.

DIG is planning to maintain the evaluation templates as living documents [DIG Action#8], and to revisit the evaluations to update information and improvements, which will be used to demonstrate progress and engagement in managing systems and data.

3.2 Feedback from governance groups

In 2018 two governance groups have been installed: the Working group on DATRAS governance (WGDG) and the Working group on SmartDots governance (WGSMART). The annual report of WGDG with the specific achievements will be available soon. The annual report of WGSMART will be available after the group holds their physical meeting in October 2019.

Although both governance groups differ in their nature -WGDG deals with a long existing database, WGSMART with an only recently developed platform combining; an image analysis tool, database and reporting module- the experiences are largely similar.

3.2.1 Benefits

First of all, the close connection between ICES Data Centre (data officers and/or programmers) and experts (submitters, users) from the ICES community is beneficial for both. There is room to explain reasons for new technical developments, and the (often biological) background of recommendations from various groups. Furthermore, group size (6-8 people) is sufficient to take well-informed decisions, even by web conference. The meeting frequency (four times a year max. 2 hours by web conference) is sufficient to keep track of the developments. The governance groups in this way facilitate the development process.

For WGDG specifically the communication on DATRAS developments has become more coherent over the groups as the WGDG representatives of the data submitting expert groups all have the same information, documented in the minutes of the meeting. The feedback from the survey expert groups can be compiled by the governance group, which leads to a more consistent advice to the ICES Data Centre. Also, the technical aspects of any change can be discussed in the governance group. This leads in the end to well-informed and feasible decisions on changes in DATRAS. Meeting reports are posted on the SharePoint site.

In the case of WGSMA^{RT} a physical meeting is held in connection with the annual meeting of the ICES Working Group on Biological Parameters (WGBIOP). There is a close working relationship between these two groups (with one overlapping chair) and feedback from WGBIOP is considered when planning SmartDots developments. A number of the participants attend both meetings. Given that SmartDots is a relatively new tool and thus to facilitate frequent developments, an annual physical meeting is deemed highly productive for tasks requiring more in depth discussions/decisions or hands on work by a number of members. Communication on developments and the work plan is maintained via the ICES expert groups's SmartDots github site (<https://github.com/ices-eg/SmartDots/projects/3>) and all meeting notes are posted on the SharePoint site.

3.2.2 Points of attention

Since the group size is limited due to the web conference set-up, it is important to recruit people from different backgrounds and prevent redundancy.

It is currently unclear what the process is to discuss and decide on topics that overarch multiple governance groups. It is proposed that the chairs of the governance groups meet annually with the head of data by web conference [Accepted as DIG Action #4] to see if there are any topics that have been dealt with by more than one governance group and of which the outcome should be agreed between the groups and conversely to discuss future prospects for overarching topics. During the meeting the governance group chairs can also exchange other information that may be of relevance for other groups.

There is one person taking part in both WGSMA^{RT} and WGDG. Although this happened accidentally, the groups see the added value of this. It is therefore suggested that the newly proposed governance groups contain one person that is also active in another governance group. Obviously, this person has to add value in itself to both governance groups.

3.3 New governance structures identified

The introduction of a governance framework can benefit new or emerging systems from having built-in methods for prioritising tasks, and having regular contact with a representative group of the user community.

The ICES Data Centre have recognised the benefits of closer collaboration and contact with existing governance groups, and two new groups have been identified to widen the approach. Two resolutions will be submitted to establish governance groups for the Transparent Assessment Framework, and the Acoustics Data portal. DIG welcomed the wider adoption of the approach, and fully supported the initiative.

In addition, it is recognised that for a number of systems, there are already groups that can naturally adopt the governance function without the need for new groups. These include:

- Steering Committee of the Regional Fisheries Database (SCRDB) for the Regional Database (RDB) and Regional Database and Estimation System (RDBES).
- Working group on Bycatch of Protected Species (WGBYC) for the data system for by catch of protected species.
- Working group on Deep-water Ecology for the Vulnerable Marine Ecosystems data
- Working group on Spatial Fisheries Data (WGSFD) for the spatial fisheries data management.

Many of these groups deal with systems that have access limitations or separate data policies, so have often been more directly involved in defining system requirements and needs more directly with the ICES Data Centre than other, broad systems with many different expert groups supplying and using data.

3.4 New governance evaluations

Having piloted the governance evaluation process in the past year, DIG will now progress to evaluate a number of platforms within ICES. It is important to recognise that a governance evaluation is not the equivalent of forming new governance groups. Often there are already established groups that largely oversee and define data standards and requirements of systems or workflows. The evaluation is an approach that spans a little wider to ensure that all considerations of best practise in data management are considered.

This year, DIG will progress a number of governance evaluations:

- Spatial Fisheries Data workflow
- Marine environment database (DOME)
- Vulnerable Marine Ecosystems

Further systems will be examined for feasibility, or the process will be initiated, but might not complete within the year:

- Bird database (ESAS) application
- Bycatch Database

Each governance evaluation will follow a similar structure:

1. Initial evaluation, following the categories and questions
2. Reviewer scoring and identifying broad improvement areas
3. Share initial findings with developers and groups governing the data structure to reach consensus on the state/scoring and identified improvements
4. Governance structure identifies actions to prioritise improvements and takes forward the improvement programme
5. DIG revisits governance evaluation, specifically to see how categories/questions with identified improvements have been progressed (1-3 years later)

3.5 Governance summary

The collaboration between the ICES Data Centre and DIG have refined the processes and terminology around the concept of data governance in the last 2-3 years, culminating in a more formal structure and approach over the previous year.

The progress is now starting to be more tangible, and with proposals for new groups and more evaluations of existing systems progressing, the coming year will bring more alignment and results. The progress is also motivated by the supportive statements received from SCICOM, ACOM, and council at the times when this approach has been presented.

One of the key challenges is to ensure that the way data are acquired, organised, used, shared, and maintained can be reasonably evaluated to define improvements and track progress while keeping in touch with the community. It is important to avoid making governance structures too formulaic as the specific work will vary from group to group and across data collections.

Moving this process from the more theoretical definitions of categories to concrete evaluations and groups dedicating time to improve data processes is a significant step for ICES, and it should be recognised that the ICES Data Centre has played a central role in facilitating this by meeting the process with openness, assisting governance groups with establishing github repositories, and a more detailed dialogue with users. In time, this approach will pay off with more effective and defined solutions as users are clearer on roles, and thus expectations and progress will be on realistic scales.

4 Data Policy Review

DIG routinely performs a review of the ICES Data Policy. This is done to ensure that the data policy reflects current considerations and reflects changes in ways to access or work with data. There are now additional data policies that cover areas where the default open access cannot be provided due to the sensitive or commercial nature of the data being used in certain workflows. These data policies were also reviewed in order to ensure there is alignment and consistency in the use of terminology across the policies.

4.1 ICES Data Policy

DIG reviewed the ICES Data Policy. One of the key areas of discussion was around the appropriateness of the term data policy and what it means compared to data license.

Overall it was recognised that a policy typically sets out a set of principles to guide decisions or achieve outcomes, while a license is a permission to do, use, or own something.

Currently the ICES Data Policy encompasses both aspects, stating the principles in operation for data from ICES, as well outlining the permissions for use and redistribution of data.

Looking ahead, DIG believes a separation of license and data policy will be better and clearer, and will also better align with current practises elsewhere. As a result, DIG will draw up an overview of existing open data licensing models and evaluate their benefits and drawbacks in the ICES context. DIG will also examine if there are potential challenges or conflicts with increased adoption of GitHub with a new data license, and provide an update to SCICOM in March 2020. It is recognised that this can mean some more substantial changes to the existing Data Policy in due course.

In the meantime, some edits for consistency of language was identified in the existing data policy, and these edits will be brought to the attention of the Bureau and/or Council by the ICES Data Centre.

4.2 Conditions for VMS Data Use

The VMS Logbook data policy was reviewed, and some minor suggestions for change of wording were made. The suggestions will be submitted as a recommendation to WGSFD to review and implement.

4.3 RDB and RDBES Data Policy

The RDB and RDBES data policy was generally found to be clear and comprehensible. It was updated just recently (Jan 2018) by SCRDB. DIG found a couple of minor wording issues and a missing reference. The suggested edits will be submitted as a recommendation to SCRDB to review and implement.

5 Activities relevant to DIG

DIG is, via its members, recommendations and requests, involved in a range of inter sessional activities both directly within ICES, and outside ICES. This session of the DIG meeting enabled the wider group to hear about some of these activities, and how they align with the work of DIG (Also see Annex 9 for diagram of connectivity of DIG to other ICES groups and initiatives).

5.1 ICES Annual Science Conference 2018 – Data’s Den

DIG organised an open session event at the ICES ASC 2018 in Hamburg. The event was an informal presentation of pitches for ideas for products, tools, or systems relating to improving how we work with data. Malin Werner presented and summarised the event outcome.

Six participants presented pitches in a highly compressed three-minute format, followed by questions from a judging panel, with subsequent “investment” of sweets from each panel member. Finally the audience vote, based on the volume of applause awarded a further investment to each of the participants.

Overall, there was approximately 120 conference delegates attending the event, and it was felt that this was a very positive way to engage in an otherwise technically complex and varied subject, while also acting as a social event that facilitated discussions.

Many members of DIG were actively involved in the planning and execution of the event, and were thanked for their efforts.

5.2 WKSEATEC 2

DIG members attended the second WKSEATEC workshop (Workshop on Technical Development to Support Fisheries Data Collection) held at ICES, Copenhagen in 2018, and the co-chair of the Workshop, Marcellus Rödiger, presented on work to date. The focus of the workshop is on the data acquisition stage for fisheries data – especially the use of electronic measurement boards or data capture to obtain data during fisheries independent surveys or during commercial catch sampling. The workshops have built up an overview of current solutions in operation or being under development, and have identified opportunities where shared or open development of standards or software could benefit the community. Recommendations to utilise ICES vocabulary servers to assist in developing a common fisheries data language (FDL) was one of the recommendations that would make use of the new semantic capabilities for which DIG have been seeking use cases (in addition to the one included in this report). Looking ahead, the work is looking into the feasibility of either further meetings, or linking up with recently established groups with overlaps, such as WGTIFD (Working group on Technology Integration for Fisheries Dependent Data).

5.3 OpenSeaLab 19

After holding an internal, small-scale hackathon, WKINVITED, in 2018, ICES were presented with an opportunity to take part in a larger, collaborative hackathon hosted by EMODNet during the 2018 DIG meeting. WKINVITED was successful in generating some insights and demonstrations of the positive synergies of putting people with different skillsets together. However, it was also recognised that to make the most successful event, there was a need for significant investment of time, and some financial investment to support such an event.

So when the opportunity to take part in a hackathon of different scale was presented, it was received positively, and the head of the ICES Data Centre has been coordinating the ICES contribution and participation in the OpenSeaLab hackathon event 4-6 September 2019 in Ghent, Belgium (<http://www.opensealab.eu/2019>).

The presentation by Neil Holdsworth, Head of ICES Data Centre, summarised the preparatory work already done and underway. DIG members were encouraged to promote the event and encourage participation from national institutes or local universities (DIG Action#10).

5.4 OSPAR ICG-DATA

DIG received an update from Chris Moulton, OSPAR on the creation of the intersessional correspondence group (ICG-Data). Much like DIG, the OSPAR ICG-Data is established to work on cross-cutting data issues, and to implement the organisational strategy. In the case of OSPAR, there is a specific Data and Information Management Strategy. The chair of DIG (in the capacity of a national representative), and the head of ICES Data Centre are both members of ICG-Data, and it was recognised that there is likely considerably synergies between activities in DIG and in ICG-Data.

There is also direct support for encouraging good data management practises directly for key deliverables, such as the Quality Status Report (QSR2023).

5.5 IODE/UN Decade of Ocean Science

Co-chair of the IODE, Taco de Bruin, provided DIG with an overview of the UN Decade of Ocean Science, and in particular data management considerations. Timelines are now progressing, and the first global meeting and stakeholder event had just recently taken place in Copenhagen, Denmark. This event will be followed by an increasing number of regional workshops that will all feed into the planning process before the official launch of the decade in 2021.

There are specific data management aspects that have already been highlighted in this preparatory phase, such as an emphasis on open data, the integration of social science and human dimensions data, and the promotion of adoptions of standards and interoperability. Many of these priorities align well with ICES activities, and while the delivery side is currently less clear (e.g. “an information system” for sharing and using data), it is clear that the current focus in ICES on enabling good governance and enhancing interoperability will enable active participation in, and contribution to, the decade of ocean science.

5.6 Dark Data work in WGZE

The working group on zooplankton ecology (WGZE) raised the issue of dark data at last year’s DIG meeting. Dark data refers to datasets that are not publically available or fully digitised. In 2018, DIG made a recommendation to WGZE to consider using the ICES metadata catalogue that already has a historical resource set, built up by WGHIST. DIG received an update by correspondence from Peter Wiebe, who was unable to attend DIG this year. WGZE has continued work on identifying sources of dark data, and has started defining relevant metadata utilising ICES vocabularies and structures. The data has not yet been committed to ICES metadata, but progress is clearly being made. In addition, concrete work towards making dark data publically available is also progressing with the help of Todd O’Brien from NOAA (US).

6 Data Centre Activity Update

The ICES Data Centre provides DIG with an activity update. This is reciprocally beneficial in that it provides DIG with up to date knowledge on the priorities of the ICES Data Centre, and there is an opportunity to get input from the various data managers in DIG on best practises, practical experiences, and other initiatives that may be relevant.

This year, the ICES Data Centre provided five presentations, four of which are demonstrating the increasing integration and interoperability of applications in ICES, and one demonstrating the use of a collaborative working environment that opens up possibilities for enhanced collaboration with governance groups and other members of the ICES community.

6.1 New ICES Data portal

Architectural planning and technical design for the new ICES data portal has begun. The new design centres on recognition of an extended number of data types, identified a minimal number of common traits (geography and time), while utilising micro-services from existing data systems to the new portal. The micro service approach is structured around an entity-attribute-value model approach (EAV), which can leverage a lot of different types of data in an efficient manner – especially when well-integrated with the ICES vocabulary services. The approach is going to be tested with two disparate datasets to begin with (acoustics and contaminants).

DIG generally welcomed the approach and recognised the strengths and flexibilities in the micro services approach – allowing a wide range of data and data products to be integrated in the portal over time without disrupting other data flows. DIG volunteered four members to act as beta testers when development progresses [DIG Action #12], and the ICES Data Centre will provide a short progress update for the DIG member newsletter distributed after SCICOM meetings (first time in September 2019) [DIG Action #11]. One final question on closer integration or use of metadata was noted by the Data Centre. While metadata in ICES generally cover very large collections and does not provide highly granular information, there is potentially a need to utilise this in providing information since the flexibility of the EAV model can provide data from multiple sources and contexts in the same results. However, some of these design considerations will be addressed as the project progresses further.

6.2 Transparent Assessment Framework (TAF)

While TAF had been presented previously in DIG, and had already been discussed in the context of governance evaluations and the continued improvement programme, this update provided key information on the increasing degree to which TAF is being used operationally.

The platform has now been formally launched at the FAO-GFCM Fishforum event in 2018, and currently more than 100 stock assessors have used TAF to complete more than 100 assessments.

There has also been recognition that most early adopters are experienced R users who have picked up the approach relatively quickly. However, there are also a number of assessments done in other ways currently, and there is now a clear need to provide training and support to a wider base of users to ensure the successful adoption of all stock assessments performed in ICES. R is one of the training needs, but also an instruction in the use of GitHub repositories is turning out to be an area where more support is needed.

There is also on-going development work to deliver performance improvements, better integration of metadata, and integration with Stock Assessment Graphs (SAG).

DIG recognised that the proposal to establish a governance group for TAF was important, and that only once a governance group is in place can discussion around targets for adoption rates in terms of number of stocks being assessed through TAF. There was also recognition that with the increased adoption, and emerging governance, there is scope for TAF to be utilised for a number of other assessments than purely fish stocks. The flexibility of the framework design is such that other assessments can also use TAF, even in a chain of steps as the platform can now reference data outputs from another TAF assessment. This may be particularly useful for ecosystem assessments, where multiple aspects are brought together.

6.3 Acoustic data portal

The acoustic data portal has been developed as part of the AtlantOS project package, which is completing soon. However, the acoustic data portal is now a core ICES data product that bridges a wide number of expert groups in ICES, both survey and technology groups.

The acoustic data portal is the central component, but has integrated with other developments in the ICES community and beyond. Assessments and data analysis can be performed in both the StoXX and EchoR tools, and it is the standardisation and validation of the acoustic data model that enables this. Validation processes makes use of Schematron validation of a data model that is based on the WGFAST metadata convention and extensive use of the updated ICES vocabulary system. At the same time support is provided for users to submit regular table data if expertise in XML structuring of submission formats is not available on national level. The portal now also includes map based data exploration, and development of the system is now focussing on the inclusion (and validation) of additional surveys, and the integration with TAF to enable data services to support assessments.

A well-defined and standards based data model and validation means that commercial software developers of scientific acoustic analysis are now embedding the ICES acoustic format directly into their software, which in turn will greatly help the community in working up data quicker without having to transform local and national data formats.

Due to the wide range of user expert groups, there is a need for a new governance group to advise the ICES Data Centre on priorities and issues related to the maintenance and development of the portal. While several of the expert groups have the relevant knowledge, it was felt that there was not one of the existing acoustic themed expert groups that would encompass the entire workflow, which is why a new group is proposed rather than making use of an existing expert group.

DIG fully supported this approach, and recognised the widespread impact this work has already had on bringing together ICES expert groups, software suppliers, and developments in national institutes.

6.4 VISA: Web based Single stock advice

In 2018, ICES received a special advice request from the EU to look at disseminating ICES advice beyond traditional pdf files. In the report (<http://doi.org/10.17895/ices.pub.4657>) ICES worked up a prototype development to demonstrate how advice can be made available in a user friendly web format. The format, in addition to presenting the advice together in a well-structured format increases accessibility to information and data that now has the scope to integrate further with other activities in the ICES Data Centre such as an emerging data portal, TAF, and acoustic

data visualisations – all through service exchanges making use of the increase in interoperability between applications.

The VISA prototype development for four stocks demonstrates how a comprehensive set of information can be brought together in a way that offers an easy navigable page with sections that include interactive visualisations and embedded data tables for download. By using tools that are already in active use in the ices community (R and GitHub) there is scope for enhancing and building on these prototypes rapidly. By using R “markdown” the same content can be presented consistently in web friendly formats as well as traditional document form (which may be required by some national or governmental departments for records management) while drawing on the benefits of available services.

A prototype integration of the stock advice in a GIS overlay was also demonstrated and shows great promise in the way the information can be generated dynamically to be used in multiple places or contexts.

Overall, this work is progressing well as a key end-point for many stakeholders and recipients of ICES advice. There is scope for a great deal of integration of data flows in ICES, and the VISA tool will ultimately represent the end point where the right data and information are brought together.

6.5 Demonstration of ICES use of GitHub

ICES organises its project management and development work across three main sections: expert groups, development, and production. Within each of these areas, there are many repositories for different types of projects and development works. The expert group area is often governed or administrated by individual expert groups themselves. Allowing shared development and maintenance of tools enable all members to access the most up to date code used in their workflows and collective improvements, documentation, and issue tracking.

In the ICES development and production repositories, code that are developed and maintained in the ICES Data Centre can be progressed while making use of project management tools, issue tracking, and bringing in either groups or individual experts on specific developments.

The adoption of GitHub as a collaborative tool was already demonstrated earlier in the DIG meeting where the WGSMA²T governance group demonstrated how they make use of GitHub to prioritise the tasks and ideas for developments.

Overall, the use of GitHub offers a very capable platform that tracks versions, changes, and contributions seamlessly, while presenting a lot of options to develop and maintain code in an open and transparent way.

DIG queried how permissions and access to various repositories were managed, and it was recognised that this is a separate process from the current SharePoint registration, so no integrated tracking of individuals. However, this also offers opportunities in terms of ICES community members that are already active on GitHub and would prefer to keep their work together under existing accounts. There are currently no plans for force a single adoption for all projects to be managed on GitHub, but rather an approach of it as a preferred tool.

DIG also queried if there was an overarching policy for which licenses code or repositories were using on the ICES GitHub repositories. Currently this is managed on a case-by-case basis, but an analysis will be carried out to summarise existing license information – as this has linkage to the work being conducted by DIG to look at data licensing as well [DIG Action#17 and #13]

7 ICES Data Centre Accreditation

ICES Data Centre tabled a document outlining the decision to seek formal accreditation as part of work to overall lift clarity about processes, policies, and data quality. Accreditation has previously been discussed in DIG (Section 5.2.3 in 2017 DIG Report) mainly in the context of ensuring that governance evaluations and groups would provide support for future accreditation. Now that formal accreditation will be sought, discussions around different accreditation models covered primarily two accreditation systems; IODE National Oceanographic data Centre accreditation, and Core Trust Seal Accreditation.

Presentations from the Marine Institute in Ireland on IODE accreditation and from the Norwegian Marine Data Centre at IMR on Core Trust Seal (CTS) provided useful experiences of going through an accreditation process.

Generally speaking, both accreditation models places clear requirements on the organisation seeking the accreditation to provide evidence of how processes are handled in a range of categories. Both IODE and CTS have strong links and memberships of the World Data System (WDS), an interdisciplinary body of the International Science Council. Both accreditation models reference the WDS catalogue of evaluation criteria that are organised into three main components:

- Organisational Framework
- Management of data, products, and services
- Technical infrastructure

While the IODE accreditation provides extensive documentation for the requirements for a quality management framework for National Oceanographic Data Centres, the Core Trust Seal provides a more transparent assessment with a 4-step compliance level evaluation. Both models have a formal process for submitting evidence for accreditation with slight variations in timelines stated on respective websites (and different experiences from Marine Institute and IMR from those).

There were different opinions in DIG in terms of which accreditation model was most exhaustive, relevant, and approachable. Overall, there was recognition that both accreditation models will offer ICES Data Centre a comprehensive exercise in documenting and defining processes for managing data. DIG examined some broad categories to help guide a decision, listed in the paragraphs below.

7.1 Perceived quality of accreditation

This is not a full-blown assessment of whether one set of quality criteria is better than the other, but rather how the perception of having an accreditation from the relevant body would be perceived by stakeholders and wider community.

The IODE accreditation was felt to be the most exhaustive in defining scope and providing content for guidance. But the organisation and accessibility of the content can be hard to locate. In addition, the scope of the IODE accreditation is limited to oceanography and biology – while there is an increasing recognition in ICES that the inclusion of social and human dimension data is important.

The Core Trust Seal is more multidisciplinary, so guidance is much more generic, and there is more emphasis on the data centre seeking accreditation to provide evidence of own processes.

The content and presentation of material for CTS is easily accessible and well presented. There are no limitations set on the type or discipline of data that are in scope for CTS accreditation.

7.2 Accreditation response time

IODE accreditation provides NODC's with a 2-year window to complete the application and a further year to remedy any shortcomings identified. There is a 5-step process outlined in the IODE Quality Management Framework for National Oceanographic Data Centres (IOC Manuals and Guides 67, Dec. 2013).

The CTS process involves payment of a fee, after which peer review of the accreditation application should be completed within two months. The applicant then has one month to return clarifications after receiving feedback. This process for clarification has a maximum number of returns of five times – so a total of maximum five months extra.

DIG did not feel that the response time was a major issue in comparison to the amount of time that would need to be invested in the provision of evidence for accreditation in either case.

7.3 Accreditation transparency

The IODE process involves submission of accreditation documents, but no publication of these. A formal recognition of the application accreditation will be aligned with an IODE meeting, but the outcome of the process can be known well ahead of this. There did not seem to be anything preventing an organisation from publishing its own accreditation information, but there is no catalogue of responses with which to compare.

The CTS application process is fully transparent, and all accredited repository reports are published on the CTS website. This approach provided IMR with useful pointers to other, similar, organisations that had obtained accreditation already. So in addition to be able to look up and compare other accreditation documents, ICES' own accreditation would become a resource for others in the future.

DIG recognised that the ICES data policy and overall ethos is to strive for transparency where possible.

7.4 Accreditation fees

IODE does not charge for its accreditation and subsequent reviews. CTS charges 1,000 Euro for accreditation and re-accreditation in a three-year cycle. DIG discussion the various merits of charge vs no charge and recognised that while there is a cost saving in the no-charge model, the relative cost compared to the staff time required to prepare material, the overall cost is small even for the CTS accreditation.

7.5 DIG decision

There was not full consensus on one particular accreditation approach in DIG. Overall there was however agreement that either of the accreditation schemes would serve ICES well in preparing the evidence for processes. DIG also observed that the accreditation process itself focusses on the existing processes, and does not in itself guarantee best data management practises. But it initiates a programme of work that will identify areas in need of improvement and areas of strength – much like what has been initiated with the governance work. Going through a formal process

provides clarity and a need to deliver – but it is equally important to use the information developed in the accreditation process to develop an improvement programme. In the end DIG members voted on their preferred option, and while some members abstained, the outcome was ten votes in favour of CTS with four members in favour of IODE. So the final DIG decision is to start accreditation with the Core Trust Seal process.

It is likely that as an IODE Associated Data Unit, ICES might decide to utilise the material prepared to also seek IODE accreditation – which will be largely complementary as both approaches are built on the same core from WDS.

It should also be noted that DIG identified ICES Data Management accreditation as a medium potential to disrupt in the tracker now used for following changes that may impact ICES data management (see report section 8). This means that there are some challenges in terms of staff resources required to meet this task, as well as opportunities in gaining recognition and increasing confidence in ICES data and advice products.

8 Tracking challenges and opportunities

The ability to identify potential pressure or new tools that can provide effective data management solutions is important for ICES. DIG has initiated a future challenges and opportunities tracker, which will be reviewed regularly. Over the previous year, the initial horizon scanning exercise was turned into a more formal tracker that allow categorisation and evaluation of technologies and developments that might pose challenges or present opportunities for more efficient solutions – or both.

During the meeting, DIG reviewed the initial entries, updated wording, categories and in some instances the potential impact, which is termed the potential to disrupt.

Not all of the concepts tracked by DIG will necessarily come to fruition, and the register may not necessarily cover every conceivable technical challenge or opportunity for the future. But it is composed by the collective expert knowledge of DIG members, and the respective groups that these members also serve both within and out with ICES.

Currently, DIG has identified 15 broad topics, most of which represents both opportunities and challenges. The full current snapshot is included in Annex 8.

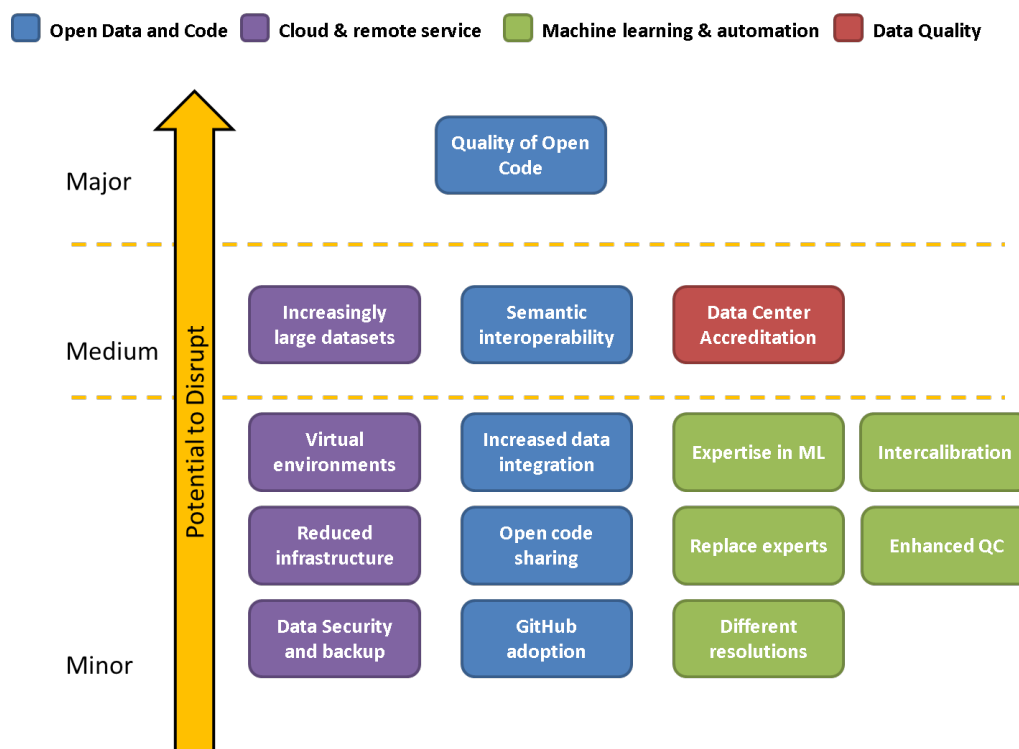


Figure 1. Categorisation of challenge and opportunity topics.

In addition to the potential to disrupt, DIG has also organised content into four categories, which are briefly discussed below.

8.1 Machine learning and automation

The machine learning and automation category has the highest number of topics, but currently all set to minor potential to disrupt. It is clear that the formation of a machine learning group in ICES is a positive step that might help clarify some of these topics and perhaps revise them accordingly. Many of the challenges and opportunities with regards to machine learning are dependent on the rate of adoption if various expert groups. ICES Data Centre and DIG will monitor recommendations and requests for any indications that this needs to be revised or updated.

8.2 Cloud and remote services

ICES already have a cloud strategy, and has been utilising cloud technology for both backups and projects with virtual research environments. One challenge topic in particular is considered to have a higher potential to disrupt. As data collection technologies are changing and often increasing sharply in storage volumes, it is likely that there will be increased community expectations and needs for ICES to also store larger volumes of data, especially if machine learning projects starts to emerge. So far, these aspects have largely been managed in identifying the right stage for ICES to ingest the analysed, tabular data to store in database. However, as methods develop, it is likely that increasing needs for different formats and granularities of data will present increasing demands for storage management in ICES. Two examples of current activities that are already increasing the data volume (albeit in a controlled manner) are the SmartDots application where images are stored, and the spatial fisheries data products, where GIS layers can consume large volumes of data as well.

8.3 Open data and code

Normally it is mainly the benefits of openness and transparency that are being discussed when the topic falls on open data and code. And the overall positive effects of having wider exposure and many eyes on data and code to help spot issues should not be underestimated. However, there are also some major concepts that needs to be considered as code becomes more intermingled between community contributions and in house developments. If the process is not carefully controlled and subject to testing, there is an increasing risk that processing errors with potentially large impacts on advice can work its way into systems. While DIG in no way want to discourage the increasing open data and code culture in ICES, it is necessary to recognise the need of ensuring that code have been tested when it is to be used in formal advice. TAF may well be able to assist in mitigating some of the challenges, as well as benchmark processes and a stronger quality management framework.

8.4 Quality assurance

This category was only added during the DIG meeting and only an entry for the upcoming accreditation process for the ICES Data Centre have been added to far. This activity is registered as a medium potential to disrupt, both from the positive side of enhancing and improving existing processes, and with a risk of presenting a drain on staff resources in the ICES Data Centre.

It is recognised that this category is not yet fully populated, and especially content from the inter-sessional ACOM-SCICOM workshop of data quality should be reviewed to add and register findings. This will be carried out in the coming months [DIG Action#20]

9 Data guidelines and standards

Through the last year, DIG has been defining a method for signposting and describing data guidelines, conventions, and standards. WGFAST has provided valuable input to this process on the basis of their experiences with working up metadata conventions and high definition data storage formats.

The work started with a collection of 13 data guidelines that DIG have maintained in its previous composition as an expert group for oceanographic data management. Now the membership of DIG is wider, and more focussed on broader, strategic issues for all of ICES Data. So the decision was made to look for alternative expert groups and external organisations to become maintainers for these guidelines. Work will progress over the coming year to ask other groups if they are willing to look after the data guidelines, and the Marine Environmental Data and Information Network (MEDIN) in the UK have agreed to collaborate on maintaining data guidelines.

The format for signposting data guidelines and formats is defined based on descriptive fields commonly used in open source code development. Each guideline of format will have the following information available:

Field	Content
Title	Meaningful title to aid searching for the most relevant guidelines
Description	Short paragraph describing the content
Maintenance status	Either "Maintained", "Looking for maintainers" (for 1 year), or "Not maintained"
Current Maintainer	Only if status is maintained or looking for maintainer will this be populated with the relevant ICES expert group
Maintainer URL/Contact	Email or web link to get in touch with maintainer
Link to Major version	The current accepted, reviewed, and published version of the data guidance. ICES publications with doi.
Link to development version	If the maintainer is developing new versions in public, links can be provided to collaborative spaces here (e.g. GitHub, Google Documents etc.)
Supporting development comments	Optional Comments to indicate to potential contributors how they can best support development

All 13 existing Data Guidelines will be completed once initial dialogue on the majority of oceanographic data guidelines have been completed [DIG Action#19].

It is envisaged that the existing webpage for data guidelines will be replaced with a table that lists the titles, maintenance status, link to most recent major version, and link to a more detailed page that lists the full information about each guideline.

This approach can subsequently be extended to include the WGFAST metadata and high definition data formats, and conceptually all other guidance developed to governance groups, other expert groups, and ICES Data Centre to document data guidance. However, the first step is to get the structure established and updated with the existing data guidelines over the coming year [DIG Action #19].

DIG will no longer maintain the data guidelines as a group, but instead aim to find the most appropriate expert groups to look after and develop these.

10 New guidance on reference data

Questions had emerged during the year relating to the change and update of reference/vocabulary data and how this should be handled in terms of support for originally submitted data and potential updated. While the questions related specifically to an example of the rework of maturity scales, it was recognised as a more general management question on how to ensure that the impact of changes to reference data is known, potentially across multiple systems.

DIG will, in collaboration with WGSMART write up guidance document for handling changes to reference data [DIG Action#3], which will be added to the list of data guidance with appropriate maintainer, and version in due course.

11 Best practise for Data Management Handbook

DIG and ICES Data Centre developed a user handbook on Best practise for Data Management (doi 10.17895/ices.pub.4889) in preparation for the Annual meeting of ICES Expert Group Chairs (WGCHAIRS) in January 2019. The handbook has generally been well received, and is already in use for guidance. For example, the handbook was referenced in the ACOM-SCICOM Data Quality document developed Feb-Mar 2019. The handbook has also been referenced in the Workshop on joint data flows in the Barents Sea (WKBAR) where data consolidation is part of the exercise.

DIG will continue to review the handbook to ensure it stays relevant [DIG Action #15]. The handbook will of course also be added to the list of guidance discussed above.

12 ICES Strategy, DIG, and alignment of activities

At various points in the agenda, DIG discussed the group's role as an operational group, and how the new ICES strategy, Science plan and emerging Advisory plan outlines specific issues for DIG to advise on or monitor. These points were brought together for reporting as they align with the overarching direction, functioning and engagement with DIG.

12.1 Strategic plans

The Chairs of SCICOM and ACOM kindly attended a session where the strategic plans were discussed. Overall it was felt from both the ACOM and SCICOM chairs that DIG was moving in the right direction by focussing on the data governance and quality aspects on a broader level. There was recognition that as an operational group, there is a variable inter-sessional workload for the group. In many instances, the initial contact to DIG for such work is via the chair, but a large number of group members indicated willingness to assist and lend expertise to issues as they arise through the year. Some members were only able to provide effort during the annual meeting, but their input and contribution is nonetheless highly valued as most of the inter-sessional work is summarised and discussed at this stage. Overall the dialogue with ACOM and SCICOM is seen to be productive and well-functioning without the need for further formalisation.

DIG identified some key areas of the ICES strategy where most activity is likely to be centred on the "Essential Data for Science and Advice" and the "Evidence for decision making" topics. This does in no way suggest that DIG will not address other strategic topics, but that these two points and considered most aligned with the work of DIG.

Likewise, two main goals were identified from the science plan; "Observation and exploration" and "Emerging techniques and technologies". DIG did however observe that for the science plan there is a wider set of goals that may all require advice or support.

12.2 DIG role and function

DIG reviewed the description of its role and function in the newly updated Guidelines for ICES Groups document, and agreed with the scope, while recognising that it is broad and may at times become resource/capacity limited. The main roles identified for DIG are:

- Advice on data management in general
- Advice and review of data policy
- Addressing technical issues
- Supporting and developing user oriented guidance
- Feedback, guidance to ICES Data Centre, SCICOM and ACOM on
 - New data products
 - Interaction with national data centres
 - Data handling and storage
 - Metadata
 - Use of IT
 - Data quality

There was full recognition and agreement on the functions of DIG. While the current work on data governance and future challenges and opportunities are not explicitly in the description, they are considered overarching approaches that DIG has adopted to carry out these functions.

12.3 Engagement with ICES community

There has been a significant increase in the way DIG engages with other groups and initiatives since becoming an operational group. Especially the attendance at the WGCHAIRS meetings and the event at the Annual Science conference have been seen as points of broader engagement.

Overall, there is an increasing number of items on the DIG agenda, and while they are already regarded as overarching tools to achieve the functions listed in section 10.2, there is a desire to ensure that it is straightforward to communicate the role, function, and activities of DIG in a clear coherent way. During the meeting, some initial sketches and discussions took place, and work to draw up a work scope diagram that will assist in future engagement [DIG Action #18].

12.4 Review of terms of references

DIG reviewed terms of references and have provided updates to these (Annex 3). The main changes from the current terms of references were to:

- Consolidate ToR's A and B into a single ToR for overall advice and feedback
- Update ToR C to reflect links to ICES strategy rather than ICES Data Strategy
- Update/replace ToR D with more explicit work to support and facilitate data governance work
- Update ToR E to include current as well as future challenges and opportunities.

Due to the consolidation, the letter codes of ToR's will change for next year

12.5 DIG Chair

The current chair of DIG is reaching the end of the initial three-year term. It is SCICOM that decides on the chair of operational groups, but DIG has traditionally always put forward a recommendation for the chair.

This year there were no forthcoming candidates to take on the chair, and the current chair indicated willingness to take on the one-year extension, which was agreed with the group to go forward as the recommendation from DIG.

Current and previous chairs observed that it is highly beneficial to have a new chair identified early on to ensure a good period for passing over and providing supporting information. And so all members of DIG were encouraged to consider options for an incoming chair for the next meeting in May 2020. This will allow an overlap period until the change of chairmanship in January 2021, subject to the approval from SCICOM during the September meeting.

Annex 1: List of participants

Name	Institute	Country (of institute)	Email
Adriana Villamor	ICES	ICES	adriana.villamor@ices.dk
Anna Osypchuk	ICES	ICES	Anna.Osypchuk@ices.dk
Arni Magnusson	ICES	ICES	arni.magnusson@ices.dk
Carlos Pinto	ICES	ICES	carlos@ices.dk
Chris Moulton	OSPAR	OSPAR	Chris.Moulton@ospar.org
Christian von Dorrien	Thünen-Institute of Baltic Sea Fisheries	Germany	christian.dorrien@thuenen.de
Colin Millar	ICES	ICES	colin.millar@ices.dk
David Currie	Marine Institute	Ireland	David.Currie@Marine.ie
Gisbert Breitbach	HZG Centre for Materials and Coastal research	Germany	Gisbert.Breitbach@hzg.de
Graham Allen	BODC	UK	graham.allen@noc.ac.uk
Hans Mose Jensen	ICES	ICES	hans.jensen@ices.dk
Helge Sagen	IMR	Norway	helge.sagen@hi.no
Henrik Kjems-Nielsen	ICES	ICES	henrik.kjems-nielsen@ices.dk
Hjalte Parner	ICES	ICES	hjalte.parner@ices.dk
Ingeborg de Boois	Wageningen Marine Research	The Netherlands	ingeborg.deboois@wur.nl
Jens Rasmussen	Marine Scotland	UK	Jens.rasmussen@gov.scot
Joana Ribeiro	ICES	ICES	joana.ribeiro@ices.dk
Johannes Johansson	Swedish Meteorological and Hydrological Institute	Sweden	johannes.johansson@smhi.se
Julie Davies	DTU Aqua	Denmark	joco@aqua.dtu.dk
Lena Szymanek	National Marine Fisheries Research Institute	Poland	lena@mir.gdynia.pl
Malin Werner	SLU	Sweden	malin.werner@slu.se
Marcellus Rödiger	Thuenen Institute	Germany	marcellus.roediger@thuenen.de
Marcin Wichorowski	Institute of Oceanology Polish Academy of Sciences	Poland	wichor@iopan.gda.pl
Mehdi Abbasi	ICES	ICES	Mehdi.Abbasi@ices.dk

Name	Institute	Country (of institute)	Email
Neil Holdsworth	ICES	ICES	neilh@ices.dk
Susanne Tamm	BSH	Germany	susanne.tamm@bsh.de
Taco de Bruin	NIOZ	Thee Netherlands	bruin@nioz.nl
Vaishav Soni	ICES	ICES	vaishav.soni@ices.dk
Wim Allegaert	ILVO	Belgium	wim.allegaert@ilvo.vlaanderen.be

Annex 2: Agenda

Tuesday 21 May

09:30 Welcome, introductions & practical arrangements (*25 minutes*)

09:55 Agenda for meeting (*5 minutes*)

10:00 Previous Action List, Recommendations given and received (*1 hour*)

11:00 Coffee break (*15 minutes*)

11:15 Governance Framework – Activity updates (*45 minutes*)

- WGDG
- TAF
- SmartDots
- Acoustic

12:00 Breakout groups – governance evaluation candidates – 2-4 groups (*1 hour*)

- Identify system/collections, Custodians and group(s) to consult
- Action plan for conducting evaluation (Assign DIG members, time plan)

13:00 Lunch break (*1 hour*)

14:00 Governance evaluation, candidates and action plan (*30 minutes*)

14:30 ICES Strategy, Science Plan, and Structure. Role of DIG as an operational Group (*30 minutes*)

15:00 Data policy review – subgroups (*1 hour 15 minutes, including coffee break*)

6. ICES Data Policy
7. VMS-Logbook Policy
8. VME Data Use
9. RDB Policy

16:15 Data policy review – plenary – summarise any recommendations (*30 minutes*)

16:45 Update DIG activities and related groups (*1 hour 10 minutes*)

- Data's Den – Annual Science Open Sessions 2018
- WKSEATEC II
- OpenSeaLab II - EMODNet/ICES/CMEMS Hackathon
- OSPAR ICG-Data
- IODE/UN Decade of Ocean Science
- Dark Data (WGZE)

17:55 DIG Chairing – options (*5 minutes*)

18:00 Close for the day

Wednesday 22 May

09:00 ICES Data Centre Projects Update (*4 hours*)

- ICES Data Portal
- TAF
- Acoustic Portal
- VISA and follow-up
- How ICES is using Github

13:00 Lunch Break (*1 hour*)

14:00 ICES Data Centre Accreditation discussion (*1 hour 30 minutes*)

- Marine Institute Presentation on IODE accreditation
- IMR Presentation on Core Trust Seal accreditation
- Data Centre Discussion Document
- Plenary discussion and shared experiences
- DIG recommendation

15:30 Coffee break (*15 minutes*)

15:45 Future challenges tracker – review and update breakout groups (*45 minutes*)

1. Machine learning & automation
2. Cloud and remote services
3. Open data and code
4. Transparency of process

16:30 Plenary & summary on future challenges tracker (*40 minutes*)

17:10 DIG Links with other data & technology groups in ICES (*20 minutes*)

17:30 Recap of the first 2 days, and action list update (*30 minutes*)

Thursday 23 May

09:00 ICES Best Practise in Data Management – Handbook & further development discussion (30 minutes)

09:30 DIG Data Guidelines – Progress update and Breakout groups (1 hour 30 minutes)

1. 1-3 groups to complete template for existing guidelines + any recommended actions
2. Search and locate candidate guidelines in ICES
3. Draft instructions/documentation for template and review process

11:00 Coffee Break (15 minutes)

11:15 Data guidelines plenary (45 minutes)

12:00 Review ToR's, final recommendation on chair (20 minutes)

12:20 Linking up Governance, Future opportunities and challenges, QA & QC, and recommendations for persistent identifiers. (1 hour)

13:20 Summary and recap of Actions, Recommendations & Dates for next meeting (10 minutes)

13:30 - 15:00 Lunch breaks, social media updates, completion of writing/actions, finalised content to Secretariat (Guideline templates, social media updates, etc.).

Annex 3: DIG terms of reference for the next meeting

The **Data and Information Group** (DIG), chaired by Jens Rasmussen, United Kingdom, will meet in ICES, Copenhagen, Denmark, 26-28 May 2020 to:

- a) Provide guidance and feedback to the ICES Data Centre
- b) Advise on data regulations and their impact on ICES Strategy, ICES Data Policies, and license considerations.
- c) Facilitate data governance by performing evaluations and encouraging dialogue between expert groups, governance groups, DIG, and the ICES Data Centre to adopt best practises in data management.
- d) Evaluate and monitor current and future challenges and opportunities in data management and new technologies for ICES.

DIG will report by 28 June 2020 to the attention of the Science Committee.

Supporting Information

Priority	The Data and Information Group provides ICES with solicited and unsolicited advice on all aspects of data management including technical, data policy and data strategy and user oriented guidance. This operational group flies the flag for ICES in setting standards for global databases. It also provides an important interface for oceanographic, environmental, and fisheries data management in ICES, and promotes good data management practice
Scientific justification	<p>Term of Reference a) Direct interfacing with the ICES Data Centre around priorities and general best practise recommendations enables the Data Centre to receive both solicited and unsolicited advice on solutions and practises from a broader international community.</p> <p>Term of Reference b) DIG is constitutes of data maangers from member coutrnies and can raise awareness of specific regulation or licensing perspectives that may impact on data sharing or collation on an ICES level.</p> <p>Term of Reference c) Promoting data governance and best practise for data management in the wider ICES community os becoming increasingly important as multi-disciplinary use of, and complexity of data are both increasing. It is important for the user community to be engaged in identifying priorities for relevant data systems, and provide ICES Data Centre with a clear line of communication for implementing changes .</p> <p>Term of Reference d) By reviewing and monitoring current and future activities, DIG can proactively advice ICES on emerging issues. Utilising the experience of the group and combining it with the operational knowledge of the ICES Data Centre, the potential benefits and challenges can be idetified and their potential impact presented.</p>
Resource requirements	The resource required to undertake additional activities in the framework of this group is negligible.
Participants	The Group is expected to be attended by some 15–30 members, with good international and topical coverage

Secretariat facilities	Meeting facilities, organization and facilitation of WebEx meetings (frequency and participants depending on topics to be discussed. Participation of ICES Data Centre
Financial	No financial implications.
Linkages to advisory committees	ACOM (indirect)
Linkages to other committees or groups	As Data is an important topic for most groups under SCICOM and ACOM, this group links to a large number of groups, although often indirect.
Linkages to other organizations	There are linkages with relevant international bodies and programmes like SeaDataNet/SeaDataCloud, EMODNet, IOC and its Working Committee on International Oceanographic Data and Information Exchange (IODE), OSPAR, and HELCOM.

Annex 4: Recommendations

Recommendation	Adressed to
Review comments, edits, and observations to the VMS-Logbook data policy made by DIG as part of routine review of ICES Data Policies. Final edits and updates to published policy is for agreement with WGSFD.	WGSFD
Review comments, edits, and observations to the RDB + RDBES data policy made by DIG as part of routine review of ICES Data Policies. Final edits and updates to published policy is for agreement with SCRDB	SCRDB

Annex 5: Actions

Number	Action	Addressed to (Lead in bold)	Deadline
1	Clarify if governance of ESAS are included in ToR's for Seabird group	Neil Holdsworth, Ingeborg de Boois	August 2019
2	Contact WGPME to get more details about Iugols DB and potentially establish contact with data centre	Jens Rasmussen	July 2019
3	Develop guidance on implementation of changes in reference data management considerations.	Ingeborg de Boois, David Currie, Joana Ribeiro, Neil Holdsworth	September 2019
4	Plan Webex between DC, DIG, SCRDB, governance group chairs for status/shared needs	Neil Holdsworth, Jens Rasmussen, Ingeborg de Boois, Julie Davies	December 2019
5	Conduct DOME governance evaluation	Chris Moulton, Hans Mose Jensen	March 2020
6	Conduct Spatial Fisheries Data governance evaluation	Christian von Dorien, Colin Millar, WGSFD Chairs	September 2019
7	Plan Bycatch of sensitive species governance evaluation	Christian von Dorien, Carlos Pinto, WGBYC Chairs	March 2020
8	Establish top level folder in DIG Sharepoint for governance evaluation	Jens Rasmussen	September 2019
9	Review data licensing options, evaluating pros and cons of different models	Chris Moulton, Joanna Ribeiro, Helge Sagen, Ingeborg de Boois	March 2020
10	Promote Awareness of OpenSeaLab 19 hackathon	All DIG members	August 2019
11	Update paragraph on the development work for ICES Data Portal for DIG newsletter to be circulated after ICES ASC	Carlos Pinto, Mehdi Abbasi	September 2019
12	DIG Beta testers for the development of the ICES data portal – contact as needed	Carlos Pinto, Mehdi Abbasi, Lena Szymanek, Ingeborg de Boois, David Currie, Jens Rasmussen	As needed – update in May 2020
13	Follow on from Action 9 – consider implications of licensing for github repositories	Jens Rasmussen, Chris Moulton	May 2020
14	Map out existing DIG group members and links to other group membership (EGs, SGs, Committess)	Neil Holdsworth, Jens Rasmussen	August 2019
15	Review best practise document and suggestions for potential updates or inclusions	Ingeborg de Boois, Wim Allegart, Lena Szymanek	May 2020
16	Add Best Practise document to IODE Ocean Best Practise Portal	Neil Holdsworth	September 2019
17	Review and summarise existing licensing in use on ICES github	Colin Millar, Chris Moulton, Jens Rasmussen	March 2020

Number	Action	Addressed to (Lead in bold)	Deadline
18	Redraw or design DIG work scope diagram	David Currie, Ingeborg de Boois, Neil Holdsworth	May 2020
19	Work plan and execution of work on Data Guidelines revisions, updates, and presentation	Hjalte Parner, Jens Rasmussen, Graham Allen and MEDIN, BODC staff	Immediate contact + updated work plan for September 2019
20	Populate future challenges and opportunities with content from ACOM-SCICOM data quality document	Jens Rasmussen, Neil Holdsworth	September 2019

Annex 6: A semantically-enabled summary of the Marine Institute’s commercial sampling data

Commercial Fisheries Sampling Summary

Using semantic data techniques makes it easier to link our data to other people's data and enrich our queries. An example is that you can filter the summary data by things like ICES Working Group or Stock (from the ICES vocabulary server) or Conservation Status (from DBPedia). It also tries to pull an image and text from DBPedia when it can. Recursive queries are also easier than in our operational databases (e.g. filtering by all samples from 27.7).

Consists of

- 1. a Java program using the Apache Jena library to convert a database query into semantic data using the ICES semantic vocabulary services. This data is then loaded into an Apache Fuseki server running in a Dock container <https://github.com/davidcurrie2001/SemanticFishData>
- 2. A Shiny app to display the data and allow filtering by some of the enriched vocabulary. The app is hosted in a Docker container. The app can either query the SPARQL endpoint from (1) directly or use pre-generated R data files (for the sake of speed) <https://github.com/davidcurrie2001/SemanticSummary>

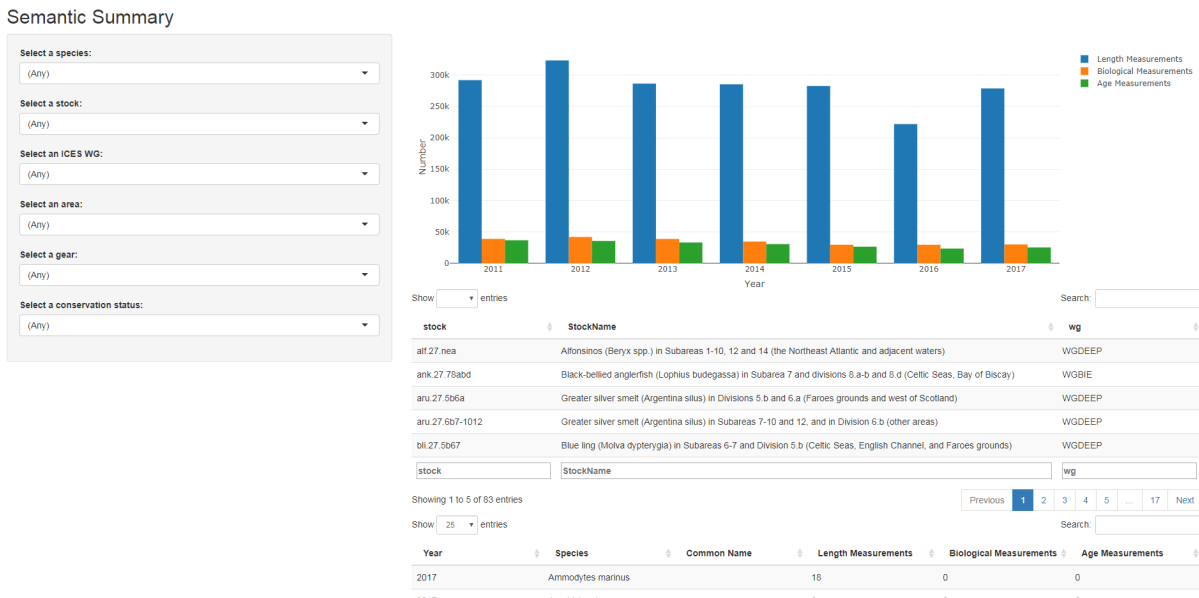


Figure 1 Default view of "all" data.

The app displays i) a plot of the number of length, biological, and age measurements per year, ii) the ICES stocks and working groups that are relevant to that data, iii) a table with the number of length, biological, and age measurements per species and year.

Semantic Summary

Select a species:

Cod

Select a stock:

(Any)

Select an ICES WG:

(Any)

Select an area:

(Any)

Select a gear:

(Any)

Select a conservation status:

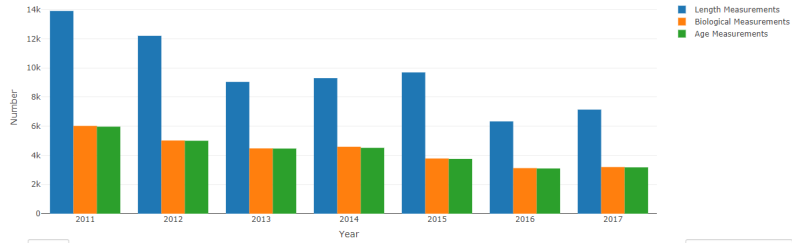
(Any)

Cod

MI Species Dashboard link



The Atlantic cod (*Gadus morhua*) is a benthopelagic fish of the family Gadidae, widely consumed by humans. It is also commercially known as cod or coding. Dry cod may be prepared as unsalted stockfish or as cured salt cod or klipfish. In the western Atlantic Ocean, cod has a distribution north of Cape Hatteras, North Carolina, and around both coasts of Greenland and the Labrador Sea. In the eastern Atlantic, it is found from the Bay of Biscay north to the Arctic Ocean, including the Baltic Sea, the North Sea, Sea of the Hebrides, areas around Iceland and the Barents Sea. It can grow to 2 m (6 ft 7 in) in length and weigh up to 96 kg (212 lb). It can live for 25 years and usually attains sexual maturity between ages two and four, but cod in the northeast Arctic can take as long as eight years to fully mature. Colouring is brown to green, with spots on the dorsal side, shading to silver ventrally. A stripe along its lateral line is clearly visible. Its habitat ranges from the shoreline down to the continental shelf. Several cod stocks collapsed in the 1990s (declined by >95% of maximum historical biomass) and have failed to recover even with the cessation of fishing. This absence of the apex predator has led to a trophic cascade in many areas. Many other cod stocks remain at risk. The Atlantic cod is labelled vulnerable on the IUCN Red List of Threatened Species.



Show	entries	Search:
stock	StockName	wg
cod 27.6a	Cod (<i>Gadus morhua</i>) in Division 6.a (West of Scotland)	WGCSE
cod 27.6b	Cod (<i>Gadus morhua</i>) in Division 6.b (Rockall)	WGCSE

Figure 2 Selection of a species.

If a user filters by species the app will attempt to fetch an image and text from DBPedia and provide a link to the Marine Institute’s “Species Dashboard” app.

Semantic Summary

Select a species:

(Any)

Select a stock:

(Any)

Select an ICES WG:

WGBIE

Select an area:

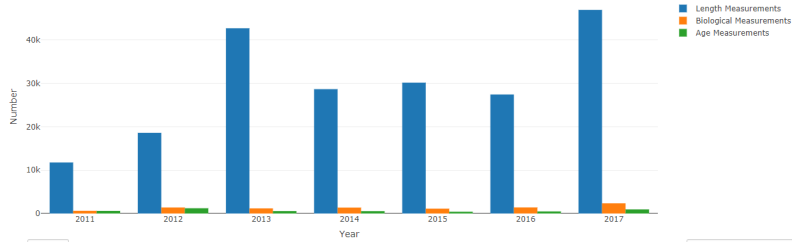
27.7

Select a gear:

All Bottom trawls

Select a conservation status:

(Any)



Show

entries

Search

stock	StockName	wg
ank 27.78abd	Black-bellied anglerfish (<i>Lophius budegassa</i>) in Subarea 7 and divisions 8.a-b and 8.d (Celtic Seas, Bay of Biscay)	WGBIE
hke 27.3a46-8abd	Hake (<i>Merluccius merluccius</i>) in Subareas 4, 6, and 7, and Divisions 3.a, 8.a-b, and 8.d, Northern stock (Greater North Sea, Celtic Seas, and the northern Bay of Biscay)	WGBIE
ldb 27.7b-k8abd	Four-spot megrim (<i>Lepidorhombus boscii</i>) in Divisions 7.b-k, 8.a-b, and 8.d (west and southwest of Ireland, Bay of Biscay)	WGBIE
meg 27.7b-k8abd	Megrim (<i>Lepidorhombus whiffiagonis</i>) in Divisions 7.b-k, 8.a-b, and 8.d (west and southwest of Ireland, Bay of Biscay)	WGBIE
mon 27.78abd	White anglerfish (<i>Lophius piscatorius</i>) in Subarea 7 and divisions 8.a-b and 8.d (Celtic Seas, Bay of Biscay)	WGBIE

stock

StockName

wg

Showing 1 to 5 of 5 entries

Show

25

entries

Previous

1

Next

Search

Year	Species	Common Name	Length Measurements	Biological Measurements	Age Measurements
2017	Lophius budegassa	Black-bellied anglerfish	6500	501	0

Figure 3 Filtering.

Users can filter by things such as stock, area, working group, gear, and conservation status.

Annex 7: DIG Data Governance Questions

The full DIG template contains 10 broad governance categories defined in the DIG 2018 report. This has been further refined with a number of questions. In the DIG working template, each question contains a number of additional fields to provide an answer, provide a rating (1-5), A written justification for the rating, Identify improvements, and to assign actions and action owners. The table below lists the questions without these additional fields to improve the layout of the table.

Category	Question
Architecture & Governance	Is there an identified group that oversees or defines the rules of the system? This includes submission formats, exchanges, and services. This will typically be done in collaboration with the data centre, but require regular review, and named members/contact persons
	Does the system/process align with a wider perspective on data integration? Is there documentation that demonstrates integration or adoption of information exchange and system design?
	Are there clear deliverables or products associated with the system? Are they maintained or reviewed on a regular basis?
	Is there a clear/documented decision process for updates and changes to the system?
	Is there reporting or progress monitoring associated with the system use?
	Are there documents for business continuity and disaster recovery processes?
Data Development	Are there data processes that allow versioning and tracking of imported data (or an audit trail)?
	Are there test reports for calculated fields confirming code/calculations?
	How are products generated from the source? Can the process be replicated at a later stage?
	Is the data structure designed to enable effective analysis, advice and decision making by analysts?
Database Operations	Is integrity of associated databases and structured data sets assured?
	Does the database solution use organisationally defined standards for naming conventions of fields, tables, etc.
	Is the database scaled to anticipated performance?
	Does ICES have the technical capability and understanding of the technology for this database/data model?
Data Security	Does the system enable appropriate access to data? Is this a process aligned with organisational permissions management tools?
	Does the access/security meet relevant regulatory standards or demands for data privacy and protection?
	Are there guidance and communication on what steps are taken to safeguard collected data?

Category	Question
Reference and Master Data	Does the system use vocabulary services for reference/lookup values, and are those vocabularies clearly identified and versioned?
	Does the system identify authoritative sources and contributors of the data?
	Does all data have clearly defined stewards/custodians?
	Is there a clear process for managing changes to reference and master data?
Data Discoverability	What processes are in place for organisational data delivery/sharing? E.g. Are there key reporting dates, regular reviews/QC periods?
	Is the data system/process supporting easy discoverability of the data to all relevant parties/stakeholders?
	Is there linkage between data discoverability and metadata management processes?
Document and content management	Are there documented/established principles for handling unstructured content associated with the system/process?
	Is business continuity in place for unstructured content?
	Are relevant documents and other content associated with the system retrievable and locatable by all relevant users?
Metadata	Is there explicit/direct linkage between metadata and source data?
	Does the source data inform/update the metadata through a recognised process (e.g. regular manual or automated)
	Are metadata records available for all data in the system/process? Is it in a recognised/INSPIRE compliant format?
	Is metadata utilised to facilitate searchability or contact to the organisation?
	Is it possible to export or access metadata through services that enable direct data access?
Data Quality	Is there a process to measure and provide quality of data in relation to data integrity, legal requirements and business priorities?
	Is the system/process registering all data quality checks on the ICES QC catalogue? And is the QC catalogue used to review QC efforts?
	Are there any standardised data quality reports or expressions that can be associated with data - either directly in database, or as part of metadata?
Compliance with FAIR principles	How "Findable" are the data within ICES? Is the search for data integrated or standardised within the wider ICES Data ecosystem?
	How Findable are the data out with ICES? (e.g. Are the data registered/submitted elsewhere, and if so, are there clear links to Master Data, originator information etc.)
	Are data appropriately accessible?
	Are the formats for accessible data described/documented?
	Does the data system/process provide interoperability through services?
	Are the formats for interoperable service formats described/documented?

Category	Question
	Are data reusable in terms of clear change history/versioning or persistent identifier?
	Is the provenance of the data described/documentated?

Annex 8: Future Challenges and Opportunities

Major Potential to Disrupt				
Category	Topic	Challenges	Opportunities	Mitigation/Action
Open data and code	Quality of open code	The increasing complexity of models and code raises questions about how code is verified and tested.	More "eyes" can mean errors in code are more likely to be found. A formal process will increase the quality of code.	Quality management framework. Code testing. Code reviews. Standardised test datasets. Benchmark process.
Medium Potential to Disrupt				
Category	Topic	Challenges	Opportunities	Mitigation/Action
Cloud and remote services	New data storage challenges: Increasingly large datasets	Capacity planning and expectation management from the ICES Data Centre will be important. Realistic timelines for developing datasets and series (optimism bias)	New or existing virtual research infrastructures and cloud solutions may offer new scalable solutions in time	DIG & Data Centre to monitor recommendations and evaluate cases.
Open data and code	Semantic interoperability of data	Still limited uptake and understanding of semantic linkage and linked data concepts in the broader ICES Community. Clarity about the major vocabulary and ontology structures that will govern and link together is not currently planned - but more opportunistic (e.g. linkage between international bodies, different fields of expertise may be developing their own ontologies).	Huge potential for connecting up data across disciplines. Improved user experiences in finding linked/connected data (e.g. biological and oceanographic data from the same survey without separate searches)	Use cases from Data Centre and DIG 2019 Report Use case from Marine Institute 2019 Report, examples of integration in new data portal & other ICES data systems
Quality Assurance	Accreditation for ICES Data Management	Large effort needed to achieve accreditation. Continuous process may take resources from other areas. May not greatly affect the underlying quality management metrics	Recognition through an internationally recognised accreditation body. Greater confidence in ICES Data management and advice outputs	Use the ICES data management community to assist/advice/provide input. Defining the scope to a manageable piece of work and working step-wise outwards. Action: Decision on CTS Accreditation

Minor Potential to Disrupt				
Category	Topic	Challenges	Opportunities	Mitigation/Action
Machine learning and automation	Familiarisation with Machine learning techniques and tools	Few experts in the ICES community, but there is a working group now. On the short term there is too few experts, but in the long run there will be more. Large expectations to newly formed group WGMLEARN	To collaborate and attract machine learning experts to the ICES community. ICES holds a lot of well-structured data that could provide good opportunities both ways	Formation of WGMLEARN DIG, WGMLEARN, and Data Centre to maintain dialogue
Machine learning and automation	Substitution of experts with Machine learning functions	There is a perceived risk that some machine learning processes may automate some processes to a stage where training new experts will become difficult. This will in turn make it difficult to provide good classification information for training in machine learning	Collaborative efforts between machine learning community and expert community can help expel fears and make more efficient use of expert time, freeing up time from more menial tasks	Discuss with WGMLEARN on their thoughts on communicating development of machine learning, to consider the issue that there could be worries. Discussions around the SmartDOTS development to allow testing of machine learning algorithms against training sets either separately or alongside human analysts
Machine learning and automation	Species not identified to the required taxonomic level	Machine learning will potentially be able to process much larger volumes of samples with known error margins. However it is also recognised that in many current examples of applying machine learning to species identification from images or videos, the taxonomic resolution will be lower	The performance in species identification may be further improved over time with e.g. deep learning, if there are resources and collaboration opportunities	Individual EG will need to discuss tradeoffs between potential volume of samples against precision/resolution Some examples of applications around zooplankton in the community (WGZE)
Machine learning and automation	Intercalibration of regional/area data for training sets	Processes developed in different regions may require new/different/expanded training sets or source data for machine learning to perform effectively across wider areas.	ICES is a natural convergence point for membership countries, so may have an opportunity to facilitate such processes	Some examples already progressing - e.g. ecotaxa for zooplankton DIG and Data Centre to monitor recommendations requests for relevant content Recommendation/request for lugols based image reference database being examined

Minor Potential to Disrupt				
Machine learning and automation	Enhanced Quality checking through more and adaptive machine learning processes.	If ML guided QC were to be implemented, it may well create expectations that ICES would employ such techniques. This in turn would require Data Centre Expertise and resources to develop such approaches	Scope to extended quality checks based on much larger/wider patterns or historic data in a manner that is often tedious or difficult to craft for individual parameters in large datasets. Also potential for faster QC making data available for advice and processing quicker.	DIG and Data Centre to monitor any trends or initiatives in this area
Cloud and remote services	Virtualised work environments	Technological change that potentially require constant connectivity or synchronisation. This will be achievable in working group settings, but may be more difficult to achieve in field conditions (e.g. during surveys)	Creates opportunities to tailor working environments for tasks. E.g. rather than relying on everyone having the right setup on a laptop, a virtual machine can be created to complete a workflow. Multiple operating systems and setups can be easily accommodated.	DIG to identify potential workflows that would benefit from adoption of virtualised work environments. ICES Servers already virtualised TAF is already running parts of processes within virtualised environments, and RDBES is adopting similar approaches
Cloud and remote services	Data security: Enhanced Data Backup facilities	Ability to backup or protect data in the cloud can also lead to concerns about data security, especially in environments that span beyond national/regional boundaries. Potential data protection issues	De-risks local data centres and distributes backup to minimise risks associated with disaster recovery	Risks applies to restricted/sensitive data i.e. commercial catch, VMS 1 - these data types should only be stored on premises 2- If there is a business case for storing offsite (size limitations), then there has to be an evaluation (due diligence) of the cloud service to ensure the service has the appropriate security/ jurisdiction for the data type in questions ICES Data Centre to monitor (this is predominantly an infrastructure service)
Cloud and remote services	Reduced infrastructure management	If most services are phased over to a cloud Infrastructure or Platform as a service solution, it removes local needs for hardware management. However, contractual arrangements and exit strategies from providers become much more important to avoid vendor lock-in. Cost may become more dynamic and part of running	A more dynamic and scalable environment that can be stretched and scaled to meet the organisations need without the need to over-procure capacity for longer term stability.	ICES cloud strategy

Minor Potential to Disrupt

budget rather than single capital investment items in IT.

Open data and code	Increased data availability & submission	Greater need for standardised and well-structured data solutions that can accommodate a potential wider range of data types. Will likely require more information requirements from originators (data context, lineage etc.) to minimise inferred validation/QC.	Enhanced compliance. Opportunity for ICES to be a forwarder/compiler of data for e.g. INSPIRE standards compliance.	Discussions around dynamic submission of data versus quality control already in progress ICES Data Centre already making substantial amounts of open data available.
Open data and code	Open code sharing	There may be reluctance from some communities to openly share code - but this is probably not an issue for most users. It can be difficult to keep track of where different tools are stored e.g. on SharePoint, on user's own repos, on ICES GitHub repos etc.	Improvement of tools over time with a wider range of contributors and participants	GitHub adoption increase TAF will increasingly catalogue/registers code used for assessment ICES Data Centre to compile overview of available code under ICES GitHub accounts
Open data and code	GitHub adoption	A large number of new ICES initiatives are using GitHub. If the licensing terms or availability of GitHub were to change it could have a significant impact on ICES work. Not all of the ICES community is familiar with using GitHub.	Many projects are making good use of GitHub for collaboration and this is having a very positive impact.	Training for using tools like the Transparent Assessment Framework should also incorporate GitHub training, if required. ICES Data Centre should monitor any changes to GitHub licensing and react appropriately.

Annex 9: Connectivity of DIG

The Diagram below illustrates the connectivity between ICES groups. All members of DIG were queried on the ICES register of memberships for links to other groups (e.g. both a member of DIG and another group)

