

ICES DIG REPORT 2018

ICES SCIENCE COMMITTEE

ICES CM 2018/SCICOM:02

REF ACOM, SCICOM

Report of the Data and Information Group (DIG)

30 May to 1 June 2018

ICES Headquarters, Copenhagen, Denmark



ICES
CIEM

International Council for
the Exploration of the Sea

Conseil International pour
l'Exploration de la Mer

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

H. C. Andersens Boulevard 44–46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

Recommended format for purposes of citation:

ICES. 2018. Report of the Data and Information Group (DIG), 30 May to 1 June 2018, ICES Headquarters, Copenhagen, Denmark. ICES CM 2018/SCICOM:02. 32 pp.

The material in this report may be reused 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 the ICES website. All extracts must be acknowledged. For other reproduction requests please contact the General Secretary.

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

<https://doi.org/10.17895/ices.pub.8362>

© 2018 International Council for the Exploration of the Sea

Contents

1	Executive summary.....	1
2	Opening of the meeting.....	2
3	Adoption of the agenda	2
3.1	Review of actions and previous recommendations	2
3.1.1	Recommendations	2
3.1.2	Actions	3
3.2	Feedback from other working groups	6
3.2.1	DATRAS Governance Group.....	6
3.2.2	WKINVITED	6
3.2.3	WGSFD	7
4	Data Guidelines	8
5	Future Data Challenges and Opportunities.....	10
6	ICES Data Centre Update	12
6.1	Transparent Assessment Framework (TAF)	12
6.2	Regional DataBase and Estimation System (RDBES)	12
6.3	Acoustic data portal and output challenges	12
6.4	Eutrophication Tools.....	13
6.5	Marine Aggregates Extraction	13
6.6	New generation of vocab.ices.dk.....	14
6.7	Reports Data Mining	14
6.8	New ICES Data Portal.....	14
6.9	Bycatch (of protected species).....	15
6.10	European Seabirds at Sea (ESAS)	15
6.11	SmartDots (Otolith Annotation)	16
7	Dark Data Recovery and the role of ICES	17
8	Data Legislation updates, Data Policy, and Impact on ICES	18
9	Data Traceability and Provenance in ICES	19
10	Data Governance evaluation and reporting.....	20
11	Joint ASC Session with PGData	21
	Annex 1: List of participants.....	22
	Annex 2: Agenda.....	23
	Annex 3: DIG terms of reference for the next meeting	25

Annex 4: Recommendations	27
Annex 5: Actions.....	28
Annex 6: DATRAS Governance Group (DGG) Report.....	30
Annex 7: DIG Data Governance Framework Categories	32

1 Executive summary

The Data and Information Group (DIG) met in ICES headquarters, Copenhagen, Denmark 30 May – 1 June 2018. 17 members, representing 9 different countries took part in the meeting. A further 10 members of the ICES Data Centre took part in all or parts of the meeting.

DIG and ICES Data Centre submitted a horizon scanning paper to SCICOM in February 2018 on future data challenges and opportunities. The topics of this paper were detailed further and turned into a risk-based monitoring approach, which DIG will use to track concepts of strategic importance in the ICES Data Management sphere.

ICES Data Centre presented a range of current activities and applications in development. The overall direction of the ICES Data Centre is towards a service-oriented architecture, where exchange of data across multiple platforms can be realised. In particular, the work on linking together new advice support applications such as the Transparent Assessment Framework (TAF) and the Regional DataBase and Estimation System (RDBES) demonstrate progress towards this. The update of the ICES vocabulary servers to provide semantic linkage and services is a significant change that will enable open linked data to be provided by ICES, and improve the ways in which an ICES data portal update will be developed. DIG will assist by providing use cases that will demonstrate how these new services and concepts can be realised within the ICES community.

The service-oriented approach also links directly to the tool that DIG has been developing to help profile systems and initiatives through evaluating data governance and management. DIG finalised this internal checklist, and will be applying it to a number of applications in the coming year, and will report summary findings in a user-friendly format.

DIG has had an ongoing discussion around data identifiers and traceability, both within ICES, and back to national submitters. The discussion also included ongoing activities within the ICES community to identify and leverage dark data. Overall, DIG has made recommendations to ensure that data will be adequately described through metadata, and that improved linkage to persistent identifiers can be implemented over time to traceability.

DIG also took a look at the existing Data Type Guidelines, along with other initiatives to produce data format descriptions within the ICES community. There was a recognition that the process needs to be reviewed with an emphasis on balancing flexibility and dynamic content with the ability for ICES to review and publish clear versions of data guides. The current Data Guidelines will not be updated or revised further, but instead, DIG will draft a framework for a new and more open process. This approach will aim to balance the need of the community to work flexibly with the ability to review and publish fixed/major versions.

Finally, DIG discussed and reviewed impact of legislative changes, such as GDPR in preparation for next years' data policy review, which will include both the general data policy and policies for restricted data.

2 Opening of the meeting

The Data and Information Group (DIG) met in ICES headquarters, Copenhagen, Denmark, 30 May–1 June 2018.

17 members, representing 9 different countries took part in the meeting. A further 10 members of the ICES Data Centre attended (parts of) the meeting.

The opening of the DIG meeting was slightly unusual this year since the hackathon workshop WKINVITED, originally proposed by DIG, completed by presenting the outcomes and products developed to DIG in an overlap session on the morning of the 30 May. After this overlap session with WKINVITED, the main DIG meeting started, following a more traditional agenda.

3 Adoption of the agenda

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

DIG supports the ICES strategic goals of “Underpinning Science and Advice through data and information services”:

- ICES Strategic Goal 4: Promote the advancement of data and information services for science and advice needs.
- ICES Strategic Goal 5: Catalyse best practises in marine data management, and promote ICES data nodes as a global resource

The terms of reference for DIG were:

- a) Review the priorities of the Data Centre Action list.
- b) Provide guidance and feedback to the ICES Data Centre
- c) Advice on other 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 SSGs and/or EGs, and review the outcome of those ad-hoc groups.
- e) Promote new technologies and data management infrastructure development.

DIG will report to SCICOM at the next SCICOM meeting during the 2018 Annual Science Conference, and the SCICOM mid-term meeting in March 2019.

The full agenda is available in Annex 2.

3.1 Review of actions and previous recommendations

3.1.1 Recommendations

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, three recommendations were received between the DIG meetings – all from 2017:

Year	Expert group	Recommendation	Recipient(s)
2017	WGBIOP	Quality indicators have been and are being developed for biological parameters. WGBIOP recommends that the ability to flag issues with data, and to record quality indicators is incorporated into the ICES databases. For the incorporation in the databases cooperation with the ICES data centre is vital (see chapter 5.2)	Data Centre DIG
2017	WGEGBS2	Update egg and larvae database gear description and the metadata survey description given in chapter 5.4	Data Centre DIG
2017	WKSEATEC	WKSEATEC 2017 seeks feedback from DIG as to where the work of SEATEC might integrate with data collections and various QC and assessment tools being developed at the ICES 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 from WGBIOP was recognised as being directly relevant to previous discussions held in DIG with regards to quality flagging of data within ICES. ICES Data Centre have developed a QC database, listing the explicit and detailed quality checks performed on each field in ICES data collections. However, this recommendation extends to request that quality indicators can be added to multiple data collections, potentially by multiple groups – although the current recommendation only reflects the intention of one group. DIG expressed some concern about the potential scale of the task, and has formed a small group of members of DIG and ICES Data Centre to engage in a dialogue with WGBIOP to ensure the recommendation can be realised, but also consolidated with current direction of development of systems and quality approach (Action 1).

The recommendation from WGEGBS2 had already been completed by the ICES Data Centre by the time of the DIG meeting, so required no further action.

A number of DIG members took part in the 2017 WKSEATEC meeting, and are intended to also engage with the WKSEATEC 2018 meeting. There is already an ongoing dialogue about the developments and intentions of creating more universal data exchanges for electronic measurements and data capture for fisheries data. DIG will explicitly maintain the dialogue, and also ensure alignment with current or ongoing activities in the ICES Data Centre (Action 2).

3.1.2 Actions

DIG has had a busy year with a range of inter-sessional activities. The DIG Action list has been used to track both actions arising from the previous years' main meeting as well as inter-sessional work. The table below summarises all actions 2017–2018, along with status or report section describing progress.

Number	Action	Addressed to	Status
1	Update framework text following additions mentioned in Annex 7 (of 2017 report)	Ingeborg de Boois	Completed. See report section 9
2	Investigate if Data mining of ICES library tool is useful to specify relations between ICES groups and the data portals those groups use. Specific task: to link the acoustic data portal working groups to the correct reports. Furthermore, the ICES database use by the various ICES Integrated Ecosystem Assessment groups may be investigated	Sjur Ringheim Lid, Ingeborg de Boois, David Currie, Jens Rasmussen, Carlos Pinto	Updated and merged with new work See report section 5.1
3	Install a DATRAS governance group, consisting of DIG representative (Ingeborg de Boois), WGBIFS, IBTSWG, WGBEAM representatives (tbd), ICES Data Centre (Anna Osypchuk, Vaishav Soni)	Jens Rasmussen, Ingeborg de Boois, Neil Holdsworth	Completed. See report section 2.2.1
4	Sort out how access restriction is organised when a scientist changes jobs leading to different access permissions to restricted data	Ingeborg de Boois, Neil Holdsworth	Completed. Managed by secretariat and National representatives.
5	Consider how to best Combine 'dataset collections' and 'data portals' at ices.dk/marine data	Jens Rasmussen, Neil Holdsworth	Completed See report section 5.8
6	Create standard checklist for the development of new ICES data portals	Jens Rasmussen, David Currie, Hjalte Parner	Ongoing See report section 9
7	Circulate IODE Meeting report to DIG Members	Lesley Rickards	Completed
8	Arrange an ICES Data Centre contact person for the guidelines shared with IODE Ocean Data Practises	Neil Holdsworth, Taco de Bruin	Completed. ICES now has a generic contact email for this.
9	Advance review of CTD, Bottle, and biological plankton guidelines – locate experts to review.	Taco de Bruin, Hjalte Parner, Lesley Rickards, Else Green, Ruth Lagring, Friedrich Nast	Updated See report section 3
10	Complete resolution for forming WKINVITED and follow up on arranging Workshop and promoting it	David Currie, Sjur Ringheim Lid	Completed. (2017 resolutions)

Number	Action	Addressed to	Status
11	Take forward proposal for alternative ICES ASC session at ICES ASC 2018 in Hamburg	Jens Rasmussen, Malin Werner, Christian von Dorien, Ingeborg de Boois	Completed. Additional planning during this meeting to firm up on tasks. 2 New Actions on the 2018 meeting list to complete the planning (Actions 12 & 13)
12	Intersessional group to monitor GDPR implementation and potential impact on ICES Data Policy	Taco de Bruin, Christian von Dorien	Completed See report section 7
13	Intersessional evaluation of effort required versus benefit of storing seabed litter data recordings from non-trawl survey sources (e.g. Video data)	Helge Sagen, Neil Holdsworth	Completed. WGML now established. New action for DIG chair to contact working group to raise topic (Action 17).
14	Intersessional discussion/investigation of provenance and traceability. Can data be traced back to the originators? Both on metadata and record level. Also include considerations on DOI reference integration.	Jens Rasmussen, Neil Holdsworth, Friedrich Nast	Updated. See report section 8
15	Build and maintain a list of issues, considerations and potential changes to ICES Data Plan and Policy to aid in the next review (2019)	Jens Rasmussen, Neil Holdsworth	Ongoing See report section 7
16	Consult with Steering group chair of EOSG on membership of DATRAS governance group.	Ingeborg de Boois	Completed
17	DIG to recommend format for Data standards publication type for ICES (requests to publish data standards documents has currently gone in as CRR, but it is recognised that it's not a good fit)	Jens Rasmussen, Neil Holdsworth	Completed (Forum correspondence with SCICOM) Also see report section 3 on data guides
18	DIG and Data Centre to draft document of future challenges for Marine Data Management in ICES	Jens Rasmussen, Neil Holdsworth, Consultation with all DIG	Completed document. Follow up activities see report section 4

3.2 Feedback from other working groups

DIG receives reports from expert groups and workshops that are closely linked to the work on data, or are formally reporting to DIG. This year, the newly created DATRAS Governance group, the WKINVITED hackathon, and WGSFD provided updates during the DIG meeting.

3.2.1 DATRAS Governance Group

The DATRAS Governance Group (DGG) has organised members in 2017 and met twice by web conference in 2018 prior to the DIG meeting. The tasks of the governance group has been refined and clarified during these meeting to allow the group members to communicate with their respective survey groups.

In addition, the DGG has also provided updates to the wording of the governance approach that DIG and associated Governance Groups and governance exercises will work through to evaluate systems.

In the following year, the DGG will do a more in depth evaluation of the DATRAS workflows following the governance framework approach. This will in turn be condensed into an assessment with identifiable issues/actions that will help move the overall data collection governance to an improved position. The full report from DGG is available in Annex 6.

3.2.2 WKINVITED

The first ICES hackathon workshop concluded just as the DIG meeting started, and overlapping presentations of the products and ideas developed during the event was presented directly to DIG. The WKINVITED workshop will produce its own report with recommendations and lessons learned from the experience.

Overall DIG was impressed with the amount of work achieved in WKINVITED, and subsequently discussed the merits in repeating the event, taking into consideration the forthcoming report from WKINVITED. However, other organisations are also progressing hackathons, and during the DIG meeting, it emerged that EMODnet are considering a joint hackathon event in 2019, where ICES could potentially be a partner. The benefits of a joint hackathon are that more resources can be provided for participants as well as demonstrating wider integration in data products. However, the final details and agreements of such an event would still have to be organised and agreed, mainly between the ICES Data Centre and EMODnet. DIG felt that it would be a positive way to try out a joint exercise, but also stressed that it would be important to maintain a momentum within ICES, even if the joint event did not go ahead.

Thus two actions were identified for the ICES Data Centre and EMODnet to determine if a joint hackathon will go ahead (Action 3). At the time of writing, ICES has received an official invitation from the EMODnet Secretariat to join the joint hackathon in 2019 (also with Copernicus). As a precautionary measure, a resolution for a new ICES workshop format hackathon has been prepared only to be submitted if the joint event does not progress (Action 4). The decision on whether to progress with a joint hackathon concept or a workshop will be finalised for the September SCICOM meeting, so a single suggestion can be brought forward.

3.2.3 WGSFD

The working group on Spatial Fisheries Data (WGSFD) reported on the recent process of making data calls for VMS and logbook data from member countries. This type of data is confidential and requires separate data handling and data policy in order to effectively bring the source data together and create the aggregated products that are used for advice.

The nationally submitted data are subjected to a Quality Control (QC) process that allow reporting back to submitters in case of issues, and ultimately is drawn together in an overview that allow expressions of confidence in the ICES advice that is produced.

The workflow of controlled and protected data submissions, QC and advisory product development is an exemplar of effective collaboration between an Expert Group and the ICES Data Centre, where the process has been developed in collaboration and matured rapidly. It is however resource intensive for both national data submitters and the ICES Data Centre, indicating that handling of confidential and controlled data is an elaborate process which requires planning and resource to realise. This point may be of relevance to other working groups considering processes for handling data that cannot be submitted or shared openly.

From a DIG perspective, the Data Policy agreement for handling VMS and logbook data is separate from the more generic data policy for open/accessible data. Given that the general data policy will undergo its planned review in 2019, DIG felt that it would make sense to also review the VMS/Logbook data policy at the same time to bring them into an aligned review cycle. The review of the VMS/Logbook Data policy will be added to the DIG agenda 2019 along with the general data policy review (Action 5).

4 Data Guidelines

DIG is an operational group, but has historically evolved from a former expert group that was focussed on Oceanographic data management. Thirteen Data Type Guidelines for physical-chemical-biological data types were developed by that former group (WGMDM). DIG has automatically carried an element of ownership of these data type guidelines, and it does make sense for DIG to be involved in ensuring that descriptions and guidance is available for types of data dealt with in the ICES community. However, with the change of focus to a more strategic level within DIG, combined with a diversification of membership, the availability of members to review guidelines has reduced. A sub group reported back from the past year, identifying key issues that were discussed in plenary:

- An inventory of databases and data types within ICES should be prepared and compared with existing guidelines.
- Existing Guidelines were reviewed in 2006, so is overdue for a review to reflect changes and updates to methods and equipment used to capture data.
- A comparison with the UK MEDIN guidelines would potentially reduce the workload as these have been reviewed more recently, and there is a degree of overlap.
- References to the IOC's Group of Experts on the Technical Aspects of Data Exchange (GETADE) are out of date, and should be removed. GETADE was merged with the JCOMM/IODE Expert Team on Data Management Practises in 2005.
- Review of text for website had been carried out, and minor updates will be circulated
- The ICES Data Type Guidelines were submitted to the IDOE Ocean Best Practises Repository. One of the requirements was to provide a single point of contact for queries. The ICES Data Centre has created a mailing list (guidelines@ices.dk) that will act as the contact email, and will be distributed to relevant staff members.

In addition to the review by the sub group, DIG was also requested to make recommendations on publication processes for data format guides in the future on the SCICOM forum. The general recommendation was that the data guideline format should be used for describing data formats as well as collection and submission processes. However, it was recognised that the existing format (originally created in 1999) requires changes to accommodate this.

Other working groups, predominantly WGFASST, are developing standards and recommendations for data or metadata, that are submitted, but do not fit well into the current process. In addition, it is recognised that work on updating and revising formats and guidelines does not stand still, and there is a need to enable more dynamic work on such guidelines, while ensuring that ICES only formally recognises and publishes content when it has been appropriately reviewed or quality audited.

DIG decided to not progress reviews of the existing data type guidelines for the time being. It is recognised that they still need to be available as they are referenced in various other organisations, including HELCOM. However, in light of the growing calls for a more dynamic approach, and more diverse collection of data, format, and processing guides, DIG prioritised work on defining a more flexible structure/format for data guides as well and emphasis on the process.

There was not adequate time during the meeting to fully write out all aspects of this, so a DIG sub group will work to prepare a document covering format, process, and how to handle more dynamic development without impacting on ICES reputation (Action 6). This document will then be circulated for general discussion/acceptance in DIG between meetings. A progress report will be provided to SCICOM in September, while the final recommendations will be reported to SCICOM at the mid-term March meeting 2019.

5 Future Data Challenges and Opportunities

DIG and ICES Data Centre collaboratively submitted a paper to SCICOM in February 2018 outlining the three main future challenges and opportunities for managing data in the ICES community¹, being:

- Machine Learning
- Cloud environments and related services
- Open data and code sharing

DIG discussed each of these topics, seeking to specify more concrete examples of both challenges and opportunities. It was generally recognised that in order to structure and monitor these topics, a risk-based management format would be well suited, both for tracking and updating the issues within the ICES community as well as for simpler and more approachable reporting to the rest of the ICES community.

A template for using a risk-based management approach was drawn up based on the feedback from sub-groups identifying challenges and opportunities in each of the main topics. The headings of the risk-based management template are described in the table below while the initial populated matrix will be set up on the ICES SharePoint system for assigning actions and monitoring/updating entries (Action 11). It should be stressed that the matrix is currently a draft, and that the WKMLEARN report was not yet available, so will likely add further topics to the matrix. The matrix will predominantly be used by DIG to perform a routine update and monitoring activity, which will be reflected in the Terms of Reference for the group.

Matrix Heading	Explanation
Priority Topic	Which of the three priority topics the line is relating to. As far as possible, this will be constrained to one topic, but there will be issues that clearly span across multiple topics.
Concept Description	The main concept/capability that is being addressed.
Challenges	Challenges associated with the concept (if any).
Opportunities	Opportunities associated with the concept (if any).
Existing Activities	Current or ongoing ICES activities that are of direct relevant to the concept
Recommendations and actions	More regularly updated section clarifying who is monitoring or carrying out any actions relating to this concept.
Likelihood	A simple 1 – 3 score that demonstrates the expected likelihood of this concept having an effect on ICES work. Where 1 = Long term (5+ years), 2 = Medium term (3–5 years), and 3 = Short term (within the next 1–2 years).
Impact	A broad descriptor of the perceived impact of the concept directly on ICES Data Centre and services. 3 = Complete rebuild of existing service or need to bring in new expertise, 2 = Substantial re-engineering of existing solutions, or need to consult external expertise as part of projects, 1 = Some

¹<http://ices.dk/community/groups/Documents/Future%20data%20challenges%20and%20opportunities%20in%20ICES.pdf>

Matrix Heading	Explanation
	changes likely, but mainly in demand or smaller tweaks to existing solutions.
Change Score	Product of Likelihood x Impact. So a score from 1-9. 1-2 considered low effect, 3-4 medium and 6-9 high effect. DIG will review and update these scores and raise issues that are escalating in risk or reward. It should be stressed that the scores reflect potential for both positive and negative effects. The higher the score, the bigger the change and potential work required.

Where specific activities or concepts indicate a high change score, DIG will aim to report and make recommendation to alleviate risk and maximise benefit.

6 ICES Data Centre Update

The update on current and future activities in the ICES Data Centre is a recurring item of the agenda of DIG meetings. It provides an opportunity for DIG members to gain and overview of current priorities and for ICES Data Centre staff to obtain advice on best practise or governance related issues to specific developments. The sections below summarises discussions around presented activities.

6.1 Transparent Assessment Framework (TAF)

The approach for developing TAF to enable location and reproduction of assessment processes through a data-model-output workflow is a strong step towards a structure approach for some of ICES' main advisory products. The framework has been presented to DIG on previous occasions, but the updates demonstrated a clear approach towards a well-considered process model that is flexible enough to cope with the realities of diverse data sources while moving increasingly towards a service-oriented architecture. DIG recognised that TAF would be an ideal candidate for profiling with the DIG governance checklist, and a sub group of DIG and ICES Data Centre staff will carry out this exercise in the current calendar year (Action 7).

6.2 Regional DataBase and Estimation System (RDBES)

An update on the development of a new Regional DataBase and Estimation System indicated that several workshops had been taking place over the year, now resulting in a near finalised data model. The new system will bridge functionality from the prior Regional DataBases to InterCatch, where two separate data calls were necessary to first provide detailed data, and subsequently nationally raised data. The new RDBES will simplify the process and data calls will be reduced to a single call, while estimation will be performed based on reviewed and accepted scripts directly embedded in the system. The REDBES is identified as linking up to the TAF framework as well, providing a big potential for a more streamlined process. The current project aim is for completion in summer 2019, but it was recognised that some aspects of the project to date has taken longer to finalise. It will however significantly change the system capabilities, since the new RDBES will accommodate different sampling designs and allow estimation processes to be linked directly to these designs.

DIG queried how long the transition period would be moving from the current RDB + InterCatch solution to the new RDBES system. It was recognised that not all submitters would be able to move immediately to the new solution, but currently the length of the transition period is not specified. DIG recommended that a clear transition period should be set out to allow countries as much time to prepare as possible, while avoiding long periods of running systems in parallel (Recommendation 1).

6.3 Acoustic data portal and output challenges

This presentation covered the integration of acoustic data into the ICES Acoustic Data portal (acoustic.ices.dk). There are well-defined workflows, and the data model for the Acoustic Data portal integrates WGFASST SISP-4 metadata convention for processed acoustic data from active acoustic systems.

The biggest issues experienced around the current workflow and documentation is the alignment between publication and reviews of data formats or conventions. The existing processes are too slow for development, but there is still a strong desire to retain the ICES review process/quality stamp on the products. This aspect of the presentation heavily influenced DIG's approach towards data guidelines in report section 3.

6.4 Eutrophication Tools

The presentation demonstrated the work between ICES Data Centre and HELCOM to develop an integrated approach between data systems and assessment tools, specifically for eutrophication relevant parameters in this instance (nutrients, chlorophyll, oxygen, etc.). The approach married large volumes of ICES held data with external data sources also needed for assessment, such as satellite data. The process also encapsulated the indicator data extraction and development with multiple acceptance levels before the finalised products could be loaded onto the HELCOM assessments database. This is the culmination of a long process spanning from 2014–2018, but is now recognised as an effective and well-documented process that links up with the data sources.

The ICES Data Centre will shortly be working with OSPAR to apply the same methodology (with minor alternations) initially to a limited test area, with scope for expansion if successful. Similarly, ICES Data Centre is working with EEA on Core Set Indicators (CSI's) on Nutrients, Chlorophyll a, and Dissolved Oxygen, adopting and reutilising similar approaches. Fundamentally, the approach adopted for the various eutrophication tools are increasingly aligning with the TAF approach, and strategically, it is welcome that ICES Data Centre is developing multiple projects along very similar workflows, even if alterations for specific contexts may still be necessary.

6.5 Marine Aggregates Extraction

ICES Data Centre is working with the Working Group on the Effects of Extraction of Marine Sediments on the Marine Ecosystem (WGEXT) to develop a database of volumes of marine sediment extracted and the associated licenses to do so. Previously, WGEXT have worked on this data in spread sheet solutions, and had initially drafted a database within the group. Now the ICES Data Centre has supported the group and developed a fully-fledged database and revised data submission standard. Initially, the database has been presented and received good feedback and suggestions from the working group. Next stage is to import historical data, and work towards adoption of the new data submissions format. Ultimately, the database will support web services and mapping tools as the product develops.

DIG asked about linkage to other requests for wider marine spatial planning tools, and the ICES Data Centre is aware that other groups are working on related topics. Generally, it is recognised that the database will either need to be aggregated into or linked via services to a wider marine spatial planning data system. However, this will be a gradual development – but the awareness and ability to link the products together are there. Anticipated requests for information on MSFD indicator 6 for assessing disturbance is also likely to impact and drive the development towards wider aggregation and standardisation.

6.6 New generation of vocab.ices.dk

Behind this title is perhaps one of the most substantial changes in ICES Data Centre handling of reference data for a long time. The revised vocabulary services now feature mapping of semantic concepts, advanced search, direct linkage to code types and vocabulary terms, as well as the ability to download lists of related codes. While many data users and submitters in the ICES community may not notice any immediate changes on the basis of this update, it has dramatically altered ICES' capability for creating and exposing linked open data. With a single service call, developers and data analysts can now retrieve ICES' vocabularies and their inter-linkage in a simple knowledge organisation system (SKOS) as well as resource description framework (RDF) formats. An exercise is already underway with the British Oceanographic Data Centre (BODC), who maintains similar vocabulary structures for UK Marine organisations and SeaDataNet, to link the vocabularies across the organisations.

DIG recognised the importance of this development, but also that many users in a ICES community is less familiar with semantic solutions, and the potential power that can be leveraged from analysis and linkage of data with these services. As such, DIG agreed to merge existing work on Text mining of ICES report with a wider action to develop 2-3 concrete use cases that demonstrate how data location and/or extraction with the ICES data holdings could be leveraged and improved with the utilisation of the new vocabulary services (Action 8).

6.7 Reports Data Mining

The presentation of the ICES Reports Data Mining Tool demonstrated capabilities of using a "Bag-of-words" model that enable searching across a large amount of reports to find expressions and connections between e.g. concepts and working groups. DIG is exploring the capabilities of the report data mining tool, and have previously extracted lists of associations between working group reports and data portals held in ICES. This work is now being rolled together with the use case examples for the newly updated vocabulary services, but will also seek to utilise the report data mining tool (Action 8).

6.8 New ICES Data Portal

The ICES Data Centre presented their early plans for rebuilding the ICES data portal. There is recognition that the existing portal does not support new data types, and that there is a discontinuity between the portal and the thematic portals from a user perspective. The revision of the data portal is adopting a service-oriented architecture, leveraging more unified service exchanges between thematic databases and the main data portal. This will enable users to search and explore the data more directly as well as access downloads – realising the FAIR principles of all data on the portal (and in ICES) being Findable, Accessible (where appropriate), Interoperable (through services), and Reusable. A key requirement is also to be able to demonstrate full INSPIRE compliance, so that future ICES requests for advice or data developments can be delivered in fully INSPIRE compliant formats without clients needing to perform additional transformations. ICES Data Centre has also identified the need to be able to provide access not only to data, but also indicators, and advisory products.

Overall, the presentation was positively received, and the service oriented architecture approach is encouraged by DIG. Questions on how tightly the new product will incorporate metadata, persistent identifiers, and potential external data linkages were relevant to feed into the process, but too early to provide concrete answers to.

DIG will review the current document outline for user stories, and collate comments back to ICES Data Centre (Action 9)

6.9 Bycatch (of protected species)

ICES Data Centre has created a database for recording by-catch of protected endangered and threatened species, working with the Working Group on By-catch of Protected Species (WGBYC). A data call was issued, and a template was provided to facilitate submissions. The main function of the template is to transform data from a more familiar spread sheet or table format to an XML format that can be easily ingested in the database. Some issues were encountered during the data call with submitters not completing all mandatory fields or using reserved characters in fields that would cause the transformation to XML to break. However, after working with submitters from 18 countries, the majority of issues have been ironed out, and ICES Data Centre will review the current format to ensure better alignment in submissions for future data calls. The approach of working with expert groups to introduce improved data formats, thus enabling better quality control, is leading to a major improvement for the working group. They are now able to work off a single centralised copy of data, reducing risks of variations between analysts, and it allows the working group to focus on quality control and consistency rather than technical problems.

DIG queried how the by-catch database of protected species aligns with the current information about by-catch in the Regional DataBase systems and the new RDBES. The answer was that currently, data are extracted from the RDB to provide effort estimations. DIG observed that this is leading to slight duplication of data, and that in the longer run, the by-catch data should form part of RDBES, allowing the current by-catch system for protected species to only store indicators instead. However, it is recognised that there are different organisations involved in the two workflows, and that it may require some time to fully merge the workflows. There is currently communication between WGBYC and the RDBES Core Group on this issue. DIG recommended that the potential requirements from the WGBYC database should be added to the functional requirements of RDBES for consideration (Recommendation 2).

6.10 European Seabirds at Sea (ESAS)

ICES Data Centre is working with the Joint Nature Conservation Committee (JNCC) in the United Kingdom, who currently holds a database of aerial and ship based surveys of seabirds at sea (e.g. not at breeding sites). ICES were contacted in 2017 to explore the possibilities of taking over hosting and development of the data system. ICES Data Centre has been working to make an online version of the data portal. The governance structure for the data portal has been established and will be overseen by the joint working group on birds (JWGBIRDS). The Data Policy for the portal will fit within the overall ICES Data Policy, however like the VME and Biodiversity data portals, there will be the concept of restricted records as noted on the data policy landing page. DIG recommended to take the current plans through the governance framework evaluation

to ensure all relevant aspects of the portal development and associated data management has been covered (Action 10).

6.11 SmartDots (Otolith Annotation)

The SmartDots tool for testing and annotating images of otoliths allow cross-organisation comparison and consistency checks. The project is now an international online platform for age reading workshops and calibrations consisting of a database, an open source software package, and a reporting tool. The database is developed by ICES Data Centre, while the SmartDots software itself is developed by ILVO in Belgium, and the R-based reporting tool has been developed by DTU Aqua in Denmark.

The system allows coordinators to set up events that will allow age readers to work through a set of images of otoliths and enter their readings. Reports on consistency between age readers can subsequently be produced to help look for problem areas of inconsistencies. Future plans for further development include integration of a previous legacy system (WebGR), facilitation of maturation calibration events, exploration of otolith shape analysis, and allowing machine learning algorithms to take part in the ring trial events.

The current application is for registered users only, and since it is supporting consistency tests, this makes sense.

DIG queried if the image-based tool was applicable for all types of otoliths. In Belgium, all otoliths are now analysed from images, and the Netherlands are heading in the same direction. Neither have yet experienced any issues with using image based reading for any type of otoliths.

Follow on discussions about the current structure of the ICES Marine Data section of the website also concluded that the revision of the Data portal would effectively replace and update the current structure. Thus, DIG did not see any need for making recommendations on changes to the current organisation, since the revision and restructuring will be tightly coupled with this redevelopment.

7 Dark Data Recovery and the role of ICES

DIG tabled a discussion of dark data recovery activities, raised by the Working Group on Zooplankton Ecology (WGZE) who are actively investigating the recovery and archiving of historic data into recognised repositories along with documentation and metadata.

Dark data are “dormant” data that are not currently used or shared in a meaningful way. The data may be in digital or analogue formats, but are not forming part of a standardised, quality controlled collection or archive.

Several DIG members have been directly involved in data recovery or archaeology projects, and recognised the important and potentially huge value of dark data held in various institutes. When discussing Dark Data, the term covers historical data that are not publically available, or part of a recognised archive or repository where the data can be requested.

WGZE has a specific ToR to examine the potential recovery of dark data prior to the WGZE time series that are held on the NOAA COEPOD database, and requested input from DIG on the role of ICES in the process.

DIG members identified several external initiatives that either provide funding or support to the recovery of dark data. In particular, the EMODNet ingestion programme also contains options for recovery of historic data.

ICES have also done a number of recovery projects over the years, including historic plankton data and stomach data. The ICES Data Centre clarified that there is a process in ICES to work on this, but it will mean that standards and QC needs to be enforced, which may sometimes be difficult for dark data where there are no staff with direct knowledge of data collection left in organisations.

DIG observed that the Working Group on the History of Fish and Fisheries (WGHIST) has been very active in identifying historical data sets/dark data, and have been utilising the ICES Metadata catalogue to register historical datasets. 95 historic fish related datasets are registered on the metadata catalogue.

DIG recommends that WGZE also register the dark data/historic data they identify on the ICES metadata catalogue (Recommendation 3). While DIG recognises that registering the metadata will not necessarily mean that the data becomes immediately accessible or recovered into standard data formats, it will identify the datasets, and create a list that would allow potential future collaborative projects on data recovery to rapidly identify datasets that could be included.

8 Data Legislation updates, Data Policy, and Impact on ICES

DIG members investigated potential impacts of legislative changes on ICES as part of an inter-sessional sub-group. The presentation during the DIG meeting and subsequent discussion centred on the EU's General Data Protection Regulation (GDPR), and evaluating if, and where, there may be aspects of data processing at ICES that contains information relating to an identifiable living person within the EU or EEA who can be directly or indirectly identified. Further, that clear consent is obtained and explanation of the reason for holding or collecting the data is provided to individuals, along with information on its intended use and how long it will be kept for. While ICES has an international status, the GDPR is applicable if data on EU and EEA citizens are being processed, so there is still a likely impact and legal obligation on ICES.

ICES did circulate privacy statements to committee members and national delegates, and made them available on the ICES website, with separate privacy statements for meetings, events, and recruitment.

Thus, in terms of handling personal information relating to the work of ICES, there seems to be adequate coverage. However, DIG did query if the use of web services were logged and stored, and whether data calls need to expressly contain statements on purpose and retention, where even pseudonymised data are included.

DIG recommended that the ICES Secretariat develops and maintains frequently asked questions (FAQ) about the position on data submission matters – especially where data may contain personal information. DIG does not suggest that ICES necessarily takes on responsibility for large volumes of personal data, but that it might be easier for national data submitters to reference a FAQ page when processing data and conferring with national data controllers (Recommendation 4).

It is not envisaged that GDPR legislation will require modification of the existing data policy. However, the ICES Data Policy is due for review in 2019, and the topic will be included in the consideration. DIG members were also encouraged to raise or add any other legislative or policy related changes that may impact on the review of the ICES Data Policy (if any). A list of relevant points to consider during the review will be established.

9 Data Traceability and Provenance in ICES

As the catalogue of data products and data collections are growing within the ICES community, the ability to link between a derived product and the associated data becomes more important. ICES Data Centre are already working hard on this concept, and many of the new and existing developments, such as TAF, RDBES, and vocabulary updates are enabling this very linkage.

However, there is a parallel question about the traceability to national submissions. A query to the DIG members indicated that virtually all organisations represented were either actively publishing data with persistent identifiers themselves, or investigating and considering the possibilities. In most instances, there are not currently large overlaps between the data submitted to ICES, and the data published with DOI, but there were some instances. In addition, it was also recognised that some data centres hold data for which they create persistent identifiers, and at least parts of these datasets were also submitted to ICES.

Overall, ICES is progressing well on enabling linkage and traceability within own systems. But there is currently no ability to take in or register links back to national submitters' persistent identifiers. This means the same data may be submitted or presented in multiple places without the ability to recognise them as identical sources.

There are still questions about the level of granularity at which persistent identifiers are applied to datasets. Some countries may opt for a single DOI for a survey that both contains fisheries and oceanographic data, while another country may choose to mint separate DOI's for the two subsets. However, in either case, it means there would be a trace back to metadata, and contact information relating to the submitted data. As such the granularity may present a challenge, but will not prevent the ability to recognise and incorporate persistent identifiers in the submitted data formats.

DIG did recognise that it would not be feasible to alter existing data submission formats only to capture persistent identifiers, but stressed that the ability to provide this traceability is becoming more important as more members adopt publishing and persistent identifiers themselves as well. The final recommendation from DIG was that ICES Data Centre should consider incorporating the ability to submit persistent identifiers for data into all new or modified data submission formats (Recommendation 5), but not as a mandatory requirement (since not all member countries will develop the capabilities at the same pace). It was further agreed that ICES Data Centre would provide a list of data systems with a status on provenance and traceability of nationally submitted persistent identifiers (Action 14).

DIG also discussed that national submitters and ICES itself will enhance potential metadata and standards compliance by adopting persistent identifiers. For example, to mint a DataCite DOI, a particular schema of metadata must be submitted, and it was recognised that there would be overlaps with e.g. INSPIRE metadata requirements, which may assist in improving both traceability and profiling of data at the same time. However, there was not time during the meeting to do a full profile comparison between the DataCite and INSPIRE metadata schema. This comparison will instead be carried out after the meeting (Action 16).

10 Data Governance evaluation and reporting

DIG has over the past two years developed and refined a checklist of topics to evaluate data management and governance of systems or work processes. The list contains a number of topics highly relevant for data management; and is aligned with the Data Management Association's body of knowledge. However, it is also recognised that the terminology of this assessment method is not particularly approachable for non-specialists. DIG investigated different options on how to collate and condense its governance framework evaluations into a more user-friendly format.

Subgroups evaluated possible mapping to the FAIR principles of Findable, Accessible, Interoperable, and Reusable concepts, but found that many of the more technical data management categories mapped poorly into these principles. The FAIR principles are centred on making data available and possible to work with, while the DIG governance framework spans wider, also addressing more technical issues like development standards, technical architecture etc.

Instead, the FAIR principles evaluation was added to the DIG checklist for the governance framework approach. It was agreed that the governance framework topic list with slightly revised wording and the additional category for FAIR added will only be used internally in DIG for the technical evaluation of systems and work flows.

The DATRAS Governance Group had also discussed the governance framework and provided minor revisions to wording to help make the intention of different evaluation categories more clear. The updated wording and categories of the governance framework can be found in Annex 7.

The DIG chair agreed to draft a single line summary format for the evaluations instead. The draft format will be tested and refined when several systems are evaluated in the coming year (DATRAS, TAF, ESAS) (Action 15).

11 Joint ASC Session with PGDATA

DIG received a question from PGDATA about whether it would like to contribute towards a theme session on the subject of Data Quality – this would be a follow-up to a previous PGDATA convened 2016 ASC session called “When is enough, enough?” Methods for optimising, evaluating, and prioritising of marine data collection.”²

This was discussed and although the session did indeed sound interesting it was felt that DIG could not contribute any extra expertise beyond that already provided by PGDATA.

² <http://www.ices.dk/news-and-events/asc/ASC2016/Pages/Theme-session-O.aspx>

Annex 1: List of participants

Name	Institute	Country (of institute)	Email
Anna Osypchuk	ICES	ICES	Anna.Osypchuk@ices.dk
Carlos Pinto	ICES	ICES	carlos@ices.dk
Chris Moulton	OSPAR	UK	Chris.Moulton@ospar.org
Christian von Dörien	Thuenen-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
Else Juul Green	ICES	ICES	else.green@ices.dk
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.parnar@ices.dk
Ingeborg de Boois	Wageningen Marine Research	The Netherlands	ingeborg.deboois@wur.nl
Jens Rasmussen	Marine Scotland	UK	Jens.rasmussen@gov.scot
Malin Werner	SLU	Sweden	malin.werner@slu.se
Marcellus Rödiger	Thuenen Institute	Germany	marcellus.roediger@thuenen.de
Medhi Abbasi	ICES	ICES	Mehdi.Abbasi@ices.dk
Neil Holdsworth	ICES	ICES	neil.holdsworth@ices.dk
Örjan Bäck	SMHI	Sweden	orjan.back@smhi.se
Peter Wiebe	WHOI	US	pwiebe@whoi.edu
Ruth Lagring	Royal Belgian Institute of Natural Sciences	Belgium	ruth.lagring@naturalsciences.be
Sara Almeida	Hydrographic Institute	Portugal	sara.almeida@hidrografico.pt
Signe Bakker	ICES	ICES	signe.bagger@ices.dk
Simon Claus	VLIZ	Belgium	simon.claus@vliz.be
Sjur Ringheim Lid	IMR	Norway	sjur.ringheim.lid@imr.no
Susanne Tamm	BSH	Germany	susanne.tamm@bsh.de
Taco de Bruin	NIOZ	The Netherlands	bruin@nioz.nl
Wim Allegaert	ILVO	Belgium	wim.allegaert@ilvo.vlaanderen.be

Annex 2: Agenda

Wednesday 30 May 2018

11:30 – 12:00: Arrival, Welcomes & Introductions, Safety

12:00 – 13:00: WKINVITED presentations and judging

13:00-13:30: WKINVITED outcome, recommendations, and future considerations of hackathon(s)

13:30-14:00: Short lunch break

14:00-15:00: Review action list (some items directly on agenda – others brief update during this slot)

15:00-15:30: Updates from SCICOM meetings

15:30-16:00: Coffee break

16:00 – 16:15: Activity update: WGSFD (Christian)

16:15-16:45: Activity update: Data Guidelines progress (Taco, Hjalte, Else, Ruth) (*Action 9*)

16:45-17:30: Data Future Challenges paper (*Action 18*) – Introduction + Breakout groups for:

- Machine Learning
- Cloud Infrastructure
- Open Data/Code sharing.

17:30-17:50: Plenary from breakout groups

17:50-18:00: Plenary. Wrap up actions & reporting tasks for the day

Thursday 31 May

09:00-13:00: ICES Data Centre Update (including plenary)

13:00 – 14:00 Lunch Break

14:00 – 15:00: Current activities and coordination (Breakout groups)

- DATRAS Governance Group (*Action 3,16*)
- Data's Den Open session at ASC (*Action 11*)
- ICES Data Mining Tool and Group linkage (*Action 2*)
- WKINVITED + WLMLEARN. Lessons and recommendations (*Action 10*)

15:00-16:00: Plenary on breakout groups

16:00-16:30: Coffee break

16:30 – 17:15: Breakout groups: Future activities

- Dark Data, Recovery, and ICES role (Peter Weibe).
- ICES Data publication structure: Datasets vs. Portals etc. (*Action 5*)
- Joint ASC session 2019 with PGDATA (Dave Currie)
- WKASMSF Guidance/outcomes.

17:15 – 17:45: Plenary from Breakouts.

17:45-18:00: Plenary. Wrap up actions & reporting tasks for the day.

Friday 1 June

09:00 – 09:30: GDPR and other legislative drivers. (*Action 12*) and look ahead to Data Policy Revision/Review

09:30 – 10:30: Provenance progress (*Action 14*)

13:00-12:30: Governance framework – finalise recommendations and actions.

12:30 – 13:00: Next meeting dates and inter-sessional activities roundup.

13:00-14:00: Breakout groups & finalising report sections

- Highlights & social media

14:00: Meeting close

Annex 3: DIG terms of reference for the next meeting

The **Data and Information Group (DIG)**, chaired by Jens Rasmussen, United Kingdom, will meet at ICES, Copenhagen, Denmark, from 21–23 May 2019 to:

- 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 by 24 June 2019 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) & b)</p> <p>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 c)</p> <p>By monitoring and reviewing, legislative drivers, and the ICES Data Policy, there is a clear review process that ensures data policies and processes stay up to date and reduces risk for the organisation as a whole.</p> <p>Term of Reference d) & e)</p> <p>By reviewing and monitoring current and future activities, DIG can proactively advice ICES on emerging issues. From time to time, identification of issues will lead to the formation of small groups for governance, or workshops to allow adequate time to tackle specific challenges and opportunities. DIG will increasingly adopt a risk-based management approach to monitoring, review and reporting of these topics.</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 PICES, GOOS, SeaDataNet/SeaDataCloud, EMODNet, IOC and its Working Committee on International Oceanographic Data and Information Exchange (IODE), OSPAR, and HELCOM

Annex 4: Recommendations

RECOMMENDATION	ADDRESSED TO
1. 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
2. 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,
3. 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
4. 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
5. 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

Annex 5: Actions

Number	Action	Addressed to	Deadline
1	Initiate dialogue with WGBIOP to clarify recommendation and understand potential scale of work	Ingeborg deBoois, Neil Holdsworth, Carlos Pinto	September 2018 + update May 2019
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
3	Discuss and decide if a joint hackathon event between EMOD-Net and ICES (and potentially other partners also) can go ahead	Simon Claus, Neil Holdsworth	End August 2018
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
5	Include review of VMS/Logbook Data policy in the DIG 2019 agenda/programme of work	Jens Rasmussen	Mar 2019
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
7	Complete a governance framework profiling of TAF	Jens Rasmussen, Christian von Dorrien, Colin Millar, Arni Magnusson	December 2018
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
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
10	Complete a governance framework profiling of ESAS	Ingeborg de Boois, Neil Holdsworth	December 2018
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	September 2018

12	Advertise and encourage submission of pitches to the Data's Den open session during ASC 2018. Specifically contact ICES Comms team, WKINVITED, WKM-LEARN to encourage participation	Mail Werner, Christian von Dorrien, David Currie	July 2018
13	Identify judging panel for the Data's Den open session during ASC 2018	Malin Werner, Neil Holdsworth, Christian von Dorrien	August 2018
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 2019
15	Draft a single line reporting format for the outcome of governance framework evaluations	Jens Rasmussen	December 2018
16	Compare and map out overlaps between DataCite and INSPIRE metadata schemas	Ruth Lagring	October 2018
17	DIG Chair to write to chair of WGML to highlight potential other sources of macro litter data than trawl data.	Jens Rasmussen	September 2018

Annex 6: DATRAS Governance Group (DGG) Report

The DATRAS governance group (DGG) was installed in January 2018 and consists of ICES Data Centre (Anna Osypchuk, Vaishav Soni), representatives from the main working groups supplying data to DATRAS – IBTSWG, WGBIFS, WGBEAM (Finlay Burns, Henrik Degel, Wim Allegaert) and DIG (Ingeborg de Boois, chair).

The group meets approx. four times a year by web conference. At DIG three members of the group met to further work on actions in progress.

1.1 Tasks

The tasks of DGG are:

1. Elaborate the framework on the governance of DATRAS;
 - a. Discuss and fill in the DIG governance framework for DATRAS (trawl survey data as well as litter data)
 - b. Provide suggestions to ICES Data Centre for implementation of the improvements
2. Oversee and advise on the interpretation and prioritisation of recommendations from expert groups addressed to DATRAS;
 - a. Based on the compilation of recommendations from other expert groups and committees, with a focus on synergy and coherence of similar re-requests for products and services.
3. Align DATRAS over the different surveys;
 - a. Align DATRAS input formats for the surveys where possible
 - b. Align QC and QA protocols between the surveys where possible
 - c. Align DATRAS CA input format with Acoustic data portal biological data format where possible
 - d. Align products for the DATRAS surveys where possible
4. Provide a platform for end user feedback to the DATRAS system.
 - a. Seek and collate feedback from end users of DATRAS via interaction in working groups and committees, targeted questions, through the ICES websites, or feedback given directly to the Data Centre
 - b. Provide responses to the end user feedback, and create recommendations to the relevant entities if a follow-up action is appropriate and practical

1.2 Progress

Related to task 1:

DGG has made slight amendments to the governance framework as developed in DIG.

Related to task 2: in progress, but no urgent issues

Related to task 3:

- a. WGBIFS, IBTSWG and WGBEAM have been asked for feedback on the current input Exchange format. Based on that, DGG develops an update of that format. Baseline is to limit the number of input formats. When the proposed new format is ready, the data

submitters as well as the chairs of WGBEAM, WGBIFS, IBTSWG will be asked for feedback and additions.

b. Feedback from WGBEAM, IBTSWG, WGBIFS that arose during their respective meetings has been noted and will feed into this item.

c. Will be taken into account in the proposal for the new format

d. Feedback from WGBEAM, IBTSWG, WGBIFS that arose during their respective meetings has been noted and will feed into this item

Related to task 4:

a. Feedback from end users in WGBEAM, IBTSWG, WGBIFS that arose during their respective meetings has been noted and will feed into this item.

b. No action

1.3 Further Actions for 2018

September 2018: discuss the governance framework for DATRAS. Also take into account:

- the GDPR, e.g. for commercial vessels owned by one person hired by institutes for the survey;
- provenance, and the possibility to store DOIs from data submitting institutes as follow-up of DIG discussion

Before end 2018: propose new DATRAS input Exchange format, send it to data submitters and relevant survey group chairs for feedback.

Annex 7: DIG Data Governance Framework Categories

Topic Area	What is included
Architecture and governance	Understanding integration and linkage between underlying data, data products and associated working groups in ICES
Data Development	Updates including versioning to structures, formats, products and calculation methods of data either as requirements arising from new use cases or legal requirements
Database Operations	Understanding the ICES responsibilities in terms of maintaining databases versus data coming from outside or other data providers.
Data Security	Ensure that you can enable appropriate access to data and prevent inappropriate access, including user management. This also touches on potential limitations on data use and/or further dissemination.
Reference and Master Data Management	Identify the authoritative copies of the master data and understand where shared references codes are used and who maintains and develops these. This should include versioning the copies.
Data discoverability	How data are made available for sharing and integration through presentation within the ICES working community, more broadly on websites, and how different types of users need to interact with the data.
Document and content management	How documents, guidelines, web services, web interfaces and other unstructured ³ content are maintained. This should include versioning of documents.
Metadata management	How well data structures and information is profiled via metadata. This links to both legal compliance obligations (e.g. INSPIRE) and improvements in data sharing and citation (e.g. minting DOI for reports, datasets etc.).
Data Quality Management	Consideration of how quality of data, data products and calculation methods is managed for the given collection. Responsibilities may be shared between expert groups and data centre, and the key thing is to understand how decisions on quality management are made, and how they align across ICES data handling.
Compliance with FAIR principles	Findable Accessible Interoperable Reusable

³ Unstructured in this context simply means that the content is not part of a relational database – it can still be well organised.