UNESCO

United Nations
Educational, Scientific and
Cultural Organization

Intergovernmental
Oceanographic
Commission

SCOR/IODE/MBLWHOI Library Workshop on Data Publication

UNESCO Headquarters, Paris, France
2 April 2010

SCOR
International Council for Science
Scientific Committee on Oceanic Research

UNESCO 2010

IOC Workshop Report No. 230
Oostende, 5 May 2010
English

Participants

Abstract:

A meeting was convened by the Scientific Committee on Oceanic Research (SCOR), the International Oceanographic Data and Information Exchange (IODE) of the Intergovernmental Oceanographic Commission (IOC) and the Marine Biological Laboratory/Woods Hole Oceanographic Institution Library (MBLWHOI Library) on 2 April 2010 to advance the two use cases that had been developed over the past two years by the SCOR/IODE activity on data publication. Increased access to data is becoming increasingly important to stimulate data reuse and to get data into the public domain, where it can contribute to discussions of important science issues, including the ocean's role in climate change and the global carbon cycle, marine biodiversity, fisheries, harmful algal blooms, and others. The meeting noted progress in the two use cases, on publishing data related to traditional journal articles and data held by data centres, and made plans for future cooperation between data centres and libraries.

TABLE OF CONTENTS

ANNEXES

_____

The Intergovernmental Oceanographic Commission (IOC) of UNESCO celebrates its 50th anniversary in 2010. Since taking the lead in coordinating the International Indian Ocean Expedition in 1960, the IOC has worked to promote marine research, protection of the ocean, and international cooperation. Today the Commission is also developing marine services and capacity building, and is instrumental in monitoring the ocean through the Global Ocean Observing System (GOOS) and developing marine-hazards warning systems in vulnerable regions. Recognized as the UN focal point and mechanism for global cooperation in the study of the ocean, a key climate driver, IOC is a key player in the study of climate change. Through promoting international cooperation, the IOC assists Member States in their decisions towards improved management, sustainable development, and protection of the marine environment.

1. SESSION 1: PILOT PROJECT REPORTS

This Session was chaired by Dr Gwen Moncoiffé (see agenda in Annex I). She invited all participants to introduce themselves. The list of participants is attached as Annex II. It was noted during the introductions that the approaches developed should strive for "larger carrots and small sticks", in other words, that there should be a focus on incentives for data publication and less on penalties for not submitting data. It was also noted that the Global Biodiversity Information Facility (GBIF: http://www.gbif.org/) is about to issue a report on data publication.

1.1 Summary of the SCOR/IODE activities

This agenda item was introduced by Dr Moncoiffé. She recalled that what was to become the SCOR/IODE initiative on data publication started in December 2006 at the Second SCOR Summit of International Marine Research Projects (see http://www.scor-int.org/Project_Summit_2/ProjCoord2.htm). The meeting brought together representatives of most large-scale international ocean research projects (e.g. SOLAS, GEOTRACES, IMBER, GLOBEC, etc.). The meeting considered what constituted the major barriers to data sharing and greater data submission to national and global databases, and identified a need to create a formal process to ensure that scientists get credit for releasing their data and for every time the data are used by others. To this aim, the meeting recommended that SCOR form a Panel on Ocean Data Publication and Incentives.

In 2008, SCOR and IODE began to work together on this issue and the "SCOR/IODE Workshop on Data Publishing" (17-18 June 2008, IOC Project Office for IODE) was organised. The objectives of the meeting were to (i) Describe current status of data citation and publication in oceanography; (ii) Identify problem areas; (iii) Identify interoperability issues of current data citation and publication practices; and (iv) Formulate suggestions to address problem areas. The meeting concluded that the following actions were needed:

(i)   Build new infrastructure and new approaches to data publication;

(ii)   Increase the availability of the data used for figures, tables and statistical analyses in traditional journal articles;

(iii)   Encourage the expansion of journals which specialise in "data publications" or "data briefs"; i.e. journals where data/datasets are described rather than interpreted.

(iv)   Requirement for data repositories to serve as archive of data related to journal articles;

(v)   Use persistent identifiers to anchor data in repositories to be used/published in publications;

(vi)   SCOR and IODE to work with existing data centres to promote the development of data repositories at the institutional, national and/or regional level.

The report of the meeting is available as IOC Workshop Report No. 207 at http://www.iode.org/wr207.

In December 2008, a meeting was organized with ocean science journal editors in San Francisco, USA with representatives present from *Earth System Science Data* (ESSD), *Fisheries Oceanography*, *Earth and Planetary Science Letters*, *Journal of Geophysical Research-Oceans*, *Biogeosciences* and *Progress in Oceanography*. Written input was received also from the *Journal of Plankton Research*, *Oceanography*, and *Palaeoceanography*.

The meeting came to the following conclusions:

(i)   The effort was worthwhile and many of the editors consulted want to stay involved in the discussions and participate in the development of the ideas;

_____

(ii)   The idea of the peer review of datasets requires more discussion and careful consideration;

(iii)   The process of publication of data briefs or other stand-alone papers describing data sets is being tested by ESSD. The SCOR/IODE effort should focus for now on issues related to providing the digital backbone for data related to traditional publications;

(iv)   More attention needs to be given to how digital object identifiers (DOIs) can best be used to link journal articles and datasets;

(v)   It is important to know whether fields outside ocean sciences are pursuing data publishing.

In March 2009 (IOC Project Office for IODE, Oostende, Belgium, 9-11 March 2009) a meeting was held to develop use cases. The meeting included participants from BCO-DMO, BODC, WDC-MARE, IODE, and Elsevier. The use cases identified were:

- Use Case 1: Creating data publications from existing and future holdings at national data centres.

- Use Case 2: Providing the "digital backbone" for traditional journal publications.

Pilot projects were chosen to test the processes for data publication in the two cases. While BODC worked on developing Use Case 1, a meeting was organized by the MBLWHOI Library in Woods Hole, Massachusetts, USA to further develop the pilot project for Use Case 2. Progress on these two Use Cases will now be presented.

1.2 Use Case 2: MBLWHOI Library - A Woods Hole Data Repository for Data Supporting Published Articles

Ms Lisa Raymond introduced the efforts of the MBLWHOI Library of the Woods Hole Oceanographic Institution and Marine Biological Laboratory to develop a data repository for data supporting published articles, specifically focusing on the "backbone data" used to generate the figures and tables in an article. This work is supported by a George Frederick Jewett Foundation grant. Ms Raymond recalled that the "Data Attribution and Provenance Workshop" was held in Woods Hole, in April 2009 and involved stakeholders that included scientists, data managers and librarians. It was concluded that data must be discoverable, citeable and available on the Internet. Resources, standards and workflows must be defined to support publisher and funding agency mandates. It was agreed that as an action item the Library should develop a process to deposit data.

A number of important questions were raised: (i) Where should data be deposited? It was decided to use the Woods Hole Open Access Server (WHOAS) as the data repository; (ii) What metadata scheme should be used? The group came to the consensus that Dublin Core alone was not sufficient, so it was augmented with Darwin Core and "Woods Hole Core"; (iii) Provenance was identified as important as well. In this regard it was noted that the selected repository software platform DSpace generates provenance information upon ingestion of data items and is widely available in the library community. Information generated includes file size, date, depositor name and checksum. Other information can be accommodated in existing Dublin Core fields; (iv) For persistent identifiers (essential!) it was noted that the MBLWHOI Library has an existing relationship with CrossRef (http://www.crossref.org); and (v) An important hurdle identified was the issue of versioning of data sets. Concerns were raised at the meeting about how corrections and updates to a data set would be handled. For the purposes of this project any change to a data set would lead to a new data set, new metadata and a new DOI.

The project developed a workflow as illustrated in Figure 1.
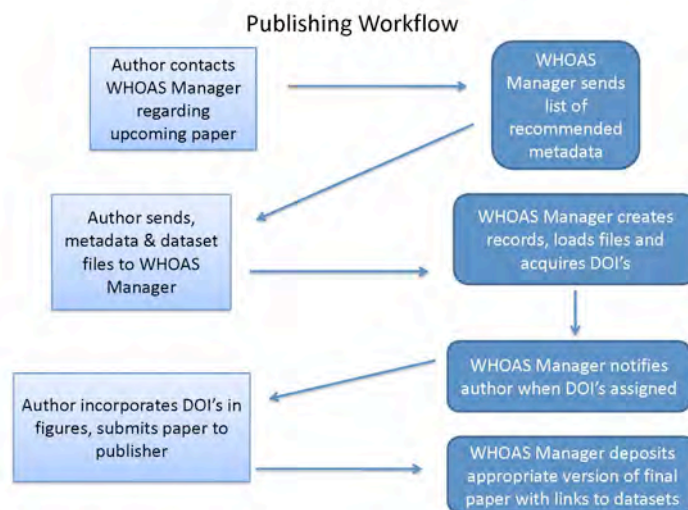
Figure 1: MBLWHOI repository workflow

Preliminary work on trials with the repository has been carried out by Peter Wiebe (WHOI), and by Peter Smith and his group (MBL). Experience so far has suggested that it is important to start discussions of data publication early in the publication process and for test cases to select papers that have relatively few figures and tables, to ease the author burden of assembling the datasets.

Ms Raymond identified the following ongoing challenges:

- Researcher concerns regarding reuse and misuse of data

- Proprietary file types

- Responsibility for quality control of data

- Additional work for authors

The Meeting congratulated the MBLWHOI Library on the results of the pilot activity and called on the Library to prepare documentation that will assist others to implement similar systems. The MBLWHOI Library has plans to share their experiences and progress at meetings of marine libraries. It was further noted that the creation of metadata remains a stumbling block at the scientist level. The meeting remarked, however, that metadata creation should be considered a process rather than a one-time event. Dr Lowry remarked further that, in the UK case, it is the task of BODC, and not scientists, to define the metadata. The Meeting thus concluded that there exist several models and mechanisms for the creation of the metadata. One participant noted that the more automated the creation of metadata, the better.

1.3 Use Case 1: Data publication and citation from the point of view of BODC as a "typical" IODE data centre

This agenda item was introduced by Dr Roy Lowry.

Dr Lowry explained that he had considered the issue of data publication as a data centre and how their method of work would fit the needs. He recalled that BODC is the IODE National Oceanographic Data Centre for the United Kingdom. BODC is funded by the National Environment Research Council (NERC) to be the UK's marine data centre and part of the Marine Environmental Data Information Network (MEDIN). In addition, BODC ingests data from international research projects (e.g. GEOTRACES) and receives funding from international activities such as SeaDataNet. Dr Lowry

identified the important difference between data set serving and data publication, as they do not follow the same process.

Dataset Serving—Data centres are primarily in the business of dataset serving. In general, BODC takes data from many sources and "atomizes" these into a common schema. Several processes are then applied, such as quality control, adding flags, etc. Data thus evolve as a continuous process and there is a never a final form of a "data atom". This process is resource intensive. Data are discovered via an x,y.z,t parameter search and the data set delivered will vary depending on the parameters specified and when the data are served, as they may be updated in the system over time.
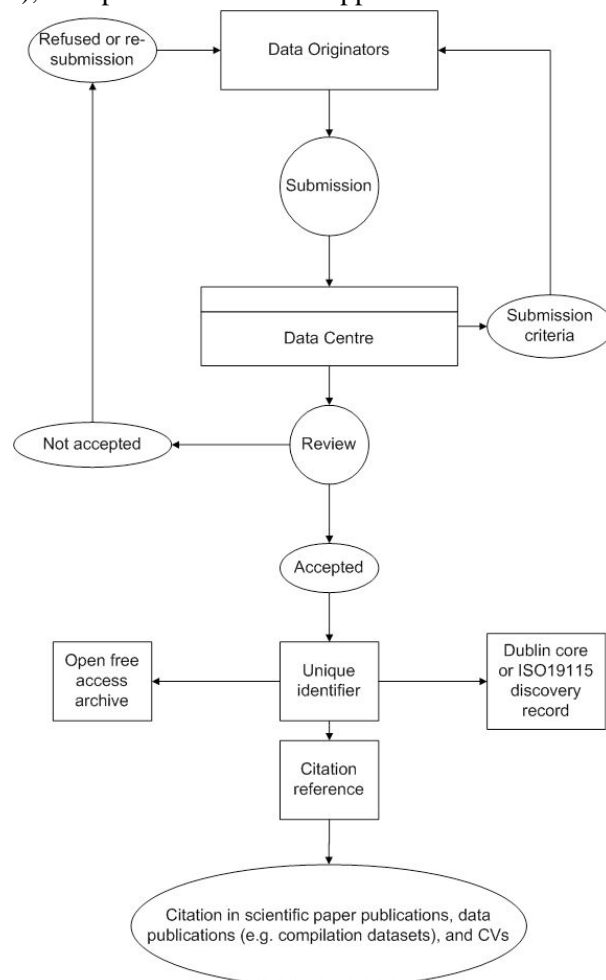
Data Publication—On the other hand, in the concept of the data publication the data set must be static and a request for a particular data set should always result in the same response. The dataset has a literature citation conforming to a standardised, globally recognised syntax. The dataset has a permanent identifier (URN) that resolves to a location on the internet (URL) from where the data may be obtained without authentication or authorisation.

Data centre involvement in data publication can be considered at three levels of service: bronze, silver, and gold.

Bronze Level—The data centre publishes whatever is submitted, as it was submitted, with no quality control or processing. The data centre would obtain a DOI, prepare and publish a discovery metadata record (Dublin Core, ISO19115 profile or FGDC), and place the data as supplied on a server at the DOI resolution URL.

Silver Level—The data centre checks that data submissions conform to data publication standards. At this level, a data centre would provide a service that is a direct analogue of the peer review of a journal article and is described in the 2008 SCOR/IODE report. This is the level of service that would probably be of interest to data centres, as it has a good balance of costs to the centre and benefits in terms of ingestion of data and service to their national ocean science communities and agencies. This level of service would require that data centres be linked to DSpace repositories because data centres cannot provide the e-repository function. The data centre would carry out the functions as for bronze level of service, plus would conduct standard conformance checking, work with the originator to rectify shortcomings, and certify standards conformance. Figure 2 shows the silver-level process.

Gold Level—The data centre ingests the submission with full quality control and metadata enhancement before publishing it as a snapshot. This level of service implies a full accession ingestion process, including reformatting data, however it is submitted; point-by-point quality assessment and issue flagging; standardising semantics such as

Figure 2: Silver Level Service

parameter labels; preparation of evaluation and usage metadata; incorporation into a common schema;

_____

and snapshot instantiation for publication. This gold level service is often provided by data centres for datasets which are supported either by the data centre's core funding or through project mandate. In the past, however, many data sets ingested by the data centre were never published and thus no credit was given. Such data sets could now be published and thus revived using the following process:

- Identification of candidate subsets of ingested data
- Snapshot instantiation to a standard that would pass silver-level quality checks
- Publication in a DSpace repository

This process could possibly be undertaken in close collaboration with data originators wishing to produce data publications such as an ESSD paper or a data brief in the journal **Geochemistry, Geophysics, Geosystems** (http://www.agu.org/journals/gc/) from data they have previously lodged with the data centre.

The following table summarizes the advantages and disadvantages of the three service levels.

	Advantages	**Disadvantages**
Bronze Level	• Cheap • Turnaround within days feasible	• Published datasets exposed to technological redundancy • No protection against retention of worthless data
Silver Level	• Relatively cheap • Easily integrated into scientific publication process, including timescale compatibility • Enhanced data management standards in the scientific community	• Risk of datasets being exposed to technological redundancy, particularly the semantics
Gold Level	• Future reusability of data virtually guaranteed	• Expensive • Long timescales

Dr Lowry noted that Argo had expressed interest in data publishing. An important question would be how to define a data set in this case, as the Argo database changes daily as profiles are added. Dr Lowry also noted that data centres are well equipped to help create metadata and to help scientists prepare data publications for deposit in DSpace-type repositories (see http://www.dspace.org/). Qualified groups of experts can also create metadata through the process of "qualified social tagging".

BODC (and probably most data centres) would not be interested in facilitating the management of highly processed data relating to tables and figures in papers (Use Case 2 of the SCOR/IODE/MBLWHOI Library activity), as there would be no advantage for data ingestion. But, it should be of interest to libraries.

Data centres cannot necessarily provide all the e-Repository infrastructure for their data publication activities because of physical limitations on storage capacity, resource issues related to developing the necessary infrastructure, and management unwillingness to support data duplication. The SCOR/IODE/MBLWHOI Library process needs to engage in the development of an international network of providers of e-Repository infrastructure for data publication. Under what terms and conditions would such resource be available? Standards would be required to ensure interoperability among repositories.

Dr Lowry presented the following conclusions:

_____

1. Guidelines are needed to direct data centres on how they can become engaged in the data publication process and used to engage a network of data centres in the activity.
2. Standards need to be developed for dataset submissions that are to be published as is by data centres.
3. International e-Repository infrastructure is required and the SCOR/IODE/MBLWHOI Library process needs to find ways to develop this.
4. The digital library community and data editors need to be engaged in the publication of processed data behind tables and figures (Use Case 2).
5. Journal editors and the library community need to declare the standard way for citing datasets used in journal publications.

Mr Pissierssens informed the participants that IODE developed the OceanDocs e-repository system (http://www.oceandocs.net). OceanDocs is an e-repository (electronic repository) of scientific publications related to marine science and oceanography. The IODE OceanDocs is the central repository of a global network of national and regional repositories. OceanDocs enables researchers to deposit the full text of their works: articles, conference papers, technical reports, working papers, theses, and more. The information managers and librarians of the participating institutes will help the researchers with the submission of their publications. OceanDocs includes not only a central repository hosted by the IOC Project Office for IODE in Oostende, Belgium, but also national and regional nodes (e.g. the Institute of Biology of the Southern Seas E-repository in Ukraine, the Digital Repository Service at the National Institute of Oceanography in India, etc.). The central repository in Oostende, Belgium, includes institutional repositories of 15 countries in Africa and 9 countries in Latin America).

2. SESSION 2: TAKING THE PROCESS FORWARD

This Session was chaired by Lisa Raymond. She invited Roy Lowry to present his Road Map Proposal. Dr Lowry identified the following requirements:

(i) Guidelines explaining to data centres what is involved in engagement with data publication
(ii) Data centre engagement in the data publication process
(iii) Published dataset standards guidelines
(iv) Establishing e-Repository infrastructure
(v) Published dataset citation standard
(vi) Engagement of journal editors in processed dataset publication

Creating awareness, interest and commitment of the data centres was identified as a considerable issue. In this regard it was noted that interest and commitment must be achieved beyond the IODE NODC community, as well as beyond oceanography. The IODE network was therefore seen as one of several conduits; the World Data System (WDS) of the International Council for Science (ICSU) and American Geophysical Union and European Geosciences Union would be others to consider.

The Meeting noted that the upcoming 22nd International CODATA Conference, 24-27 October 2010 (Stellenbosch, Cape Town, South Africa) (http://www.codata2010.com/call-for-session-organizers.php) would be an appropriate venue to promote the SCOR/IODE/MBLWHOI Library initiative and to avoid too many parallel initiatives related to data publishing. It was therefore recommended to submit a paper to the CODATA Conference on the initiative.

Dr Ed Urban was requested to investigate the possible presentation of a paper on the SCOR/IODE/MBL/WHOI initiative at the 22nd International CODATA Conference.

After lengthy discussions the Meeting agreed on the following timeline for the way forward:

(i)   **Prepare project brief that describes objectives, technical approach and expected outcomes of the data publication pilot activity (existing documentation prepared by MBLWHOI Library should be used). Deadline: May 2010 (Roy Lowry, Lisa Raymond)**

(ii)   **Organize technical discussions on compatibility/modification of OceanDocs metadata scheme for data publication pilot project. Deadline: April/May 2010 (Lisa Raymond and Marc Goovaerts; Project Leader OceanDocs)**

(iii)   **Propose a session for the Fall AGU Meeting (Peter Fox, Cyndy Chandler, Lisa Raymond , Anna Gold? Session proposals are due by 27 May)**

(iv)   **Identify a small number of marine libraries and data centres (IODE and others) willing and able to participate in a data citation pilot activity. The marine libraries should already be operating or using an e-repository (OceanDocs). The two use cases will be put out as a challenge for joint work between data centres and libraries. The challenge will be elaborated in a more detailed document. Deadline: May/June 2010 (IODE Secretariat and SCOR to coordinate)**

(v)   **Meeting/Training Course for participating marine libraries/data centres. Deadline: July-September 2010 (IODE Secretariat to implement through OceanTeacher Academy)**

(vi)   **Start pilot activity using MBLWHOI Library methodology and work flow with identified marine libraries and data centres. Deadline: start from August 2010 (coordinated by MBL/WHOI and BODC)**

(vii)   **Present preliminary report on progress of pilot project at 22nd International CODATA Conference (24-27 October 2010) (presenter to be identified; abstract must be submitted by 30 April)**

(viii)   **Present progress report on pilot project at IODE 50th Anniversary Scientific Conference (provisional 21-22 March 2011)**

(ix)   **Present progress report on pilot project at IODE-XXI and discuss formal establishment of Pilot Project on Data Citation (with more extended participation) (provisional 23-26 March 2011)**

It was noted that it would be desirable to bring together data centres to promote the data publication concept. The Meeting did not come to a conclusion on the best way to do so. A dedicated conference could be considered. Other possible opportunities could be the IODE 50th Anniversary Scientific Conference and/or IODE-XXI in March 2011.

The Meeting stressed the importance of involving publishers in the pilot activity as soon as possible, but noted that it would desirable to achieve practical progress and demonstrate feasibility prior to involving publishers. The IODE 50th Anniversary Scientific Conference could be an appropriate venue to invite the Publishing community. The Meeting also noted the importance to involve not only commercial publishers but also open-source publishers such as Copernicus (http://publications.copernicus.org/) as their goals and objectives may be closer to the project's (**Ed Urban to coordinate**)**.**

The Meeting also recommended contacting ICSTI, the International Council for Scientific and Technical Information (http://www.icsti.org) as they may be interested in participating. They have done a pilot project on data publication. In this regard it was noted that ICSTI's 2010 Summer Conference will be held in Helsinki on 10-11 June. (**Ed Urban to coordinate**).

3. SESSION 3: DATA PUBLICATION TECHNICAL ISSUES

3.1 What should lie at the end of a DOI?

The Meeting considered the question "what lies at the end of a DOI" and identified the following answers: (i) URL of the homepage of a Web data delivery system; (ii) Deep URL into a Web data delivery system; (iii) URL of a zip file; (iv) URL of a CF-compliant NetCDF file; and (vi) Archival Information Package (AIP) as specified in the Open Archival Information System (OAIS) model. One important point to note is that DOIs are one tool that we are trying to use for many different jobs. A major issue is whether the link from the DOI is accessible to a human only, or also to a software agent. The ideal situation for a link to data would be to be able to download data with only a few clicks of the mouse. CrossRef is an example of a system that provides machine-readable metadata.

It was noted that there are a number of DOI registration agencies. They are listed on http://www.doi.org/registration_agencies.html. "The primary role of Registration Agencies (RAs) is to provide services to Registrants - allocating DOI name prefixes, registering DOI names and providing the necessary infrastructure to allow Registrants to declare and maintain metadata and state data. This service is expected to encompass quality assurance measures, so that the integrity of the DOI system as a whole is maintained at the highest possible level (delivering reliable and consistent results to users). This includes ensuring that state data are accurate and up-to-date and that metadata are consistent and complies with both DOI system Kernel and appropriate Application Profile standards".

The journal *Earth System Science Data* (ESSD) (http://earth-system-science-data.net/) was mentioned as an example of a Journal that provides DOIs to data sets. Example article: http://www.earth-syst-sci-data-discuss.net/3/55/2010/essdd-3-55-2010.html

Example: doi: 10.3334/CDIAC/otg.CARINA.AMS.V1.2

Links to: http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA_Database/CARINA.AMS.V1.2/

3.2 Data publication content standards and quality assurance

This agenda item was introduced by Peter Fox.

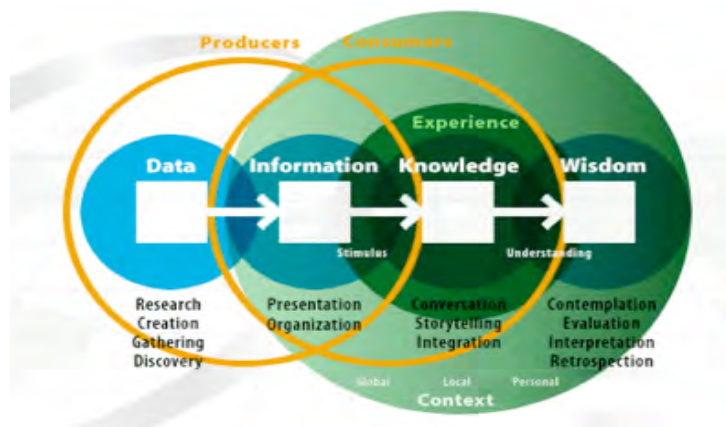Dr Fox started his presentation with the Micro Life Cycle of data, illustrated in Figure 3.

Figure 3 : Micro Life Cycle of Data

Publication is both a noun and a verb. What we are trying to publish is a complex e-science object. Dr Fox suggested the term "publication data publication." An important consideration is how we define the relevant authorities and the continuum between self-publication and authoritative (e.g. by a third-

party source) publication? A related issue is what is the content that is to be published and eventually cited. A well-worn phrase is that one person's metadata is another person's data and vice versa. For the topic of this section, that is, content standards, the real key is "mashing up" of standards – meaning that as many, any or all standards can be used, especially when encoded in XML-based languages (i.e. use of namespaces). The goal is that standards should be invisible to users.

In establishing a reputable publication, should the quality be defined as an engineering process or a science process? The answer is likely to depend on who is using the citation. In the emerging world of open data (i.e. by publishing), can *transparency* be a proxy for quality assurance? The expertise, knowledge, or measurement abilities of the person that collected the data often engenders the trust. Dr Fox also introduced the theme of workflow. In other communities people publish "workflows," but the term workflow varies by communities. How could one test if a workflow is competent without data?

In regard to the architectural foundations of publication and citation if there was to be wide-spread adoption, the Web/Internet world is one of the few scale-free environments supporting many application fields (including science). If we can use "Web" principles, including the peer and social sense of "trust," the issue will be solved. We have not yet reached that point in scientific circles.

Dr Fox concluded by identifying a few questions to stimulate discussion:

- There are enough standards already, pick one and use it!
- What do you mean there are no applications/tools to work with the standards?
- What do you mean "use a library (bibliometric) format, this is my science data"?
- Cross-walking metadata "standards," can we (or do we need to) do it?
- Are you questioning the quality of my data?

One of the emerging questions was: Who are the stakeholders of the libraries and the data centres? There are a variety of stakeholders of libraries, including individual scientists, government agencies, the public, policymakers, educational institutions, environmental organizations, and others, potentially the same type of stakeholders that access traditional libraries. As access becomes more straightforward, the same users may increasingly seek information from data centres.

Regarding the importance of standards, the conclusion was that users do not usually care about standards but they do care about being able to import data into their familiar tools (e.g. ODV, Matlab, etc). Having a standard that enables interoperability is therefore more important than having a standard for the sake of having a standard. Using the scale-free nature of the Web, we might be able to address some of our issues that are otherwise unscalable.

4. GAP ANALYSIS AND NAVEL GAZING

This Agenda Item was chaired by Dr Ed Urban.

4.1 Critical analysis of the SCOR/IODE data citation and publication process so far

Dr Lowry noted that during the past two years little practical progress had been achieved in the activity, possibly because at previous meetings the marine libraries had not been fully involved. The involvement of the MBLWHOI Library, their small-scale project, and the realization by the data centres that this issue was an ideal candidate for close collaboration between data centres and marine libraries made the concrete output of this Meeting very promising.

4.2 Data publication issues not so far addressed

This Agenda Item was introduced by Alex Dorsk. Mr Dorsk presented a brief report on the experience achieved with the WHOAS (Woods Hole Open Access Server). A first question was "Why aren't people banging down our doors to deposit data?" During the WHOAS experiences it was found that users encounter challenges during the data citation process. Responding to many of these challenges involve understanding users and their specific situations. Often these challenges do not relate to the repository but to general problems in data management.

Data depositors have to place data in a repository, which may take time to do, or they many not even know that repositories exist. Some of these barriers include data organization (deciding which data are backbone data, separating data, packaging data), metadata management (deciding which metadata are necessary, packaging metadata, generating metadata), and knowledge (knowing what a repository is, knowing which repositories exist, knowing how to use a repository). We have begun to define the intersections between data centres and libraries in data publication, which will be crucial for moving forward.

These were some of the experiences with WHOAS that exemplified these barriers:

On Understanding repositories:

- Users don't know that repositories exist
- Users think a repository is a tool for managing data throughout its lifecycle, like a virtual workdesk
- Users think that they have to know how to use repositories

On managing data:

- Users can't easily separate out publishable chunks of data
- Users have concerns about data ownership and control

On managing metadata:

- Users don't know what metadata to include
- Users don't have time or expertise to generate metadata

What methods do we have available to overcome the barriers?

Tools: tools can be created to help users overcome barriers. An example was given that described creating a Webform to guides users through the process of creating metadata.

Outreach: efforts can be made to educate scientists about data citation and repositories. An example was given, developing a "*Scientist's guide to repository metadata*". The Meeting agreed on the need for such a guide. **Alex Dorsk agreed to prepare content for OceanTeacher in this regard. It is expected that this will be a "phase 2" project after we have feedback from successful use cases.**

The Meeting further agreed to share the report of this Meeting with publishers and requested all participants to contact publishers and inform them of the URL of the report when available. (Deadline: when report is available, By: All)

5. CLOSING OF THE MEETING

Dr Lowry summarized that it has been hard to make progress in bringing data centre and library paradigms together, but we have begun to see how the two institutions could work together, without replicating roles or competition. It is important to think about end users of the data in the process, including active researchers.

The participants expressed their great appreciation to Cathy Norton, who had started the WHOAS initiative and had approached IOC/IODE to have this joint Meeting. It was regretted that Cathy had not been able to participate in the meeting, due to health reasons. The Meeting expressed the hope that Cathy would soon be able to participate in the planned activities as well as in the 2011 IODE events.

The Meeting thanked the IOC for hosting the Meeting at IOC/UNESCO Headquarters. The Meeting was closed on 2 April 2010 at 16h45.

_____

ANNEX I

AGENDA OF THE MEETING

1.   SESSION 1: PILOT PROJECT REPORTS

     1.1   Summary of the SCOR/IODE Activities

     1.2   Use Case 2: MBLWHOI Library - A Woods Hole Data Repository for Data Supporting Published Articles

     1.3   Use Case 1: Data publication and citation from the point of view of BODC as a 'Typical' IODE Data Centre

2.   SESSION 2: TAKING THE PROCESS FORWARD

3.   SESSION 3: DATA PUBLICATION TECHNICAL ISSUES

     3.1   What should lie at the end of a DOI?

     3.2   Data publication content standards and quality assurance

4.   GAP ANALYSIS AND NAVEL GAZING

     4.1   Critical analysis of the SCOR/IODE data citation and publication process so far

     4.2   Data publication issues not so far addressed

5.   CLOSING OF THE MEETING

ANNEX II

LIST OF PARTICIPANTS

Cyndy CHANDLER
Informatics Specialist
Woods Hole Oceanographic Institution
(WHOI)
MS 36, Woods Hole, MA, 02543
United States
Tel: +1-508-289-2765
Fax: +1-508-457-2169
Email: cchandler@whoi.edu

Alexander DORSK
Systems/Data Librarian
Woods Hole Oceanographic Institution
MBLWHOI Library
United States
Tel: +1 508-289-2538
Email: adorsk@whoi.edu

Peter FOX
Professor
Rensselaer Polytechnic Institute
Tetherless World Constellation
110 8th St, Troy, NY 12180
United States
Tel: +1-518-276-4862
Fax: +1-518-276-4464
Email: pfox@cs.rpi.edu

Christopher FREELAND
Director
Center for Biodiversity Informatics
4344 Shaw Blvd., St. Louis, MO
63110
United States
Tel: +1-314-518-2412
Fax: +1-314-577-0897
Email: chris.freeland@mobot.org

Anthony GODDARD
Informatics Analyst
Marine Biological Laboratory (MBL)
WMBLWHOI Library
7 MBL St, Woods Hole, MA, 02543
United States
Tel: +1-508-564-3855
Email: agoddard@mbl.edu

Roy LOWRY
British Oceanographic Data Centre
Joseph Proudman Building
6 Brownlow Street
Liverpool
L3 5DA
United Kingdom
Email: rkl@bodc.ac.uk

Mike MCCANN
Software Engineer
Monterey Bay Aquarium Research
Institute
Information Applications
7700 Sandholdt Road
Moss Landing CA 95039
United States
Tel: +1-408-775-1769
Fax: +1-408-775-1646
Email: mccann@mbari.org

Holly MILLER
Project Leader
Marine Biological Laboratory
MBLWHOI Library
7 MBL Street, Woods Hole, MA
02540
United States
Tel: +1-508-289-7632
Email: hmiller@mbl.edu

Stephen MILLER
Head, Geological Data Center
Scripps Institution of Oceanography,
UC San Diego (SIO)
9500 Gilman Drive Mail Code 0214
La Jolla 92093
United States
Tel: +1-619-548-0784
Email: spmiller@ucsd.edu

Gwenaëlle MONCOIFFÉ
Marine Data Scientist
British Oceanographic Data Centre
Joseph Proudman Building
6 Brownlow Street
Liverpool
L3 5DA

United Kingdom
Email: gmon@bodc.ac.uk

Thomas MORITZ
1968 1-2 South Shenadoah Street, Los
Angeles, CA 90034-1208
United States
Tel: +-1-310-963-0199
Email: tom.moritz@gmail.com

Linda PIKULA
Chair, IODE Group of Experts on
Marine Information Management (GE-
MIM)
Regional Librarian
NOAA Central and Regional Libraries
(NCL)
4301 Rickenbacker Causeway
Miami Florida 33149
United States
Tel: +1-305-361-4429
Fax: +1-305-361-4552
Email: linda.pikula@noaa.gov

Lisa RAYMOND
Assistant Library Director, WHOI
Woods Hole Oceanographic Institution
MBLWHOI Library
360 Woods Hole Road, Woods Hole,
MA 02543
United States
Tel: +1-508-289-3557
Fax: +1-508-457-2156
Email: lraymond@whoi.edu

Greg REED
Co-Chair IODE
Australian Ocean Data Centre Joint
Facility Fleet Headquarters Wylde
Street Building 89
Garden Island Potts Point NSW 2011
Australia
Tel: +61 2 9359 3141
Fax: +61 2 9359 3120
Email: greg@metoc.gov.au

Kate ROBERTS
Assistant Director Information
Systems
University of Tasmania (eMarine
Information Infrastructure)
University of Tasmania
Private Bag 21
Hobart TAS 7001
Australia

Email: Kate.Roberts@utas.edu.au

Ed URBAN
Executive Director
Scientific Committee on Oceanic
Research (SCOR)
College of Earth, Ocean, and
Environment
Robinson Hall, University of Delaware
Newark, DE 19716
United States
Tel: +1-302-831-7011
Fax: +1-302-831-7012
Email: ed.urban@scor-int.org

Ian WOOD
PhD Student
Australian National University (ANU)
Canberra, ACT 0200
Australia
Tel: +61 4 240 17499
Fax: +61 2 612 50010
Email: ian.wood@anu.edu.au

IOC Secretariat

IOC SECRETARIAT
Peter Pissierssens
Head, IOC Project Office for IODE
Intergovernmental Oceanographic
Commission of UNESCO (IOC)
Wandelaarkaai 7/61
8400 Oostende
BELGIUM
Tel: +32 59 34 01 58
Fax: +32 59 79 52 20 or Fax: +32 59
34 01 52
E-mail: p.pissierssens@unesco.org