

ICES WGDIM REPORT 2009

ICES SCIENCE COMMITTEE

ICES CM 2009/SCICOM:03

REF. SCICOM, ACOM

Report of the Working Group on Data and Information Management (WGDIM)

26–28 May 2009

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. 2009. Report of the Working Group on Data and Information Management (WGDIM), 26–28 May 2009, ICES Headquarters, Copenhagen, Denmark. ICES CM 2009/SCICOM:03. 26 pp. <https://doi.org/10.17895/ices.pub.8726>

For permission to reproduce material from this publication, please apply to the General Secretary.

The document is a report of an Expert Group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the views of the Council.

© 2009 International Council for the Exploration of the Sea

Contents

Executive summary	1
1 Opening of the meeting.....	2
2 ToR a – User Engagement	2
3 ToR b – Availability and Accessibility	3
4 ToR c – Quality and Transparency	5
5 ToR d – Interoperability	5
6 ToR e – Data progress	6
7 ToR f – Year of the Stomach	6
8 ACOM Proposed ToR – VMS data	7
9 Other Work	8
Annex 1: List of participants.....	9
Annex 2: Terms of Reference (ToRs) 2008/2009	11
Annex 3: Agenda.....	14
Annex 4: WGDIM terms of reference for the meeting in 2010.....	17
Annex 5: Recommendations	19
Annex 6: Agreed Actions.....	20
Annex 7: Ideas generated from the audience at the ASC 2008	22
Annex 8: Response to VMS ToR.....	23

Executive summary

WGDIM was brought into being in 2007 as an amalgamation of the Working Group on Marine Data Management (WGMDM) and the Study Group on Data and Information Management (SGDIM). It has been evolving over the last 3 years and refining its role within ICES. The meeting this year saw a significant shift away from detailed technical discussions towards more strategic themes covering the majority of ICES' data holdings and systems.

User interaction with ICES Data Centre and WGDIM has been improved as a result of the Theme Session at ASC and the much improved ICES web based data services. All of this is solidly underpinned by the Data Policy.

With increased use of data across multiple disciplines the provision of clear guidance on all aspects of data, from its initial capture through to long-term curation and use, are essential, the group currently maintains a number of guidelines under the ICES banner and will continue to develop, recommend development of or implementation of relevant standards.

1 Opening of the meeting

The 2009 meeting of the Working Group on Data and Information Management (WGDIM) was held at ICES Headquarters, Copenhagen from 26 to 28 May 2009. The meeting was attended by 24 scientists (Annex 1) and members of the ICES Data Centre. The agenda (Annex 3) was adopted. The terms of reference for the meeting are given in Annex 2.

The working group thanks ICES (Vivian Piil) for making the local arrangements and support during the meeting

2 ToR a – User Engagement

- a) Assess progress with, and update, users engagement plan - the plan developed in 2008 sets out a number of activities and pilot projects - these will be evaluated and the plan updated to ensure full user engagement and appropriate development of products;

Discussion during the meeting regarding this ToR covered the interfaces between ICES' data consumers (expert groups, individual scientists, outside organisations such as HELCOM, OSPAR etc) and ICES as a data provider. Discussion also explored the relationship with data submitters.

Presentations were given during the course of the meeting by members of the ICES Data Centre (DC) highlighting progress on a variety of applications, database systems and data products. The DC has made considerable progress over the last year and the group commends the efforts of the all the staff involved. Of particular note were the progress reports regarding Platform Codes, the EcoSystemData inventory module, and implementation of GIS.

A subgroup was tasked with evaluating and providing feedback on the current ICES web site and data portal and the plans for their redevelopment. This group included a number of members of the DC and was considered to be very productive. The comments made are already being fed into the design and implementation processes.

A presentation was given regarding the ASC 2008 theme Session which included a novel interactive session. The session attracted 25 presentations and a further 12 posters. The interactive section was very well attended and provided a range of ideas and proposals for future work (Annex 7), some of these are being actioned presently and their completion will reinforce with users that interacting with ICES, specifically WGDIM and the DC, can produce products and changes they need.

Important conclusions from the theme session were that:

- End users, regardless of purpose, demand more than a simple delivery of data to their computer, there is a need for visualization and GIS type products.
- There is a continual and urgent need to address the number of standards, best practices, and interoperability procedures

The theme session was so productive that the group agreed to propose a further theme session, provisionally titled 'Data for the Masses', to be included in the 2010 ASC.

At ASC 2009, Berlin, the DC will be presenting 'Data Centre Live' where users can interact directly with members of the Data Centre, extracting data, providing data products and discussing the associated issues.

The Data Centre E-news bulletin was launched in 2008 and is working towards its 4th edition. Short articles highlighting data and information management are welcomed.

New actions under this ToR, detailed in Annex 6/items: 12, 15, 16, 19, 22.

3 ToR b – Availability and Accessibility

- b) Identify major gaps in data availability or data accessibility, including legacy data, in the ICES data management system or data needed but not currently held at ICES;

Presentations were made by G Evans (BODC), D Johnson (NOAA), F Nast (BSH), P Wiebe (Woods Hole) and members of the DC (Maria Zarecki, Else Green, Marilyn Sørensen).

Data flows to and from the data centres represented within the group have continued as expected, there has been a notable shift in the types of instruments being used for Oceanographic data. The shift is from 'classical' devices such as Bottle Casts and MBT towards newer technologies such as Gliders (>1700% increase between World Ocean Database (WOD) 2005 and WOD 2009).

Access to data at ICES (EcosystemData) and World Data Center for Oceanography, Silver Spring (WDC), amongst others, is becoming very user-oriented with flexible spatial selection tools and quick delivery of data. Data quantities served are expanding, for example to March 2009 WDC has served 1007Gb of CTD data back out to the scientific community.

A presentation on 'Rolling Deck to Repository (R2R)' highlighted the stream of real-time and 'not-quite real time' data that are being passed directly from the point of collection to data repositories. This is an area where visible guidelines, Quality Control and Assurance (QA / QC) procedures are vital for end users of the data, the group agreed to review the level of this R2R activity within member institutes to enable appropriate levels of guidance to be given.

The discussion in the group has, due to its history and membership profile, been significantly slanted towards physical and chemical oceanography, the group agreed on the need to expand discussion to include all data types that ICES holds or is involved with in future data ToRs.

The group considered the nature of this ToR and felt that the task of identifying gaps in data should be addressed by the ICES expert groups that are providing advice, assessments or products based on ICES data holdings. Experts working with the data are more likely to be able to identify, prioritise, and suggest sources for additional data. WGDIM will continue to address data accessibility issues by providing guidance on distributed databases, data provision and emerging technologies; this conclusion is reflected in the suggested ToRs for 2010.

Gaynor Evans reported on the "Data flow to BODC current meter inventory". The International Current Meter Inventory (ICMI) application was launched on 7 March 2008, and there have been several hundred visits to the web site since then. There were no new metadata submitted to the international current meter inventory in 2009, but she encouraged the submission of new data sets.

Regarding data flow to ICES, there were three reports by data centre personnel. Maria Zarecki briefly described the requirements for submitting fisheries data to the data centre. For the most part, data contributors are getting the data into ICES by set deadlines. These data are going into DATRAS. There were some difficulties screening the incoming data and this was requiring some extra time doing the work by the DATRAS team.

Else Green said the hydrography data provides a different set of problems because while OSPAR and HELCOM are required to submit data, the submissions by national institutes, individual scientists and others are done on a voluntary basis. Those countries that submit data generally do so on a yearly basis (e.g. Norway, France, Sweden, and Poland) and the ICES data centre general maintains the data sets in similar format as recorded. The data sets include temperature, salinity, pressure, oxygen, fluorometry, and nutrients. There has been some decline in new submissions, but the data centre is expecting an increase in getting the data into the database. It was emphasized that this is very much dependent on voluntary efforts and there needs to be more encouragement for these efforts.

Gaynor asked what could be done to encourage people to submit data? Else said ICES will be proactive in chasing the data. The CSRs will enable them to chase down the data because they have information about when and where CTDs were taken.

Marilynn Sørensen discussed the input of environmental data into the environmental database, which includes a large number of parameters (~860) and a number (7) of data types. Monitoring data are provided on a yearly basis from the HELCOM and OSPAR areas and data supporting working group activities comes from national institutes. Legacy data are a point of concern. Conversion software now exists, but not for all HELCOM data. Various versions of the data need different processing. A number of issues with phytoplankton data were described. For example, error checking software has been highlighting lots of errors and when ICES asks the submitters to provide fixes, users become discouraged, even if number of fixes might be small. Other issues highlighted include the fact that the Phytoplankton Expert Group (PEG) biovolume list does not have all the species for which data are submitted and HELCOM guidelines are not always followed when submitting data. There are also some issues with zooplankton data submission. Not all species are covered by size to volume algorithms, and it is hard to know where the conversions are coming from. Size is apparently being reported in different ways and there needs to be standardization. There have been discussions about these issues between HELCOM and the ICES data centre, and an action plan has been formulated to deal with them. In addition for the need to have better communication between the HELCOM data contributors and the Data Centre, there may be ways to use Web Services technology to improve the input of the data at the point of input. It was recognized that the old data will never be as high quality as new data coming in, and tools are in place and guidelines exist, but they need to be utilized more effectively.

Marilynn said there have been significant improvements in data submission from OSPAR data and she was optimistic that the problems in HELCOM could in part be resolved by having standardized station and platform codes. One suggestion was that a workshop be held bringing together representatives from the HELCOM countries to discuss the current problems and seek consensus for their resolution.

New actions under this ToR, detailed in Annex 6/items: 5, 6, 7.

4 ToR c – Quality and Transparency

- c) Identify and resolve issues related to transparency, traceability and quality (use of data quality indicators) of data in relation to their use at ICES to formulate advice;

Presentations were made by G Evans (BODC), T. DeBruin and members of the DC.

The WGDIM guidelines are currently labelled with the group name, it was agreed that they should really be labelled as ICES Data Centre Data Guidelines to reinforce their application across the ICES community.

The group discussed the wide range of standards and guidelines developed within and outside the ICES community. Within ICES there is no central repository, each guideline sits within the workspace of its owner (Expert groups etc). It was agreed that a central repository or clearing house for ICES guidelines would greatly enhance and promote their use. Once guidelines are centralised WGDIM could also easily review the data curation sections of these and provide useful guidance to the guideline owners. There are several EGs that have ToRs involving development of guidelines, databases and data formats, they could all benefit from easy access to existing documents with WGDIM providing overviews of best practice.

The group agreed that getting sensor manufacturers to provide references to existing international guidelines relating to their instruments would be beneficial to everyone in the field.

The SeaDataNet QC manual is planned for release later this year, the WGDIM CTD questionnaire which was to be based around the released manual has not taken place.

Quality flag systems are in place for some but not all of ICES' data holdings, WGDIM recommend a flag system be adopted or developed that could be applied across all data holdings within ICES. This flag system should be able to harmonise with projects that ICES is already involved in e.g. SeaDataNet.

New actions under this ToR, detailed in Annex 6/items: 1, 2, 3, 4, 8, 9, 10, 11, 18, 20, 21.

5 ToR d – Interoperability

- d) Identify and promote relevant standards for metadata, data structures, dictionaries and data dissemination in the ICES data management system; develop recommendations for interoperability between the ICES data management system and relevant international data management bodies and programmes (e.g. PICES, IOC/IODE, GOOS, SeaDataNet, International Polar Year) to ensure rational and optimal endeavour;

Presentations were given by N Holdsworth, F Nast and members of DC.

The IDOE Standards process was outlined as an example of a highly controlled standards process. This does result in robust widely agreed standards but has a considerable time-lag from proposal for a standard to the publication of an agreed one.

Platform codes are an integral part of marine data management, ICES is now the central clearing house for platform codes and a considerable amount of work has been done to resolve issues in the existing code list. A new online platform code system is in development and due to be released in May 2009.

6 ToR e – Data progress

- e) Review the input from other Expert Groups on the request of WGDIM;

Although WGDIM seeks interaction with other EGs within ICES it became clear that the group's role and sometimes even its existence is not widely known. Several items were brought forward at the theme session for WGDIM's attention and 1 further item directly from another EG (ToR f).

WGDIM needs to publicise itself to the other EGs to improve the intergroup cooperation.

WGDIM will liaise with WGMEG to develop a workplan aimed at bringing the Mackerel Egg Database online.

New actions under this ToR, detailed in Annex 6/items: 14, 22.

7 ToR f – Year of the Stomach

- f) Work towards making the ICES 'Year of the Stomach' datasets for North Sea and Baltic more readily available to the ICES community. This will require the creation of a standardized and quality-controlled version of the data including an updated look-up key for prey codes.

Background

A request had been received from WG SAM asking for help from WG DIM to recover data collected by the Year of the Stomach Project (1981 to 1990). This data is not held centrally or available easily, the task is to create a single dataset, QC check the data and then publish it.

Progress

The Data Centre has collated what is thought to be the complete dataset; it consists of 1.4 million observations from 11 countries, breaking down further into 8 predator species and 854 prey codes. The prey codes are in NODC format and need conversion to ICES compatible form. The conversion has highlighted a code issue; 47 NODC codes do not appear in the ITIS species list.

It is not clear if the dataset compiled by DC contains all the records for the project.

Data summaries have been made available online to allow expert review. (<http://ecosystemdata.ices.dk/stomachdata/index.htm>)

During the meeting a subgroup discussed progress and agreed to liaise with N Daan. Niels is considered to be a leading authority on the project and is likely to be able to confirm the expected number of records, may be able to supply any missing ones and could provide guidance on the prey code conversion.

Contact will also be made with the contributing countries to ascertain the completeness of the ICES dataset and whether it (the ICES dataset) is the only digital version available.

New actions under this ToR, detailed in Annex 6/items: 23, 24.

8 ACOM Proposed ToR – VMS data

- a) The current and pending legal status of VMS data in the ICES area, including any issues that may hinder data use for scientific purposes and considering the status of VMS data in relation to the present and revised EU Data Collection Regulations
- b) Estimates of data quantities that can be expected from Vessel Monitoring Systems, including a consideration of VMS data temporal resolution in relation to its potential scientific uses, and any proposed or required changes in temporal resolution
- c) Investigation of organisations within the ICES area that currently archive or intend to archive VMS data and which allow access for scientific purposes
- d) A summary of what scientific tools are being developed by existing research programmes, including EU Framework projects, to analyse and interpret VMS data
- e) Proposals for how ICES scientists and Expert Groups should gain access to VMS data in the future and what data interface, interrogation, display, analysis and interpretation tools ICES should obtain or develop.

This ToR was proposed late in the year, the chairs were unsure whether the group had the expertise to fully address the issues raised. A subgroup was tasked with reviewing the ToR and proposing further action, the full report of the subgroup is provided at Annex 8. In summary:

WGDIM does not have the relevant legal expertise to respond in anything other than a lay perspective.

It is easily conceivable that VMS records will extend to many millions per member state per year.

Collection of VMS data, but not the long-term archival of such data, is a statutory requirement for EU member states, the group was not in a position to review non-EU ICES members.

It is known that a number of national institutes are developing methods to manage and filter VMS data for scientific and fishery management purposes; however, a full inventory was not available to WGDIM. One research programme was highlighted as being of specific relevance to ICES; Under the European Commission open call for tenders (No MARE/2008/10 - Studies for carrying out the common fisheries policy), the EU invited tenders for an 18 month commercial contract to develop tools for log-book and VMS data analysis.

WGDIM was not in a position to comment authoritatively on the tool-set needed in relation to other VMS-based advisory requests that ICES may receive, other than to note that such requests are current and cannot afford to wait for the outcome of the Lot 2 project in circa 18 months time.

In order to aggregate enough experts in this field WGDIM proposes a short term study group to address issues D and E.

WGDIM also recommends that the ICES Data Centre maintains a watching brief during the development of the Lot 2 study (subject to approval of the consortium members and the EC) and to identify how best ICES can accommodate any tools/algorithms that are developed as may apply to its advisory remit.

9 Other Work

WGDIM was brought into being in 2007 as an amalgamation of the Working Group on Marine Data Management (WGMDM) and the Study Group on Data and Information Management (SGDIM). It has been evolving over the last 3 years and refining its role within ICES. A subgroup examined the work of WGDIM in relation to perceived needs within ICES and presented proposals for its future work. The group examined the strengths and weaknesses of WGDIM, what the group does well and what the group should improve on.

WGDIM has a large number of 'clients' (SCICOM, ACOM, the Data Centre, Expert Groups, Workshops etc.), it provides advice and guidance across multiple disciplines (data policy, data management, data strategy, physical oceanography, fishing surveys, plankton surveys, chemical oceanography etc).

WGDIM members are involved in all stages of the data lifetime, from initial collection through quality assurance, providing the data back to the scientific community, data analysis and onwards into metadata management and long-term data curation.

What can WGDIM offer ICES? The answer can be split into 4 areas, each equally important to the long-term success and scientific prowess of ICES.

Strategy

Draft the ICES Data Strategy 2011–2015

Provide guidance to the ICES Data Centre

Provide input to the ICES Science Plan (2014)

Policy

Keep oversight of and provide revisions to the ICES data policy

Provide guidance on implementation of the ICES data policy

Advice on other data regulations and their impact on ICES

User Guidance

Provide guidelines for quality assurance methodologies

Recommend, develop and provide advice on the use of quality flag systems

Technical

Provide advice on data management guidelines

Promote new technologies

The subgroup suggested the mission for WGDIM should be:

"To provide ICES with advice on all aspects of data management including data policy, data strategy, data quality, technical issues and user-oriented guidance."

Annex 1: List of participants

Name	institute	email	country
Pekka Alenius	Finnish Meteorological Institute	pekka.alenius@fmi.fi	Finland
Sara Almeida	Instituto Hidrografico	sara.almeida@hidrografico.pt	Portugal
Richard Ayers	Centre for Environment, Fisheries and Aquaculture Science	richard.ayers@cefas.co.uk	United Kingdom
Ingeborg de Boois	Wageningen IMARES	ingeborg.deboois@wur.nl	Netherlands
Garry J. Dawson	UK Hydrographic Office	garry.dawson@ukho.gov.uk	United Kingdom
Taco de Bruin		bruin@nioz.nl	Netherlands
Henrik Degel	National Institute of Aquatic Resources	hd@aqua.dtu.dk	Denmark
Mia Devolder	Royal Belgian Institute of Natural Sciences, Management Unit of the Mathematical Model of the North Sea	M.Devolder@mumm.ac.be	Belgium
Gaynor Evans	Natural Environment Research Council	gaev@bodc.ac.uk	United Kingdom
Lars Hansson	Swedish Meteorological and Hydrological Institute	Lars.Hansson@smhi.se	Sweden
Mirko Hauswirth	Bundesamt für Naturschutz	mirko.hauswirth@bfv-vilm.de	Germany
Neil Holdsworth	International Council for the Exploration of the Sea	neilh@ices.dk	Denmark
Georgiy Moiseenko	Russian Federal Research Institute of Fisheries & Oceanography	georgem@vniro.ru	Russian Federation
Friedrich W. Nast	Bundesamt für Seeschifffahrt und Hydrographie	friedrich.nast@bsh.de	Germany
Riitta Olsonen	Finnish Environment Institute	riitta.olsonen@ymparisto.fi	Finland
Hjalte Parner	International Council for the Exploration of the Sea	hjalte@ices.dk	Denmark
Daphne R. Johnson	National Oceanographic Data Center	Daphne.Johnson@noaa.gov	United States
Helge Sagen	Institute of Marine Research	helge.sagen@imr.no	Norway
Elena Tel Perez	Instituto Español de Oceanografía	elena.tel@md.ieo.es	Spain
Peter Wiebe	Woods Hole	pwiebe@whoi.edu	United States

	Oceanographic Institution		
Christopher Zimmermann	Johann Heinrich von Thünen-Institute, Federal Research Institute for Rural Areas, Forestry and Fisheries	christopher.zimmermann@v ti.bund.de	Germany
Liam Caffrey	Liam Caffrey Marine Institute	liam.caffrey@marine.ie	Ireland
Martin Mørk Larsen	National Environmental Research Institute	mml@dmu.dk	Denmark

Annex 2: Terms of Reference (ToRs) 2008/2009

2008/2/SCICOM03 The Working Group on Data and Information Management [WGDIM]
(Chairs: H. Sagen, Norway and R. Ayers, UK) will meet at ICES Headquarters Copenhagen, Denmark from 26–28 May 2009 to:

- a) **Assess progress with, and update, users engagement plan** - the plan developed in 2008 sets out a number of activities and pilot projects - these will be evaluated and the plan updated to ensure full user engagement and appropriate development of products;
- b) **Availability and accessibility** - identify major gaps in data availability or data accessibility, including legacy data, in the ICES data management system or data needed but not currently held at ICES;
- c) **Quality and transparency** - identify and resolve issues related to transparency, traceability and quality (use of data quality indicators) of data in relation to their use at ICES to formulate advice;
- d) **Interoperability** - identify and promote relevant standards for metadata, data structures, dictionaries and data dissemination in the ICES data management system; develop recommendations for interoperability between the ICES data management system and relevant international data management bodies and programmes (e.g. PICES, IOC/IODE, GOOS, SeaDataNet, International Polar Year) to ensure rational and optimal endeavour.
- e) **Data progress** - review the input from other Expert Groups on the request of WGDIM;
- f) work towards making the **ICES 'Year of the Stomach' datasets** for North Sea and Baltic more readily available to the ICES community. This will require the creation of a standardized and quality-controlled version of the data including an updated look-up key for prey codes.

WGDIM will report by 1 June 2009 for the attention of SciCom and ACOM.

Supporting Information:

Participants:	The 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 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.
Secretariat facilities:	<p>a) Action Plan 5.13.4, 6.1, 6.4; b) Action Plan 6.1, 6.4; c) Action Plan 4.12, 6.1, 6.4; d) Action Plan 1.10, 5.13.4, 6.1, 6.4.</p> <p>a) It is essential to ensure needs of users are met: there are a wide range of users for ICES data and products from HELCOM/OSPAR to WGs/SGs and individual scientists. WGDIM should have a key role to act as a mediator between Users/WGs and the Data Centre to prioritise activities, to ensure appropriate experts are available and to give reasons for priorities. Proper engagement with users will allow data submission problems to be resolved and integrated data products and thus advice can be provided in an appropriate form.</p> <p>b) There are major gaps in the ecosystem assessments apparently caused by lack of data. However, more data are likely available for use than currently perceived either inside the ICES system or externally. Thus, groups developing the advice may not be aware of the existence of relevant data sets either because of a lack of communication or the fact that data are not being delivered on a timely basis. In addition, those environmental assessments that are now being produced by some ICES working groups are not being effectively utilized by other groups making assessments where environmental data should be considered (NORSEPP, WGRED)</p> <p>Conclusions: i) Communication between ICES expert groups needs to be improved. ii) Data contributors need to be encouraged to submit data when they are useful, not when they are completely quality controlled.</p> <p>c) Much of the data that are being used to make the environmental assessments do not reside within the ICES and little effort is being expended to track the data used to make the assessments. If the external data are being used to formulate advice, it is often difficult to later re-establish the data sets and thus the basis for the advice. Thus the group should provide advice as to how improve this reporting.</p> <p>d) To maximize interoperability data quality must be known. It is important to evaluate the appropriateness of use of data for specific applications on the basis on data quality. Coordinate work with relevant working groups or projects like SeaDataNet, ECOOP, etc. on standards for metadata, data/data structures and vocabularies. As there is limited resource available it is essential to avoid duplication of work on data management. It is thus important to engage in collaboration with international bodies and programmes especially when the 4th International Polar Year is in progress and is seeking help and guidance on data management.</p> <p>e) This is in response to a request made by SciCom.</p> <p>f) This is in response to a request made by WGSAM, Resource Management Committee.</p>
Financial:	None
Participants:	The Group is expected to be attended by some 30–35 members and guests with half of the members from each of the two categories , data managers and data users
Secretariat facilities:	Meeting facilities.
Financial:	The Data Centre Manager should attend these meetings together with other employees at the data centre.

Linkages to advisory committees:	Report is seen by ConC and all science and advisory committees
Linkages to other committees or groups:	Oceanography and Advisory Committees.
Linkages to other organizations:	There are linkages with relevant international bodies and programmes like PICES, IOC/IODE, GOOS, SeaDatanet, IPY, etc., with emphasis on IOC and its Working Committee on International Oceanographic Data and Information Exchange (IODE).

Annex 3: Agenda

Tuesday 26 May – Rapporteur Garry Dawson

0900–0930	Opening greetings and aims for the day	[Co-chairs]
	Welcome by ICES representative	
	Local arrangements	[V. Piil]
0930–1030	Review meeting schedule and items for discussion	
	Appoint rapporteurs for meeting	[H. Sagen]
	Review action items from last year's WGDIM meeting	[R. Ayers]
1030–1100	Coffee break	
1100–1200	Quality and Transparency	[T. De Bruin]
	CTD Questionnaire	
	SeaDataNet QC manual	[G. Evans]
1300–1430	Lunch	
1430–1545	Data Centre Update, Status, Upcoming Activities	[N. Holdsworth]
	SCICOM presentation	
	EMODNET, pilot projects update	
	MSFD, ICES activities	
	Data Centre 'Live' at the ASC 2009	
	IODE Standards	
	Stomach data latest update	
	Highlights from Working group link-ups	
	GEO-BON	
1545–1615	Coffee break	
1615–1745	Agree topics, membership and focus (ToR) for subgroup work	
	(3 groups maximum); groups already suggested are listed below, other suggestions are welcome	
	VMS	
	Marine XML	
	Meta-Data Standards within open source GIS	
	Year of the stomach database	
	ICES Website / Data Portal	
1745–1800	Summary of Day 1	[Co-chairs]

Wednesday 27 May – Rapporteur Gaynor Evans

0900–0930	Summary day 1, action items next year/ ToRs	[G. Dawson]
	Opening notes, aims for the day [Co-chairs]	
0930–1000	User engagement plan – Latest version	[G. Evans]
1000–1020	User Engagement plan – Review, update and develop actions For the coming year.	
1020–1030	Theme session report ASC 2008	[P. Wiebe]
1030–1100	Coffee break	
1100–1115	Theme session proposal ASC 2010	[P. Wiebe]
	Report from IODE GE-BICH (Biological and chemical experts group)	[N Holdsworth]
1115–1145	Update on Oceanographic data system in the Data Centre To include CSR & Platforms	[Data Centre]
1145–1230	Presentation on status of EcosystemData EcosystemData Steering group issues	[Data Centre]
1230–1300	Interoperability	[F. Nast]
	Data update in SeaDataNet	[F. Nast]
1300–1430	Lunch	
1430–1500	WDC Silver Spring status/dataflow/new developments	
[D. Johnson]	Data progress – review of input from other groups	[Co-chairs]
1500–1545	Plenary discussion Publicity, future role of the group looking towards the new SCICOM.	
1545–1615	Coffee break	
1615–1800	Subgroup work	
	WGDIM Strategy	Biscay [R Ayers]
	VMS data	North Sea [P. Kunzlik]
	ICES web site/ data portal	Atlantic [N. Holdsworth]

Thursday 28 May – Rapporteur Peter Wiebe + (Subgroups to provide text for report)

0900–0915	Summary of Day 2	[T. De Bruin]
0915–0930	Opening notes, aims for the day	[Co-chairs]
	Election of Co-Chair	
0930–1030	Data availability and accessibility	[G. Evans]
	Data flow to ICES	[Else, Maria, Marilyn]
	Stomach data	
	Data flow to BODC current meter inventory	[G. Evans]
1020–1050	Coffee break and group photo	
1050–1110	Report and Discussion Subgroup Strategy	[R. Ayers]
1110–1200	Report and Discussion Subgroup VMS data	[P. Kunzlik]
1200–1245	Report and Discussion Subgroup Web site	[N. Holdsworth]
1245–1300	Report and Discussion Subgroup Stomach data	[N. Holdsworth]
1300–1315	Summary of Day 3	[P. Wiebe]
1315–1415	ToRs for next year 2010	[Co-chairs]
1415–1430	Next meeting and closure	[Co-chairs]

Annex 4: WGDIM terms of reference for the meeting in 2010

The **Working Group on Data and Information Management** (WGDIM) chaired by Helge Sagen, Norway, and Richard Ayers, UK, will meet in Copenhagen, Denmark, in late May to Early June 2010 to:

- a) **Data accessibility** – Provide advice on the functionality of the new integrated ICES data portal throughout the year. Deliverable: Review of data portal, recommendations for enhancements. (**Responsible focal point:** Chris Zimmerman)
- b) **Quality, interoperability and transparency** - Identify and resolve issues related to the use of quality flags in ICES Data Management, specifically in the areas of Biological and Chemical data, with particular reference to existing international quality flag systems or those that are envisioned as needed (e.g. SeaDataNet, EMODNET, HELCOM and OSPAR etc..) Deliverable: Guideline for development/adoption of a quality flag system suitable for application across ICES data holdings, including actions to harmonise across existing systems. (**Responsible focal point:** Gaynor Evans)
- c) **ICES Data Strategy** – Draft the 2011–2015 Data Strategy in line with ICES Strategic Plan 2009 – 2013. Identify emerging technologies that ICES Data Centre, WGDIM members and the wider ICES Community should be aware of (GIS metadata). Deliverable: Draft data strategy document, Report on emerging technologies and possible applications with ICES. (**Responsible focal point:** Richard Ayers / Helge Sagen)
- d) **VMS** – Review actions resulting from recommendations of WGDIM 2009, review progress of Lot 2 project and other developments in the VMS arena. Deliverable: Report on actions, Lot 2 progress, proposals for further work; (**Responsible focal point:** Phil Kunzlik)
- e) **Assess progress with, and update, user engagement plan** – Review the success of the ICES Data Centre Live at the ICES ASC, undertake planning of agreed workshops (e.g. mackerel egg) and propose new workshops or activities. Deliverable: Outcomes from ASC 2009, updated user plan with follow up actions, detailed workplan for workshops, further proposals. (**Responsible focal point:** Pekka Alenius)

WGDIM will report by 1 August 2010 for the attention of SCICOM.

Supporting information

Priority	The 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 Group flies the flag for ICES in setting standard 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	
Resource requirements	None
Participants	The Group is expected to be attended by some 30–35 members and guests with half of the members from each of the two categories, data managers and data users

Secretariat facilities	Meeting facilities.
Financial	The Data Centre Manager should attend these meetings together with other employees at the Data Centre.
Linkages to advisory committees	ACOM
Linkages to other committees or groups	
Linkages to other organizations	There are linkages with relevant international bodies and programmes like PICES, IOC/IODE, GOOS, SeaDataNet, IPY, etc., with emphasis on IOC and its Working Committee on International Oceanographic Data and Information Exchange (IODE).

Annex 5: Recommendations

Recommendation	For follow up by:
1. ICES to convene a short-term group to provide expert advice regarding VMS data, its storage, access and tools for analysis.	ICES

Annex 6: Agreed Actions

1	Label the WGDIM guidelines as ICES Data Centre Data Guidelines	DataCenter	30 Sept 2009
2	Obtain sensor manufacturer contact details, if possible from their technical support staff, and supply to G Dawson	All	30 Sept 2009
3	Contact the sensor manufacturers and inform them of the existence of ICES Data Centre Data Guidelines and encourage their dissemination and use	GD	31 Dec 2009
4	Supply to H Sagen details of the existence of national standards relating to oceanographic data. (e.g. British Standards Institute (BSI) and Norwegian Standards Institute (NSI))	All	30 Sept 2009
5	Ensure other data types are covered within ToRs and agenda items for future meetings (i.e. the group's focus will widen to include not just oceanographic, but also e.g. nutrient, chemical, and trawl survey data)	RA	Next Meeting
6	Recommended the R2R project refer to the ICES Data Centre Data Guidelines and implement where applicable.	PW	30 Sept 2009
7	With R2R project in mind, compile a list of institutes that are sending data direct to the relevant data repository.	TdeB	30 Oct 2009
8	Provide information to ICES Data Centre on the systems for Quality Control flags in use for all data types (including biological and chemical data)	All	30 Oct 2009
9	Recommend a flag system for currently unflagged data (biological, chemical, nutrient etc) that ICES can use and that harmonises with Sea Data Net's recommendations	NH/WGDIM	Next Meeting
10	Report to Working Group on Oceanographic Hydrography the concerns noted by WGDIM members regarding the potentially significant effects of different CTD operating methods.	HS , HP and T De B	Before next WG Oc Hydro meeting
11	Check that the CTD Data guidelines contain guidance on including, in the metadata, details on the procedures used when water samples are collected	GD	30 Sept 2009
12	Provide WGDIM report 'snippets' for inclusion in Data Center E-News	RA	in time for publication
13	Submit article to 2009 Insight on CSRs and new vocabularies.	FN	in time for publication
14	Chairs to write to Mackerel group with view to developing a workplan aimed at bringing the Mackerel Egg database online in late 2010	Chairs	Before next WGMEG meeting
15	Propose further WGDIM sponsored theme session at ASC 2010, use existing proposal as a basis, keep the ICES/PICES joint convenors	NH	Before deadline for ASC 2010 proposals
16	Prepare list of people to encourage to make a contribution (scientists and data managers) as soon as the 2010 theme session has been approved. Richard to collate list.	All (RA to collate)	

18	Make contact with GEBICH and pursue possibility of cooperation regarding Quality flags (related to action 9).	Chairs
19	Facilitate liaison between the ICES platform codes project and OceanSITES regarding mooring-codes.	T De B
20	Liase with World Ocean Data Centre regarding the algorithms and application of quality control to oceanographic data.	Hjalte
21	Design and build infrastructure to support an ICES wide guideline clearinghouse.	DataCenter
22	Contact ICES EGs, informing them of WGDIM's remit and request information on guidelines produced by the expert group.	Chairs
23	Contact N Daan regarding 'Year of the stomach' data set; completeness and code systems	NH/IdB
24	Contact 'Year of the stomach' data submitters to ascertain status of data in institutes and expected number of records in ICES dataset	NH/IdB

Annex 7: Ideas generated from the audience at the ASC 2008

- Publish survey data on web through an interface to the ICES DC. The first example would be the Mackerel Egg Survey data that is collected during an international and ICES-coordinated multi-ship effort, and that is quality controlled, but so far not accessible.
- Create standard products according to clients needs (OSPAR as an example).
- Develop ODBC interfaces for all data held under ICES auspices, aiming at making the production of products by users themselves easier; also create a set of standard views to ease access to data subsets for scientists not familiar with SQL.
- Harmonize quality flags, develop a standard at least for ICES coordinated surveys and fisheries data.
- Retrieve historic, non-digital data (metadata first) – or during a workshop, develop a strategy to retrieve these data.
- Give training courses in SQL and data handling for non-data administrators.

Annex 8: Response to VMS ToR

The following text is based on a response by Marine Scotland (Science) to a number of questions circulated by the WGDIM Co-Chairs to UK colleagues prior to the WGDIM meeting. It can be considered an initial response to some of the meeting's late-breaking meeting ToRs. The original response has been augmented by WG discussions.

The current and pending legal status of VMS data in the ICES area, including any issues that may hinder data use for scientific purposes and considering the status of VMS data in relation to the present and revised EU Data Collection Regulations.

A layman's view of the EU legislation makes this reasonably simple to interpret, but WGDIM does not have the relevant legal expertise to raise this response above that of a lay perspective. The EU VMS regulation (COMMISSION REGULATION (EC) No 2244/2003) provides for satellite monitoring data to be collected for vessels > 15m with the following access conditions under Article 14:

"Access to data:

1. Member States shall ensure that the Commission has, on specific request, remote access by online sessions to the computer files containing the data recorded by their FMC.
2. The data received in the framework of this Regulation shall be treated in a confidential manner."

This implies confidentiality between the Commission and the relevant data centre of the EU Member State. It is likely that this has hindered access for scientists to VMS data in Member States whose administrations (or fishing industries) adopt a conservative approach to data protection under this Regulation. However, under the revised Data Collection Framework (DCF) (COUNCIL REGULATION (EC) No 199/2008), VMS data provision to 'end-users' is explicitly addressed in the chapter entitled "USE OF DATA COLLECTED IN THE FRAMEWORK OF THE CFP". Article 15 comprises several clauses, amongst which is stated:

"1. This Chapter shall apply to all data collected:

- (a) under Regulations (EEC) No 2847/93, (EC) No 788/96, (EC) No 2091/98, (EC) No 104/2000, (EC) No 2347/2002, (EC) No 1954/2003, **(EC) No 2244/2003**, (EC) No 26/2004, (EC) No 812/2004, (EC) No 1921/2006, (EC) No 1966/2006 and (EC) No 1100/2007;
- (b) (i) under the framework of this Regulation: data on vessels' activity based on information from satellite monitoring and other monitoring systems with the required format;"

This clearly extends the scope of the DCF to data collected under the VMS regulation. In addition, 'this chapter' covers 8 articles that specify the nature of data provision and access. This includes Article 18 which is fairly explicit and provides only a limited exclusion clause:

"Article 18

Submission of detailed and aggregated data

1. Member States shall make detailed and aggregated data available to end-users to support scientific analysis:

(a) as a basis for advice to fisheries management, including to Regional Advisory Councils;

(b) in the interest of public debate and stakeholder participation in policy development;

(c) for scientific publication.

2. Where necessary, to ensure anonymity Member States may refuse to provide data on vessels' activity based on information from vessel satellite monitoring to end-users for the purposes referred to in paragraph 1(b)."

COUNCIL REGULATION (EC) No 199/2008 also provides certain relevant definitions:

"Article 1

(e) 'primary data' means data associated with individual vessels, natural or legal persons or individual samples;

(f) 'meta data' means data giving qualitative and quantitative information on the collected primary data;

(g) 'detailed data' means data based on primary data in a form which does not allow natural persons or legal entities to be identified directly or indirectly;

(h) 'aggregated data' means the output resulting from summarising the primary or detailed data for specific analytic purposes;

(i) 'end-users' means bodies with a research or management interest in the scientific analysis of data in the fisheries sector;"

Consequently, to a lay audience, it appears that EU Member States are obliged to provide VMS data to an organisation such as ICES, both at the detailed and aggregated level. Certain caveats may apply. For example:

- does the current DCF apply retrospectively to data collected prior to 2009 – probably not?
- the VMS regulation only requires EU Member States to retain VMS data for three years, so end-users have a finite period in which to request the data from Member States before it could potentially be disposed of.
- the original Data Collection Regulation placed an expiry date on data provided to end-users that limited the period in which it could be used. We do not think that applies under the current DCF.

This discussion relates only to EU Member States and does not reflect the position with other ICES members, for example, Norway. WGDIM was unable to provide such a perspective for non-EU ICES members.

Estimates of data quantities that can be expected from Vessel Monitoring Systems, including a consideration of VMS data temporal resolution in relation to its potential scientific uses, and any proposed or required changes in temporal resolution.

VMS data in the Marine Scotland (Science) database for 2008 contains data for 3735 vessels, representing not only all information for Scottish registered vessels, but boats of other nationalities entering or operating in Scottish waters at some point in a trip. This contains 3.4 million position/speed/heading records **after** those where the vessel is travelling at less than 0.2 knots are filtered out. The VMS data recording frequency is generally 2 hourly. Any increase in the recording frequency would increase the size of the data set proportionately, although vessel details could be held relationally. It is easily conceivable that VMS records will extend to many millions per EU Member State per year.

Much of the implementation of the DCF is specified in an EU Commission Decision (2008/949/EC). In terms of the application of VMS data to ecosystem indicators, Appendix XIII of the decision expresses a preference for a recording frequency of every 30 minutes, although this does not appear to be legislated and appears to be aspirational.

Although not referred to in the DCF, it would also be of benefit to require vessels between 10 m and 15 m to carry VMS equipment. These vessels make up a significant portion of many national fleets and are currently invisible to VMS-based analysis.

Investigation of organisations within the ICES area that currently archive or intend to archive VMS data and which allow access for scientific purposes

The statutory obligation of EU Member States has been discussed above with reference to data from 2009 onwards. No attempt has been made to identify the position for ICES member countries that are not EU Member States, or for Member State data holdings prior to 2009.

A summary of what scientific tools are being developed by existing research programmes, including EU Framework projects, to analyse and interpret VMS data

It is known that a number of national institutes are developing methods to manage and filter VMS data for scientific and fishery management purposes; however, a full inventory was not available to WGDIM.

One relevant research programme was highlighted as being of specific relevance to ICES. Under the European Commission open call for tenders (No MARE/2008/10 - Studies for carrying out the common fisheries policy), the EU invited tenders for an 18 month commercial contract to develop tools for logbook and VMS data analysis (Lot 2). The tender specification listed the main aim of the study to be:

"1. To create a method to deal with classification of Logbooks data for the fleet based approach. This method should assure a standardized approach at a Regional level assuring the criteria homogeneity between Member states. As a result, it shall be possible to automatically classify trips into métiers based on Logbooks species composition, gear or group of gears and area of operation.

2. To facilitate and develop the scientific use of VMS data as the basis for the estimation of pressure indicators in support of an ecosystem approach to fisheries management."

The specification also relates the analysis of log book and VMS data to three of the fishing pressure indicators (distribution of fishing activity, aggregation of fishing activity and areas not impacted by mobile gears) identified under Appendix XIII of EU Commission Decision (20089/949/EC). The study is therefore intended to provide end-users with the tools required to provide advice on specific issues of relevance to the EU and, by implication, it is particularly relevant to ICES as an advisory body.

Explicitly, the study shall:

- "1. Develop and test methods and produce protocols on how to present these indicators using GIS and how to link VMS databases to Logbooks.
2. Model the dependence of recording rate on the precision of the suggested indicators"

The contract was awarded in late spring 2008 to a consortium coordinated by IMARES (Netherlands) and will finish after 18 months.

In discussion, GIS experts advised WGDIM that the software platforms proposed for this study may not be suitable for fully operational purposes. A suggested action point is for the ICES Data Centre to maintain a watching brief during the development of this study (subject to approval of the consortium members and the EC) and to identify how best ICES can accommodate any tools/algorithms that are developed as may apply to its advisory remit.

Proposals for how ICES scientists and Expert Groups should gain access to VMS data in the future and what data interface, interrogation, display, analysis and interpretation tools ICES should obtain or develop.

The provision of VMS data to expert groups should be given the same priority as provision of fisheries landings and discard data. In the first instance, ICES needs to be kept in the loop regarding the Lot 2 consortium work described above, and to use that opportunity to better identify the additional database, visualisation and analytic toolset that will be required to address the most immediate of its MoU requests from the European Commission.

(This may be further complicated by the move towards electronic log book recording¹ which, by the time it is fully implemented, may render additional possibilities for tying-in the log book data to the VMS data).

WGDIM was not in a position to comment authoritatively on the tool-set needed in relation to other VMS-based advisory requests that ICES may receive, other than to note that such requests are current and cannot afford to wait for the outcome of the Lot 2 project in circa 18 months time.

¹ Corrigendum to Council Regulation (EC) No 1966/2006 of 21 December 2006 on electronic recording and reporting of fishing activities and on means of remote sensing