Fisheries and Oceans Canada | Pêches et Océans Canada
Science | Sciences

C S A S | **S C C S**

Canadian Science Advisory Secretariat | **Secrétariat canadien de consultation scientifique**

Proceedings Series 2008/024 | **Série des comptes rendus 2008/024**

Proceedings of a Conference on Ocean Biodiversity Informatics

Bedford Institute of Oceanography
Dartmouth, Nova Scotia

2-4 October 2007

Compte rendu de la conférence Ocean Biodiversity Informatics

Institut océanographique de Bedford
Dartmouth (Nouvelle-Écosse)

Du 2 au 4 octobre 2007

T. Worcester, L. Bajona, and B. Branton

T. Worcester, L. Bajona et B. Branton

Fisheries and Oceans Canada / Pêches et Océans Canada
Bedford Institute of Oceanography / Institut océanographique de Bedford
Dartmouth, Nova Scotia / Dartmouth (Nouvelle-Écosse)
B2Y 4A2 Canada

December 2008 | **décembre 2008**

Foreword

This workshop was not carried out as a formal DFO Science Advisory Process; however, it is being documented in the CSAS Proceedings series as it presents some topics of interest related to the advisory process. The purpose of these proceedings is to archive the activities and discussions of the meeting, including research recommendations, uncertainties, and to provide a place to formally archive official minority opinions. As such, interpretations and opinions presented in this report may be factually incorrect or misleading, but are included to record as faithfully as possible what transpired at the meeting. No statements are to be taken as reflecting the consensus of the meeting unless they are clearly identified as such. Moreover, additional information and further review may result in a change of decision where tentative agreement had been reached.

Avant-propos

Cet atelier n'a pas été tenu dans le cadre officiel du Processus de consultation scientifique de secteur des Sciences du MPO. Il est toutefois documenté dans la série des Comptes rendus du SCCS, car il couvre certains sujets en lien avec ce processus. Le présent compte rendu fait état des activités et des discussions qui ont eu lieu à la réunion, notamment en ce qui concerne les recommandations de recherche et les incertitudes; il sert aussi à consigner en bonne et due forme les opinions minoritaires officielles. Les interprétations et opinions qui y sont présentées peuvent être incorrectes sur le plan des faits ou trompeuses, mais elles sont intégrées au document pour que celui-ci reflète le plus fidèlement possible ce qui s'est dit à la réunion. Aucune déclaration ne doit être considérée comme une expression du consensus des participants, sauf s'il est clairement indiqué qu'elle l'est effectivement. En outre, des renseignements supplémentaires et un plus ample examen peuvent avoir pour effet de modifier une décision qui avait fait l'objet d'un accord préliminaire.

Proceedings of a Conference on Ocean Biodiversity Informatics

Bedford Institute of Oceanography
Dartmouth, Nova Scotia

2-4 October 2007

T. Worcester, L. Bajona, and B. Branton

Compte rendu de la conférence Ocean Biodiversity Informatics

Institut océanographique de Bedford
Dartmouth (Nouvelle-Écosse)

Du 2 au 4 octobre 2007

T. Worcester, L. Bajona et B. Branton

Fisheries and Oceans Canada / Pêches et Océans Canada
Bedford Institute of Oceanography / Institut océanographique de Bedford
Dartmouth, Nova Scotia / Dartmouth (Nouvelle-Écosse)
B2Y 4A2 Canada

December 2008

décembre 2008

ISSN 1701-1272 (Printed / Imprimé)

Published and available free from:
Une publication gratuite de :

Fisheries and Oceans Canada / Pêches et Océans Canada
Canadian Science Advisory Secretariat / Secrétariat canadien de consultation scientifique
200, rue Kent Street
Ottawa, Ontario
K1A 0E6

http://www.dfo-mpo.gc.ca/csas/

CSAS@DFO-MPO.GC.CA

Printed on recycled paper.
Imprimé sur papier recyclé.

Correct citation for this publication:
On doit citer cette publication comme suit :

DFO. 2008. Proceedings of a Conference on Ocean Biodiversity Informatics; 2-4 October 2007. DFO Can. Sci. Advis. Sec. Proceed. Ser. 2008/024.

TABLE OF CONTENTS

SUMMARY

The theme for the Ocean Biodiversity Informatics ’07 conference, which was the third in a series of international conferences on ocean data management, was enhancing data quality assurance, as well as extending the range of information products that biodiversity informatics practitioners must be developing and delivering to satisfy the needs of users on many different levels (e.g., general public, academics, decision makers, and industry) -- in other words, translating the data housed in databases into products that are useful for society in general and decision makers in particular. To this end, a conference was held on 2-4 October 2007.

Welcoming remarks were given by Dr. Thomas Sephton, as acting director of the Bedford Institute of Oceanography, and by Dr. Fred Grassle, founding director of the Institute of Marine and Coastal Sciences at Rutgers State University. The keynote address was given by Dr. Ron O’Dor, the lead scientist for the Census of Marine Life. The conference consisted of 37 oral presentations organised into 7 sessions: 1) Visualisation Tools, 2) Habitats and Ecosystems, 3) Species Names and Management Tools, 4) Metadata Developments, 5) Data Use and Analysis, 6) New Data Systems, and 7) Integrating Different Types of Data. Each session opened with introductory remarks by a session chair and closed with a moderated panel discussion. Over 100 researchers, data managers, and contractors from government, universities, and the private sector participated in this conference, many of whom were from out of town. In general, participants agreed that the workshop was productive, and provided clear recommendations on how to proceed on a global scale with ocean biodiversity informatics.

SOMMAIRE

La conférence Ocean Biodiversity Informatics ‘07, la troisième dans une série de conférences internationales traitant de la gestion des données sur l’océan, était axée sur l’amélioration de l’assurance de la qualité des données et sur l’élargissement de la gamme des produits d’information que doivent élaborer et produire les praticiens de l’informatique dans le domaine de la biodiversité pour combler les besoins de différents ordres d’utilisateurs (grand public, universitaires, décideurs et industrie). Autrement dit, il s’agit pour ces praticiens de transformer l’information contenue dans les bases de données en produits utiles à la société en général et aux décideurs en particulier. Tel était l’objectif de la conférence tenue du 2 au 4 octobre 2007.

Les allocutions de bienvenue ont été prononcées par Thomas Sephton (Ph.D.), en sa qualité de directeur par intérim de l’Institut océanographique de Bedford, et par Fred Grassle (Ph.D.), directeur fondateur de l’Institute of Marine & Coastal Sciences de la Rutgers State University. Quant au discours-programme, il a été prononcé par Ron O’Dor (Ph.D.), le scientifique principal du Census of Marine Life. La conférence a donné lieu à 37 exposés oraux, organisés en sept séances : 1) les outils de visualisation, 2) les habitats et les écosystèmes, 3) les noms d’espèce et les outils de gestion, 4) les progrès dans les métadonnées 5) l’utilisation et l’analyse des données, 6) les nouveaux systèmes de données et 7) l’intégration de divers types de données. Chaque séance débutait par une introduction d’un président ou d’une présidente de séance et se terminait par une discussion dirigée entre experts. Plus de 100 chercheurs, gestionnaires de données et entrepreneurs appartenant aux secteurs gouvernemental, universitaire et privé, dont beaucoup venus d’autres régions ou pays, ont pris part à cette conférence. En général, les participants ont trouvé celle-ci productive et ils ont formulé des recommandations claires sur ce qu’il convient de faire pour placer l’informatique appliquée à la biodiversité dans une perspective planétaire.

INTRODUCTION

At the International Oceanographic Data and Information Exchange (IODE XVIII) conference on 26-30 April 2005 in Ostende, Belgium, Dr. Edward Vanden Berghe reported on the 2004 Ocean Biodiversity Informatics (OBI) conference in Hamburg, Germany (http://www.vliz.be/events/obi/) and on the 2002 Colour of Ocean Data conference in Brussels, Belgium (http://www.vliz.be/events/cod/). A major point of discussion at both these conferences was data policy issues and the advantages of free and open sharing of biodiversity data. It was recognised, however, that much work remains to be done. The representative of the International Council for Exploration of the Sea (ICES), Ms. Julie Gillin, suggested a second OBI conference to focus on biological data quality issues and to continue the valuable cross-disciplinary interchange.

The resulting OBI ’07 conference (2-4 October 2007) was co-organised by IODE, ICES, and the Ocean Biogeographic Information System (OBIS). It was jointly hosted by the Centre for Marine Biodiversity and the Bedford Institute of Oceanography (BIO) in Dartmouth, Nova Scotia, Canada. The unifying theme for this conference was enhancing data quality assurance, as well as extending the range of information products that biodiversity informatics practitioners must be developing and delivering to satisfy the needs of users on many different levels (e.g., general public, academics, decision makers, and industry). In other words, translating the data housed in databases into products that are useful for society in general and decision makers in particular. The primary challenge was to show how ocean biodiversity data can be made useful beyond satisfying the curiosity of the principal investigator.

In addition to presentations on ocean biodiversity informatics, the OBI ’07 conference also included a poster session, demonstrations of new hardware and software products, guided tours of the BIO facilities, a reception at BIO (2 October), and a banquet at Pier 21 National Historic Site (3 October).

The OBI ’07 conference was attended by a wide range of individuals from a variety of backgrounds, including data managers, government, and industry (see Appendix 2 for a complete list of participants). The primary focus was on oral presentations, followed by questions and panel discussions by specialists presently engaged in ocean biodiversity informatics. This report represents a summary of the presentations and discussion from the OBI ’07 conference.

The conference website can be accessed at http://www.marinebiodiversity.ca/OBI07.

OPENING REMARKS AND KEYNOTE ADDRESS

Opening Remarks

Dr. Thomas Sephton, Fisheries and Oceans Canada (DFO) Maritimes Region, Canada

On behalf of the local management team, Dr. Thomas Sephton welcomed participants to BIO. The following is Dr. Sephton's remarks.

"There are many challenges facing ocean biodiversity informatics in both the short and long term, and the discussions coming out of your conference program will address many of these over the course of the next 3 days. The final outcomes will only be as successful as its level of participation and interaction, and you will have plenty of opportunities to get involved. From a Canadian perspective, I can state that DFO is in this for the long haul. From both a legislative regulatory mandate and a policy administrative framework, we have institutionalised biogeographic and biodiversity data management and visualisation into our way of doing business. Contrary to popular belief, DFO does not manage stocks, habitat, coastal regions, ecosystems, or oceans. What the Department does manage are activities such as fishing, aquaculture, hydrocarbon and mineral extraction, whale watching and ship transit, in an effort leading to the Integrated Management of Ocean uses. All of this is in support of the ecosystem approach to management, in consultation with our resource managers and clients, and with a thorough understanding of their activities and needs. DFO manages activities through objectives: conservation, social and economic, and their regulatory Management Plans. Conservation Objectives, based on regional assessment processes for harvest fisheries, are typically expressed as: Productivity, Biodiversity, and Habitat. More recently, this experience has moved into the world of setting conservation objectives. All this revolves around our ability to develop, update, visualise, and interrogate the relevant biodiversity datasets. Thus, the importance of your conference to our global Departmental mandate in the long term. I wish you every success over the course of your conference, and I leave you in the very capable hands of the conference and local organising committees."

Welcome

Dr. Fred Grassle, Director of the Institute of Coastal and Marine Sciences, Rutgers University, USA

Dr Fred Grassle is recognised worldwide as the principal founder of OBIS and, as such, has contributed much to the proper management of natural resources. Following is Dr. Grassle's personal description of OBIS.

"OBIS publishes primary data on marine species locations to the Internet to facilitate data discovery and exploration by species, higher taxa, time, location, and depth. Features include: mapping, overlaying species distributions on ocean environments, and modeling of potential environmental range. OBIS caches species distribution data from databases distributed around world, creates taxonomic and geographic indices, seeks out new datasets, develops standards for data exchange and management, and develops software tools for online use, such that all data are freely accessible online. OBIS adheres to standards whenever it can. OBIS biogeography is founded on the Global Biodiversity Information Facility/Taxonomic Databases Working Group (GBIF/TDWG) Darwin Core; discovery is derived from the Global Change Master Directory (GCMD) at the National Aeronautics and Space Administration (NASA), and taxonomy is based on the Catalogue of Life. OBIS currently serves data from 231 authoritative sources and has 13.6 million records on 80,000 species in its cache. Of these, 6.9 million records are at the genus level and below. OBIS is among the largest providers to GBIF. As for governance, OBIS is a federation of organisations that share a vision -- free and open access to biodiversity information. OBIS has several committees, including the International Committee for Administration and General Management (soon to be replaced by a Governance Board), a Managers Committee for the Managers of Regional OBIS Nodes, and a Science Board. Now for some nuts and bolts. OBIS is a distributed system with a provider part installed at the site of the data provider and a portal part at Rutgers University that can be consulted by the end-users. Distributed Generic Information Retrieval (DiGIR) is the software which allows semantics to be decoupled from protocol and database structures. OBIS uses 'federation schema' defined in the extensible Markup Language (XML) specifying which elements will be exchanged and how they will be labeled. Data exchange and query are all formulated through XML. As for limitations, we don't know the total biodiversity. Historic sampling was very selective, both taxonomically and geographically (mostly big things like vertebrates in surface waters in temperate zones). This, however, leaves lots of opportunity for new discoveries. Plans for the future at OBIS include: thematic portals for fisheries (zeroes and sampling effort) and invasive species (harmful algal blooms), as well as demonstrator projects showing such things as predictive modeling of changes in species distribution in response to global warming. Current priorities include: filling some of the gaps, collaboration with existing Regional OBIS Nodes, creating new Regional OBIS Nodes, and creating a World Register of Marine Species (WoRMS). Generally speaking, OBIS is set on having at least one distribution record per species, preferably the type locality. Finally, the importance of OBI to OBIS. OBI is the principal provider of new tools and of data, as well as a forum to discuss future developments, forge collaboration to encourage scientific discipline."

Keynote Address
Dr. Ron O'Dor, Chief Scientist of the Census of Marine Life

Dr Ron O'Dor is chief scientist of the Census of Marine Life (CoML) and is renowned for the having the broadest possible of global marine research. Following is his keynote address to the conference participants.

"The Census of Marine Life is a decade-long program (2000-2010) to assess and explain marine life's diversity, distribution and abundance; past, present and future; covering the known, the unknown and the unknowable. Starting from a few workshops in the late 1990s, and several pilot projects in the early 2000s, CoML now consists of an International Steering Committee, with many Regional Steering Committees, major theme projects including the History of Marine Animal Population (HMAP), Future of Marine Animal Populations (FMAP), Ocean Biogeographic Information System (OBIS), and numerous global field projects. When this, our first, census ends in 2010, we expect to have a Global Ocean Observing System for Marine Line and OBIS; together adding up to a Global Marine Assessment. Simply stated, CoML is a here and test of partnerships and mechanisms for sustainable development. The grand challenge question underlying CoML is: "What did, does, and will live in the oceans? HMAP has, for example, investigated the Scotian Shelf Salem Beverly ship logs of the nineteenth century, while FMAP has developed global visualisations in species richness, as well as tracked global decline in ocean predator diversity derived from Japanese longline fishery data. CoML field projects are basically grouped as follows:

a) Human Edges: Natural Geography In Shore Areas (NaGISA), Coral Reef Ecosystems (CReefs), Gulf of Maine Area Census (GOMA), and Pacific Ocean Shelf Tracking (POST),
b) Central Waters: Tagging of Pacific Predators (TOPP), Census of Marine Zooplankton (CMarZ), and Mid-Atlantic Ridge Ecosystems (MAR-ECO),

c) Hidden Boundaries: Continental Margins Ecosystems (CoMargE), Census of Diversity of Abyssal Marine Life (CeDAMar), Census of Seamounts (CenSeam), and Chemosynthetic Ecosystems (ChEss),
d) Ice Oceans: Arctic Ocean Diversity (ArcOD) and Census of Antarctic Marine Life (CAML), and
e) Microscopic Ocean: International Census of Marine Microbes (ICoMM)."

Continuity
Dr. Edward Vanden Berghe, Executive Director of OBIS International, Rutgers University, USA

Dr. Edward Vanden Berghe was one of the main organisers of the Hamburg OBI meeting in 2004, and one of the lead authors of the conference proceedings. The following is Dr. Vanden Berghe's overview.

"There have been 3 previous meetings: 1) Hamburg 1996 - IODE meeting of a small group, 2) Brussels 2002 - a major meeting bringing together physical and biological oceanography, and 3) Hamburg 2004 - a major meeting covering new developments in ocean biodiversity informatics. The first Hamburg meeting organised by IODE was a small meeting meant to advise IODE how to deal with biological data. After more discussions at IODE XVIII in Lisbon, IODE created the Group of Experts on Biological and Chemical Data Management and Exchange Practices. The Brussels meeting was called the 'Colour of Ocean Data'. It was organised by the Flanders Marine Institute (VLIZ), IODE, CoML, and sponsored by the European Union, Flemish, and Belgian governments. This meeting was held to bring together different communities including biologists and physical oceanographers, as well as data managers and data users. The second Hamburg meeting was the first true Ocean Biodiversity Informatics Conference. It was organised by the German National Oceanographic Data Centre, VLIZ, IODE, ICES, and CoML/OBIS. Its intention was to take stock of developments, and resulted in the 'Hamburg Declaration' (http://www.vliz.be/events/obi/statement.php). Now we have OBI '07 here in Halifax, Canada, and have an opportunity to start consolidating the community particularly by hearing about new developments and how they can be shared. Publication has always been an important element of our meetings. The Colour of Ocean Data meeting resulted in proceedings, whereas OBI gave rise to a theme section of the Marine Ecology Progress Series (http://www.int-res.com/abstracts/meps/v316/) and proceedings (http://www.vliz.be/events/obi/Proceedings20070119.pdf). As for OBI '07, we will have proceedings and perhaps a journal publication. What about the future? Should we make plans for new separate conference? Alternatively, we might have a session in another meeting, perhaps the Taxonomic Database Working Group (TDWG, now known as Biodiversity Information Standards) or the Marine Sciences World Conference. We will discuss this all at end of this conference. Let's begin!"

PRESENTATIONS AND DISCUSSION

The following is a record of the oral presentations and subsequent discussion. Presentations have been posted to the OBI '07 website (www.marinebiodiversity.ca/OBI07) and can be accessed by clicking on the presentation titles below.

Session 1: Visualisation Tools

Session Chair: Peter Lawton, DFO Maritimes Region, Canada

Visualisation of Passively Tracked Marine Species using a Web-Based Mapper and the OBIS Schema

Prepared and presented by: Jerry Black, DFO Maritimes Region, Canada

Synoptic visualisation of collections of tagged marine species data provides a view into the spatial dynamics of these species. Information on the biogeography, patterns and timing of migration, and residency may be inferred from passively tracked marine species. This work demonstrates both the creation of static maps showing aggregated species movement between detectors and the creation of movies to animate collections of animals at the individual tag level.

This work covers the visual display of *passively* tracked animals. The final product will be available through the international portal of OBIS (iOBIS, http://www.iobis.org/). OBIS was established by the Census of Marine Life program (http://www.coml.org/) and is a strategic alliance of people and organisations sharing a vision to make marine biogeographic data, from all over the world, freely available over the World Wide Web.

Other good examples of animal tracking may be found at the Tagging of Pacific Predators (TOPP) website (http://topp.org/).

The objective of this project was to provide a mechanism for visualising animal tracks based on the standard OBIS schema that provides an *interpreted view* of tracked animal distribution based on assumptions of distribution that include: avoidance of traveling across land boundaries and the following of a minimum path between observation points. The problems such as the reuse of tags and maximum swimming speed on the potential path were ignored.

The example data used in this work was from the Pacific Ocean Shelf Tracking (POST) project (http://www.postcoml.org/), also a contributor to the Census of Marine Life. A subset of the POST data is available through the international OBIS portal hosted by Rutgers University. The POST data set consists of observations of individual acoustically tagged fish as they pass arrays of acoustic detectors positioned along the northern Pacific coast of North America. The species used included: Chinook salmon *(Oncorhynchus tshawytscha)*, Coho salmon (*Oncorhynchus kisutch)*, Dolly Varden *(Salvelinus malma)*, Steelhead *(Oncorhynchus mykiss)*, and Sockeye salmon *(Oncorhynchus nerka)*.

Currently the ACON mapper (a scriptable Common Gateway Interface application written in C/C++, http://www.mar.dfo-mpo.gc.ca/science/acon/) at iOBIS is capable of showing the aggregate distribution of detections by species (Figure 1).

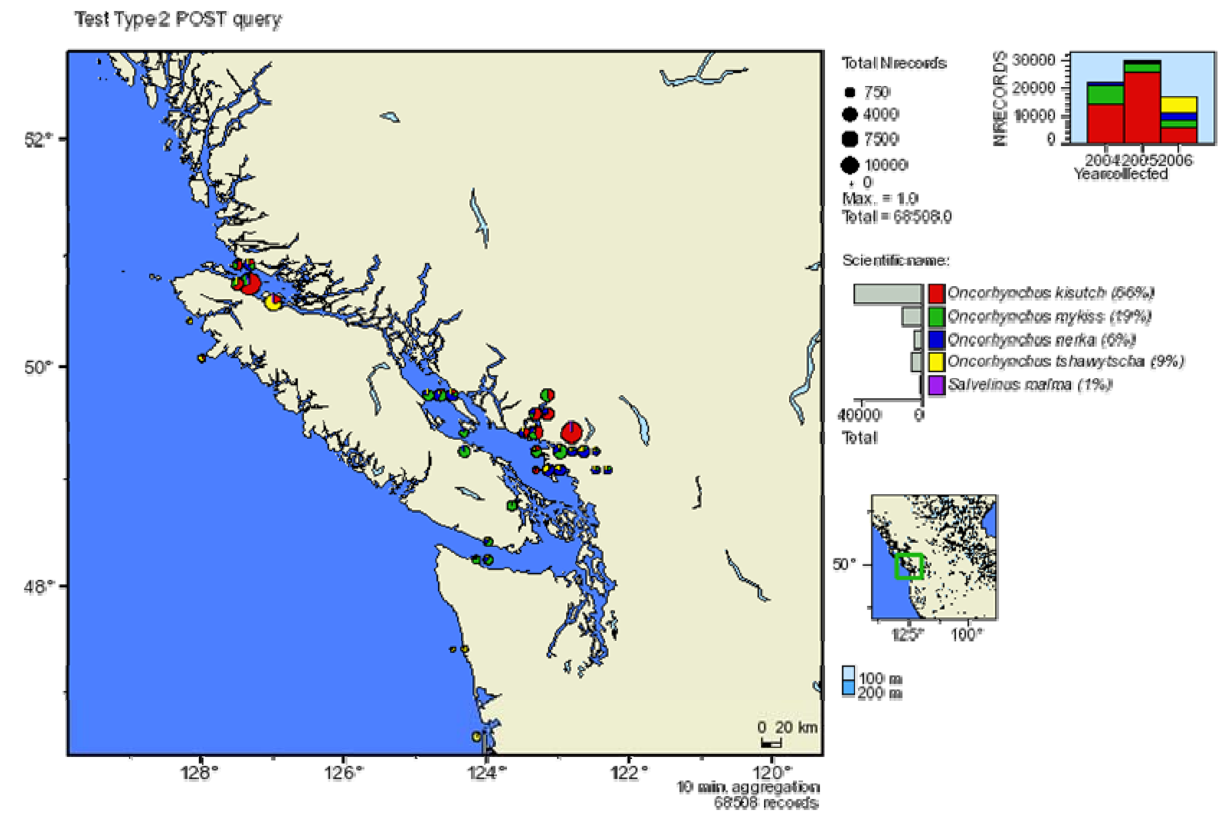


Figure 1. ACON generated map showing an example of aggregated POST detections by species.

When one attempts to show the tracks of individual animals from sparse detectors, simple linear interpolated tracks ignore the issue of intersection with shorelines (Figure 2).

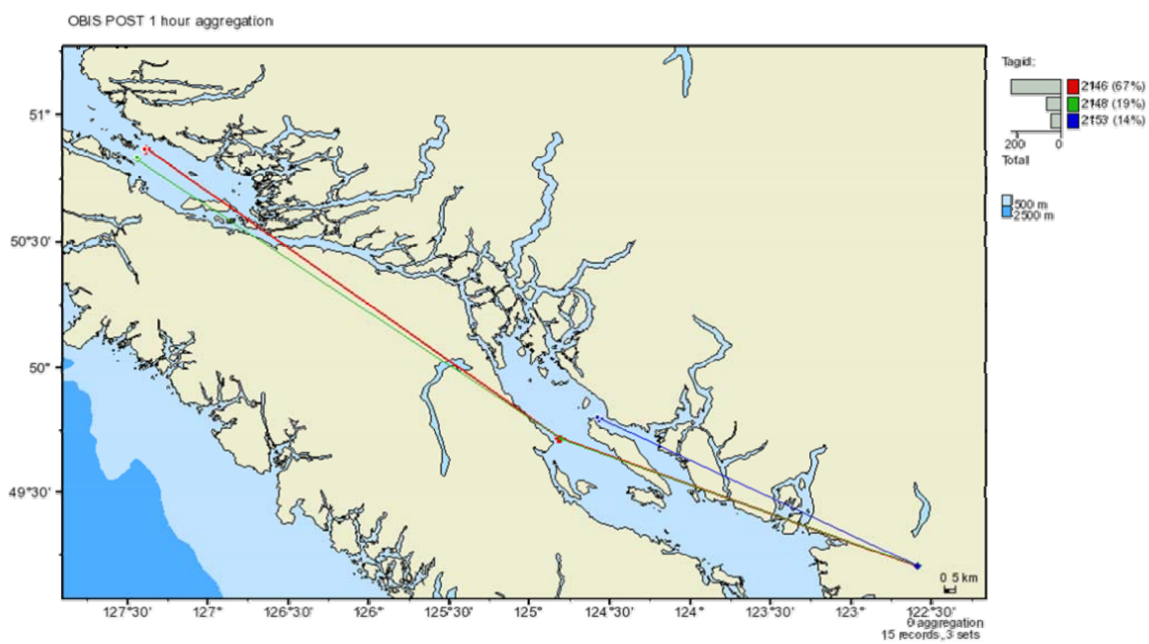


Figure 2. An ACON map showing the problem with generating a track of an individual animal using simple linear interpolation of sparse detections.

The application was extended to retrieve the selected data from the OBIS database, calculate the minimum path tracks that avoid crossing land, and generate maps, movies, or Google Earth (©2007 Google) Keyhole Markup Language (KML) files as output. This is achieved by: aggregating the POST data to 1 hour observations by tagID, creating aggregated detection sites determined from localities (white points), adding generic interpolated points that occur with the spatial extent (red points), adding interpolated connection edges (that do not intersect land), and, finally, determining minimum paths (A – B) from the available connection edges using a minimum path calculation. For animation, speed must be calculated and positions interpolated along the time domain. The result of these computational steps is individual animal tracks that appear to mimic natural behaviour (Figure 3).

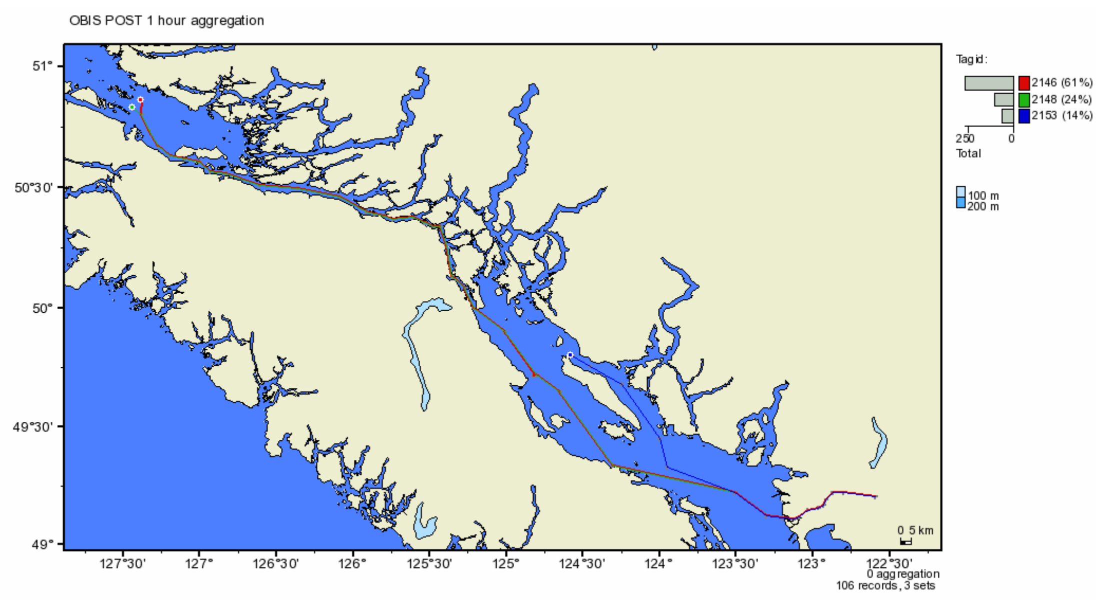


Figure 3. More realistic tag tracks generated using approach described above.

Using the KML 2.1 Time Span attribute, these tracks were able to be generated as animated tracks for use in Google Earth (©2007 Google) using the LineString primitive to describe individual line segments of a track (Figure 4).

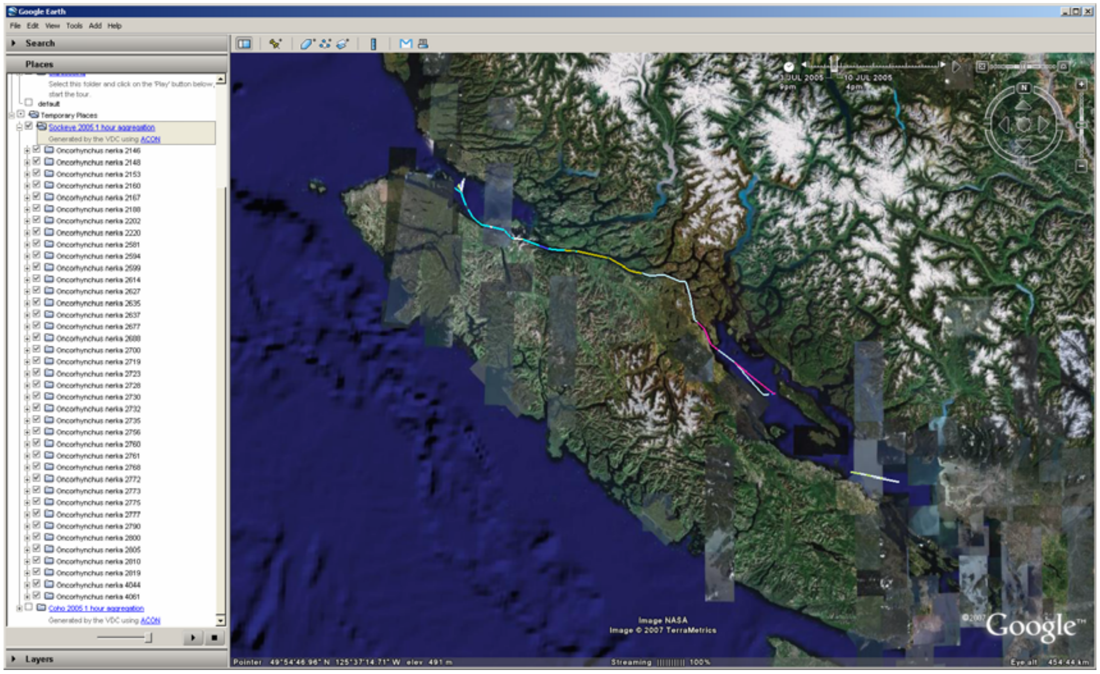


Figure 4. The same track generated using Google Earth (©2007 Google).

<u>The Image Archive: A Data Solution</u>

Prepared and presented by: Pierre Clement, DFO Maritimes Region, Canada

Imaging has been, and continues to be, a key method in investigating ocean related phenomena. Still images have been collected for since the 1920ss, and with the advent of new, less expensive digital technologies, video imagery has become a very important source of research data.

Images are collected and processed for a variety of research interests including: aquaculture interactions, stock assessment, animal production indices, undersea surface geology and benthic community structure, species identification, behaviour studies (e.g., Right Whale, mussel feeding), fish aging (otoliths and scales), invertebrate behaviour, the oxigenation of coastal sediments, etc. New discoveries, like deep sea corals and a desire for non-invasive means of assessments, suggest that the reliance on images will only increase. A systematic infrastructure is essential to move forward in the scientific interpretation of these images.

Fundamentally, all these data have both temporal and spatial components. Acquiring and preserving the metadata and integrating these data into a storage system, with 'quick' retrieval, are key to their use by present and future investigators.

This presentation will describe a pilot project to develop standard collection practices and implement a web accessible archive to manage and retrieve images collected in the field.

<u>Questions</u>

Question: There are now systems where you can mark video and note slides with items of interest so scientists can go back afterwards. Have you used these?

Answer: We hope to do this. At present, some preliminary analysis is done as the data is being collected, e.g., noting the presence of haddock.

Question: How do you decide how much of a clip to play?

Answer: This is a good question, and we will show this in more detail in a subsequent presentation.

Web Services for Mapping Marine Megavertebrates: OBIS-SEAMAP
Prepared by: P.N. Halpin, A.J. Read, B.D. Best, E. Fujioka, L.J. Hazen, and B. Donnelly
Presented by: Patrick Halpin, Duke University, USA

The OBIS-Spatial Ecological Analysis of Megavertebrate Populations (OBIS-SEAMAP) project provides a suite of advanced web services for the storage, analysis, and visualisation of geospatial datasets related to the biogeography of marine mammals, seabirds, and sea turtles. Boat, beach, and aerial surveys, in addition to telemetry data, are submitted by government, academic, industry, and non-profit providers through a Plone content management system. Taxonomies are matched and mined through the Integrated Taxonomic Information Service (ITIS) XML service. Point observations are fed to the OBIS portal via the DiGIR software and XML-based Darwin schema. Data are stored with an open-source database / spatial adapter / internet map server stack (PostgreSQL (SQL, Structured Query Language) / PostGIS (GIS, Geographic Information System) / MapServer). The rich browser-based mapping interface additionally builds upon Google Maps and asynchronous Javascript and XML (AJAX). Open Geospatial Consortium (OGC) web services provide image (Web Map Service, WMS) and XML (Web Feature Service, WFS) representations of the observations and effort. OBIS-SEAMAP archives more than 200 datasets representing over 1.1 million observations. With 3 additional years of National Oceanographic Partnership Program National Science Foundation funding secured, it is anticipated this archive to continue growing while developing new data types (acoustics, photo-id, 3D dive profiles, distribution model outputs), access controls, Open-source Project for a Network Data Access Protocol (OPeNDAP) and Simple Open Access Protocol (SOAP) web services and client libraries for environmental sampling, and automated population of metadata clearinghouses.

The Rare Biosphere Visualised
Prepared and presented by: Phillip Neal, Marine Biological Laboratory, Woods Hole, USA

The International Census of Marine Microbes (ICoMM) has been hard at work organising pilot genomics projects to census marine microbes. The amount of data generated by these projects is tremendous. As of May of 2007, there are more than 2.5 million tag sequences from over 90 samples in the ICoMM database. As a first step in analysing this data, the ICoMM bioinformatics team has developed a quick and simple online application processing system to visualise this data. This system was demonstrated and plans for enhancements were discussed.

A New Automatic Routine for Zooplankton Counting, Measurement, and Identification
Prepared by: F.G. Taboada, J. Höfer, F. Alvarez-Marques, and R. Anadon
Presented by: Fernando González Taboada, University of Oviedo, Spain

Classical zooplankton samples analysis is a time consuming labour that requires a high degree of specialisation. The development of accurate and reliable methods based on the image analysis of zooplankton samples is a promising approach, especially if the process is automated. This could change the way oceanographers design zooplankton time series programs, improving, therefore, the capability to assess the response of pelagic ecosystem to human induced perturbations such as climate change and fisheries depletion. Despite the potential usefulness of such a kind of methods, scarce research has been devoted to assess the performance of currently available tools. In this presentation, the pros and cons of different approaches to the automatic processing of zooplankton samples are reported. Based on these results, a new tool is proposed, developed under MatLab, to optimise plankton counting, measurement, and identification.

Panel Discussion

Pierre Clement was asked “How you find what you’re looking for when there is large amounts of data to browse through. Also, how do you maintain a connection between video frames and continuous audio? How does ‘freeze frame video’ impact the audio quality?”

Pierre responded by saying that you can control framing and playback, but the quality depends on the compression (i.e., codec) used. Any manipulation of framing affects the audio. They are currently doing this with the Remotely Operated Platform for Ocean Science (ROPOS). Using MPEG-4 (a collection of methods defining compression of audio and visual digital data), they were able to manipulate the frame counts but lost audio quality. He suggested that you do not need every bit of video.

The choice of variables that allows you to do classifications of benthic organisms was identified as an issue of concern. It was felt that information needs to be put into classifications that taxonomists actually use.

Jerry Black was asked whether Oracle™ provides support for his work on ACON. However, Jerry does not use Oracle™ for ACON, he uses C++. Each track was generated on demand, but they could be archived. It takes about two seconds to get a track back.

Someone asked how expensive Geo2000 4D was. It was noted that this is one possible media, but there are other options available. Other options for visualising biological data in 3 dimensions (depth and space) were discussed, including Google Ocean. Right now, if someone tries to zoom into the water using Google Earth (©2007 Google), they crash into the water. Negative depths are required to make this work properly. However, Google seems to be taking this request seriously.

Session 2: Habitats and Ecosystems
Session Chair: Ellen Kenchington, DFO Maritimes Region, Canada

<u>Advancing Ecology in Species Data Systems by including ‘Habitats’ and ‘Ecosystems</u>
Prepared and presented by: Mark J. Costello, University of Auckland, New Zealand

The exchange and combination of biodiversity data from different sources uses standardised species names. These names can be grouped in hierarchical classifications that enable browsing down to find the species of interest, and to select groups of related species for more detailed analysis. Including the ecology component of biodiversity in data systems requires classifications of habitat terms. In addition to data systems such as OBIS, a wide range of intergovernmental, conservation, and fishery organisations require classifications of marine habitats and ecosystems to enable comparisons between areas, and to organise information in maps and reports. However, all of the terms used to describe habitats are concepts whose definition is context-dependent. For example, the habitat of a benthic invertebrate is very different in spatial scale to that for a parasite, plankton, tuna, or whale. An ecosystem can be physiographically defined as a lagoon, seamount, estuary, abyssal plain, or entire ocean. Different sampling methods will define different regions, such as satellite images of ocean colour, acoustic maps of the seabed, in-situ sampling of water or sediment cores, and maps derived from analyses of species distributions that may define biogeographic regions. To date, biogeographic boundaries have been more defined by expert opinions derived from selected taxa, than analysis of species distribution data. However, it will be possible to analyse data in OBIS for such boundaries across taxa in space and time. This presentation will review classifications in use as global regions and for characterising field data, and recommend present and future options for data management.

Questions

Comment: It was suggested that there should be a Wikipedia entry for OBIS [entry added after the meeting: http://en.wikipedia.org/wiki/OBIS] and that Northwest Atlantic Fisheries Organization (NAFO) regions should be added to OBIS also.

Question: Are the geographic coordinates available for boundaries that are used as search polygons in OBIS, e.g., the Longhurst boundaries?

Answer: While OBIS does make use of these boundaries, they are not yet easily available through OBIS. However, they are available elsewhere, such as through VLIZ.

Mapping Colombian Caribbean and Pacific Bottom Seascapes and Marine Ecosystems
Prepared by: D. Rozo, M. Vides, P. Lozano, and J. Reyes
Presented by: Martha Patricia Vides Casado, Marine and Coastal Research Institute, Colombia

The present study has produced a bottom seascapes map of Colombian Caribbean and Pacific territorial waters. This map is part of a national initiative to generate an official 1:500,000 scale map where terrestrial, marine, and coastal natural ecosystems are all shown. Mapping approach is based on a benthic realms conditions classification using enduring physiographic and recurrent oceanographic features that are known to affect distribution of communities and their component organisms. Seascape modeling is based on a digital depth model using ground true points obtained from various time and scale bathymetric charts. Data for the benthic realms are assembled and classified in an ecologically meaningful way and then combined (overlaid) to derive bottom seascapes maps. Each benthic seascape is uniquely defined by the combination of bottom geomorphology, sediment types, and oceanic climate values or ranges. Through expert knowledge, undersea features were identified using slope maps, isoclines, and 3-dimensional visual analysis. Sediment types were obtained from the digitisation and generalisation of sedimentary facies charts. Ocean climate was classified using physical and chemical oceanographic records from the World Ocean Atlas 2001. Coastline, mangrove, and coastal lagoons extent was obtained by visual interpretation, region segmentation, and unsupervised classification using 3, 4, and 5 bands of Landsat 7 ETM+ dataset. Available ecosystem level information on reef areas and sea grasses was generalised and superimposed to the generated seascapes. Regionalisation of Colombian underwater territory ranges from Realms, Provinces, Ecozones to Ecoregions, and encompass a total of 155 seascapes extending 532,200 km^2 of the Caribbean Sea and 360,510 km^2 of the Pacific Ocean. Seascapes distribute up to a maximum of 4,990 metres depth. Distinctive features of the Caribbean include the Colombia Basin covering 210,000 km^2 Magdalena Fan covering an area of about 24,400 km^2 in close relation to the river delta. Extensive coral reefs top the mountainous seascapes of the Tayrona Ridge Ecozone NE of the Mesoamerican platform. Malpelo Ridge and Pacific Rise are distinctive features over the Pacific Basin. The Colombia trench, running parallel along the coast, is a characteristically very deep and asymmetrical depression of the Pacific sea floor. Upcoming work should point to seascape accuracy and verification. This map serves as a basis for future management proposals, conservation plans, and sustainable management planning.

Questions

Question: You appear to have used Principle Component Analysis to look at associations. Did you look at others?

Answer: We had to develop one layer for climate – used seamapper from NOAA and combined with own data. We tried others, but this was the best result (validation with biology).

Question: What software did you use to visualise 3-dimensional benthic maps?

Answer: Rmap, a package for plotting geographic maps in R.

Session 3: Species Names Management and Tools
Session Chair: Lou Van Guelpen, Huntsman Marine Science Centre/Atlantic Reference Centre, Canada

Improved Search and Retrieval of Ocean Biodiversity Information through NASA's Global Change Master Directory
Prepared and presented by: Melanie F. Meaux, Global Change Master Directory, USA

To improve the search for marine and non-marine species around the world, NASA's Global Change Master Directory (GCMD, http://gcmd.nasa.gov/) science keyword hierarchy has been extended, and a new Science Topic called ‘Biological Classification’ has been added. Users may now navigate a five-level hierarchy of keywords within the directory's redesigned web site for more precise results. This added functionality arose from collaboration with representatives of the Ocean Biogeographic Information System (OBIS, http://www.iobis.org) and resulted in the implementation of the OBIS taxonomic categories. The Biological Classification keywords were further extended based on ITIS, Species 2000, and GBIF taxonomic sources. The new functionality will illustrate the improved search and retrieval of ocean biodiversity information. The GCMD OBIS portal (i.e., a customised virtual subset of the directory) will also be demonstrated. Additional features will be presented such as an improved metadata display and new functionalities within the metadata authoring tools.

Applications of Fuzzy (Approximate String) Matching in Taxonomic Database Searches, with an Example Multi-Tiered Approach
Prepared and presented by: Tony Rees, CSIRO Marine and Atmospheric Research, Australia

Scientific names of species and genera can often be difficult for even the expert user to spell correctly (for example, *Cirrhitichthys oxyrhynchos* or *Syzygotettix boettcheri*); however, since most taxonomic information is keyed to a particular scientific name, user inability to correctly spell the name in a relevant ‘search’ box of a particular taxonomic information system all too frequently means that the desired information is not retrieved.

A few such systems currently incorporate a ‘fuzzy match’ or ‘near match’ option (of the “*did you mean?...*”) variety, also known in the computer literature as ‘approximate string matching’, that attempts to correct for possible user spelling errors and, thus, increase the chance of correctly retrieving the desired taxon information. A phonetic-based algorithm, devised with particular consideration for the nature of species scientific names, was developed by the author in 2001 for use in his agency’s Codes for Australian Aquatic Biota (CAAB) online taxonomic database, and has subsequently been adopted in species databases at the Flanders Marine Institute in Belgium and OBIS in the USA, with a corresponding increase in successful ‘hits’ against misspelled user input (and in some cases, misspelled or variant spelled data content, as well). More recently, this approach has been expanded further to cope with additional types of misspelled input not covered by the original algorithm, resulting in a set of algorithms that can be run in succession at user query time, with little additional time penalty, that considerably increase the chance of a ‘correct’ match against potential misspelled user input.

Prior to the application of any near match techniques, a range of checks should be applied to detect, and, where necessary, correct for structural mismatches such as the use of qualifiers cf., aff. ‘?’, near, etc; stray html tags (<i>, &, etc.); family names in genus fields; non-alphanumeric characters; and leading or trailing blank spaces. The approach presented here then applies 2 separate tests in succession, an upgraded custom (Rees 2007) phonetic algorithm and an edit distance test, that attempt to correct for phonetic and non-phonetic errors, respectively. A degree of normalisation of the endings of specific epithets (e.g., *vulgaris* versus *vulgarus* versus *vulgarum*, etc.) is also incorporated into this version of the phonetic algorithm, which can run fast because all computations on the potential target terms can be run in advance, and stored as a separate indexed database column in the relevant data table.

The edit distance testing involves the application of a standard Levenshtein Distance test (e.g., see http://en.wikipedia.org/wiki/Levenshtein_distance) separately to the genus and (where available) specific epithet components of the species name, and setting appropriate thresholds for pass/fail at each stage. On account of the dynamic nature of this test, which determines the number of single character insertions, deletions, or replacements that will turn ‘string A’ into ‘string B’, such comparisons must be done live at query time against every available target term, at a speed in the order of a few milliseconds per test. This means that to test a single input name against every target in a large database (e.g., the 1.4 million valid species names and synonyms in the 2007 Catalogue of Life compilation) would be prohibitive, in the order of 1 hour or more. The approach adopted to address this is twofold: first, build a table of genus names separate from the full taxon name index, which will typically be in the order of one tenth of the size of the latter table, and apply the test hierarchically to first genus, then species only of genera passing the initial test; and second, introduce a set of pre-filtering heuristics that limit the names to be tested (in each case) to a relatively small subset, based on word length, leading and/or trailing character matches, and more. By this means, a typical genus test can be reduced to operating on the order of 1000 names or less, and take (typically) one to a few seconds only, with a relatively small number of complete genus+species combinations to be tested fully (typically a few dozens or hundreds only).

Optimising the system for real world use involves a number of considerations, such as devising and optimising the heuristics (rules) for subsetting both at genus and specific name stages; setting the pass/fail thresholds, which can also be dynamic (e.g., with shorter words, a smaller number of characters may be allowed to be different before a ‘fail’ state is allocated); and setting of criteria for presenting the results. In the present test system, based on the author’s Interim Register of Marine and Nonmarine Genera (IRMNG) database accessible via http://www.obis.org.au/irmng/, a ‘tiered’ approach is adopted, where if no highly similar results are found (e.g., one or two characters different), the threshold is progressively widened up to a pre-set limit, so as to effectively mimic what a human user might do (i.e., if nothing detected with a narrow threshold, iteratively widen it to see what results might be obtained with a different setting).

As an example, searching for the misspelled marine species name '*Hoploxtethis atlanticum*' via the current IRMNG search entry point yields a result in a little over 2.3 seconds, pointing to the correct name '*Hoplostethus atlanticus',* as well as a misspelled version '*Hoplosthethus atlanticus*' that is listed as a synonym (misspelling) on the FishBase data compilation used as part of the source material for IRMNG. In this instance, 10,283 (only) out of a possible 238,471 genus names were tested, followed by 4 (only) out of a possible 1.4 million full species names. If a smaller test database or filter is used (for example, fishes only, comprising 11,061 genera and 74,723 species names in IRMNG at the present time), then these times are reduced by a factor of approximately ten-fold.

A further fuzzy matching routine, this time based on n-grams (which are relatively insensitive to word order), has also been developed for applying to authority names where supplied, in cases where genus and species names are broadly similar but authorities may be quite different. These similarities are reported on a 0-1 scale, as per the following (real) example:

Input name: *Halymenia dilitata* Zanardini

Near matches reported:

Halymenia dilatata Zanardini
  edit distance (genus,species): 0,1; author similarity: 1.0
Halymenia digitata J. Agardh
  edit distance (genus,species): 0,1; author similarity: 0.16
Callymenia digitata (J. Agardh) Kylin
  edit distance (genus,species): 2,1; author similarity: 0.17

In cases such as this, development of an automated routine that takes into account the authority similarity, as well as that quoted for genus and species has an improved chance of successfully identifying the best 'near match' out of several candidate ones.

These near match routines will be continued to be tested and developed further, and can be made available to interested parties on request, for incorporation into their own taxonomic database environments.

<u>Questions</u>

Question: Would you use this to correct your database?

Answer: You could. I use it to normalise my database. You could use it to help merge two databases and sort out the names that don't match. Some would be new names and some may be mistakes.

<u>Aphia: One Step towards a World Register of Marine Species</u>
Prepared by: E. Vanden Berghe, W. Appeltans, and M.J. Costello
Presented by: Edward Vanden Berghe, Rutgers University, USA

This register of marine species grew from the European Register of Marine Species (ERMS) and its combination with several other species registers maintained at the Flanders Marine Institute (VLIZ). Rather than building separate registers for all projects, and to make sure taxonomy used in these different projects is consistent, a consolidated database called 'Aphia' was developed. MarineSpecies.org is the web interface to this database. The World Register of Marine Species (WoRMS) is an idea that is being developed and will combine information from Aphia with other authoritative marine species lists which are maintained by others (e.g., AlgaeBase, FishBase, Hexacorallia, NeMys). Substantial progress towards the completion of WoRMS is imminent, and it is hoped that it will serve as the taxonomic register of OBIS. Some of the design decisions and the governance model are discussed, as are the plans for the future.

The content of the database is controlled by taxonomic experts, not by database managers. Aphia has an editorial management system, where each taxonomic group is represented by an expert who has the authority over the content and is responsible to control the quality of the information. Aphia contains valid species names, synonym and vernacular names, classification, and extra information such as literature and biogeographic data.

<u>Questions</u>

Question: What does IRMNG stand for?

Answer: IRMNG stands for Interim Register of Marine and Nonmarine Genera and is the name of Tony Rees' database.

Question: What criteria are you using to determine marine species?

Answer: Salinity limits (0.5 ppm).

Panel Discussion

Melanie Meaux was asked to elaborate on the controlled vocabulary used in GCMD. The controlled vocabulary helps to normalise searches (like misspellings) and encourages metadata writers to tag keywords. Melanie provides users with a list of keywords that users can choose from. There are also some keywords that are not controlled and new ones are being added all the time. In the future, GCMD is planning to use semantic web to link common names to species names.

Edward Vanden Berghe was asked how open Aphia's name search algorithm is. It is available for use and can be used to search using phonetic name. You can post Excel spreadsheets and get back valid names from Aphia.

Mark Costello was asked what taxonomy he would use for environmental data. Mark felt that you do not need a single hierarchy of habitats; you can use many classification systems. Mark was also asked whether OBIS can be used to investigate the transition zones between boundaries. He felt that boundaries are misleading and that a more dynamic system to explore data and see how boundaries change is needed. Many parameters define habitat. Maps are just an expression of something. There is a need to get better at describing error or assumptions. It was noted that the ICES working group on habitat mapping looked at this issue during their last meeting. Scale is important to habitat structure. It will be difficult to deconstruct habitat structure into a relational database for further exploration. Habitats are dynamic, and ways are needed to show this dynamic reality.

Martha Vides was complemented on her combination of marine and terrestrial maps. Marine and terrestrial environments merge along the coast, and it can be hard to join bathymetry with topography. This is a big challenge. In the Columbian database, results can be examined seasonally for use in a monitoring system. In order to merge terrestrial and marine environments, one needs to consider common variables.

Session 4: Metadata Developments
Session Chair: Bob Branton, DFO Maritimes Region, Canada

Development of an Australian Marine Community Profile of ISO-19115 – Outcomes and Lessons Learnt
Prepared by: G. Reed, S. Bainbridge, and T. Rees
Presented by: Tony Rees, CSIRO Marine and Atmospheric Research, Australia

Part of the understanding behind the release of the International Organization for Standardization (ISO)-19115 standard for geographic metadata was that individual communities would develop their own profile of the standard to meet the needs of the data collected and described by the community. The community profile would need to keep the mandatory elements, but could choose what other elements it required to describe the data collected, and, additionally, new elements could be added via an extension process where the standard was lacking.

In Australia, a Marine Community Profile has been developed by the Australian Ocean Data Centre Joint Facility, a virtual data centre comprising the main Australian Federal Government agencies involved in the collection and use of marine data (see: www.aodc.org.au). This standard is being used to unify the collection of marine metadata, and through this the description of marine data being collected.

A number of issues have come to light in developing a profile. The ISO-19115 standard is very general and so translating this into a specific profile that has meaning to the end users has been problematic. Related to this, is the fact that many of the ISO elements are simple text elements, and so there is a need to develop rules for how information is structured in these fields to ensure consistency of the information entered. The lack of metadata tools that support the ISO-19115 standard has also been an issue with the group being forced to develop its own using the Geonetwork open source product as its base, along with specific XML Schema Definitions to support the profile. Finally, the need to be specific so that the content is controlled and valid, versus the need to be flexible to incorporate a range of needs results in a tension that requires formal Governance to deal with – the need for formal processes to maintain the profile is one of the major lessons learnt.

The ISO-19115 standard is a beast that needs to be tamed. The Australian Marine Community Profile has taken a path of trying to make the profile as simple and small as possible knowing that individual agencies can add other elements as required. This is in contrast to other profiles that try and keep as many elements as required. At this stage, we feel that the 'lean and mean' approach makes the profile much easier to use and understand, and with good Governance, the tension between making it complex and keeping it simple can be managed.

The Marine Community Profile has been a major tool in increasing the awareness of metadata and data management, and the development of the Profile has brought a focus to how agencies collect and manage their data. The profile has also demonstrated the capacity of the marine community and has helped it be more successful in promoting data management and in gaining funding for future work.

Questions

Question: How does the Marine Profile compare to the North American profile?

Answer: The North American profile is extensive, but doesn't include some marine fields.

The SYNTHESYS Annotation System for Collection and Observation Data
Prepared by: A. Güntsch and W. Berendsohn
Presented by: Anton Güntsch, Botanic Garden and Botanical Museum Berlin-Dahlem, Department of Biodiversity Informatics and Laboratories, Germany

Over the last ten years, the biodiversity informatics community has built a number of international data networks providing open access to massive amounts of collection and observational records according to varying thematic foci. Although portal systems for retrieval of primary occurrence data have improved significantly, a convincing model for annotating records and returning scientific results and data corrections to the collection or database holders does not exist.

In the context of the European Union sixth framework project, SYNTHESYS, a new annotation system has been prototyped and linked to the SYNTHESYS portal for European biodiversity. Users who wish to contribute annotations have to first sign in to receive a username and a password. Annotations are made in the raw XML document belonging to the collection or observation object to be annotated. Thus, the annotation is clearly associated with the annotated concept (e.g., identification or collection site). Finally, the annotated document is stored in a concurrent versions system and made accessible on the World Wide Web. Additionally, annotations are cyclically sent back to data providers as XML files.

The system is generic with regard to the data standard in use, and works with the Access to Biological Collection Data (ABCD) standard, as well as various flavours of Darwin Core.

Further developments of the system will focus on user-friendly data entry forms hiding the underlying XML documents, as well as on software supporting automated re-import of annotated XML-documents into the provider database.

Questions

Question: How do you annotate a whole series of records?

Answer: This is not supported yet, but this would be good to have.

Question: I hate editing raw xml. Why not create forms to do this?

Answer: Software doesn't exist yet. Geonetwork may do this. XML forms may do this.

Rescuing Data Assets and Allowing Discovery using Metadata
Prepared by: A.W. Isenor, C. Young, J. Osler, LCdr W. Renaud, and G. Reed
Presented by: Anthony Isenor, Defence Research and Development Canada Atlantic Region, Canada

For over a decade, marine data and information requirements have been experiencing a transition from an organisation-centric view to a broader community-based view. The community-based view is driven in part by the enabling technology, but also by the realisation that the community has many common data requirements. Groups, such as academia, scientific research, the military, and many government departments, are beginning to share data and information through web-based delivery systems. However, such sharing requires considerable prior effort in the management of these data. As well, methods to discover the data assets are becoming increasingly important.

This presentation outlined three data management activities currently underway at Defence Research and Development Canada (DRDC) Atlantic. DRDC Atlantic conducts research and development in support of operations, acquisition, maintenance, and requirements planning by Canada's forces. Often this requires at-sea trials of research and development systems, including acquisition of various environmental data. Historically, these data have been used to address the specific needs of the trial, with minimal effort placed on the long term management of the data. However, more recently, the recognition of the data as an asset for planning future activities and the need for rapid assessment of the environmental space, has resulted in a coordinated effort to assemble, manage, and provide these data assets.

Activities currently underway at DRDC Atlantic are best characterised as a threefold approach. First, a data rescue activity is underway to identify, consolidate, digitise, and quality control the marine environmental data resident at DRDC Atlantic. This effort involves considerable manual effort as many historic records are only available in paper copy. A second effort involves the creation of an inventory of data collected during DRDC Atlantic field trials and supporting metadata. This effort will allow identification and planning of additional rescue activities by helping prioritise activities based on data importance, level of effort, and funding. A third activity concentrates on providing discovery of data assets via metadata descriptions. Planning is currently underway to provide metadata descriptions using the Australian Marine metadata profile. The profile is based on ISO-19115, and is currently being suggested as the marine metadata profile for the IODE system. The descriptions will be served using service-oriented architecture using Simple Object Access Protocol (SOAP) message protocol. This is a collaborative Canadian/Australian effort dealing with the Meteorological and Oceanographic offices at the two navies.

Metadata - Challenges in Capturing Changes Over Time for a Long Time Series Dataset
Prepared by: D.T. Rutherford and J.A. Boutillier
Presented by: Dennis Rutherford, DFO Pacific Region, Canada

The West Coast Vancouver Island 'shrimp' survey was implemented in 1972, and now serves as one of the longest continuous time series in existence for fish species on the west coast of Vancouver Island, British Columbia, Canada. Basic data collected have remained constant, but changes in technology with respect to vessels, fishing gear, and data capture have occurred. These present challenges when interpreting, analysing, and documenting data. Initiatives such as species as risk, ecosystem assessment, and posting for internet access have resulted in additional demands for the use of this dataset. The need for detailed documentation of this dataset, including details of changes over time, is required, and was presented here. Capturing this time series information in a metadata format is essential.

Questions

Question: Have attempts been made to normalise measurements?

Answer: Yes, these have been peer reviewed for shrimp.

Question: Isn't there a wealth of literature on this dataset already published?

Answer: Yes, but it is broken up into different places. OBIS Canada provides links to the references.

Comment: Metadata uses a one size fits all approach that won't be able to accommodate all issues. There is a need to annotate, but at what level? When I use data off OBIS, I don't always understand what I'm getting. Comments on each record may not be useful, but perhaps providing a comments field for large groups of data would be helpful.

Question: If data was colour-coded by quality, how would the data from this dataset be described?

Answer: Data on shrimp species would be green (i.e., good quality) and data on other species would be yellow (i.e., moderate quality).

Tailoring Fisheries Global Information System Infrastructure to Global and Regional Needs for Fisheries Management
Prepared by: M. Taconet, A. Bensch, R. Lefebure, F. Carocci, A. Gentile, and S. Caillot
Presented by: Marc Taconet, Food and Agriculture Organisation (FAO)

There is an imperious need to revert the current increasing trends of fish stocks overexploitation. Promoting the use of relevant information for sound policy making and management is central to this challenge, and FAO has set-up an enabling framework to new information needs in support to fisheries management, relying on two pillars: the Fisheries Global Information System (FIGIS) – an information system infrastructure, and the Fisheries Resources Monitoring system (FIRMS) – a formal information Partnership.

The famous slogan of the Code of Conduct for Responsible Fisheries claiming that 'Sound policy making and fisheries management should be based on the best available information' has pushed the international community to adopt a new paradigm of fisheries monitoring (Figure 5): the policy and management framework has to identify a comprehensive list of Resources and Fisheries units and to ensure that relevant indicators are available for monitoring and decision making purpose. This paradigm, contributing to the requirements for an Ecosystem Approach to Fisheries, can be implemented in information system terms through a monitoring platform articulated around inventories of Resources and Fisheries.

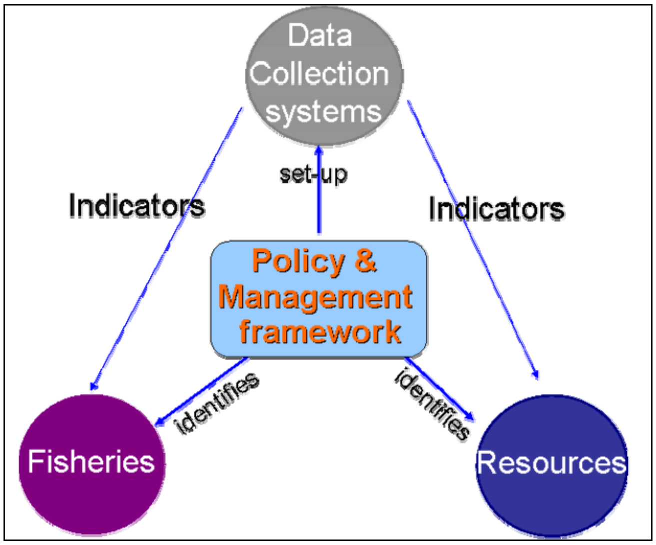


Figure 5. A new paradigm of fisheries monitoring framework.

The FIGIS infrastructure has to respond to these needs at global and regional levels, and should encourage similar moves at national level. A case study on high seas deep-water demersal fisheries in the Northeast Atlantic illustrates how FIGIS brings together the available scarce and dispersed knowledge collated on this subject. This knowledge is organised in four modules (statistics, species, stocks, fisheries) designed to respond to these needs.

The Fishery Statistics module manages workflow and dissemination of regional catch time series.

The Species module compiles within fact sheets taxonomic and biological information, including environmental criteria driving species distribution and related species geographical distribution maps.

The Resource module hosts scientific knowledge about structure of populations, stock units, resources, or stocks assessment. In the area covered by the case study, ICES is mandated to provide information on the structure of deep sea resources, or stocks such as roundnose grenadier for which three stocks units were defined.

The Fisheries module is both the most complex information domain, and that from which breakthrough move is expected for information sharing. Depending on the situations and needs, fisheries are defined through various methods including from interviews with professionals, statistical analyses, literature, and from various thematic approaches. The case study privileges two main thematic approaches for the description of North Atlantic deep sea fisheries. The Production System approach allows understanding of major fleet range of action and strategies, and should be documented with socio-economic performance indicators; states would usually provide this kind of information. The Management System approach should lead to define fishery units (or fishery management units) for management purpose, by considering management objectives within Jurisdictional boundaries combined with the knowledge collated on Resources structure. An example of such fishery unit in the case of Northeast Atlantic Fisheries Commission (NEAFC) would be the *High Seas Rockal and Hatton banks deep water demersal fishery*, where the management rules are certainly distinct from those applying on the *deep water demersal resources* of the same bank but within national jurisdictions. At the cross road of these approaches, more elaborated fisheries tailored to management needs can be defined at fishing operations level; fisheries of concern here would occur on the fishing ground within NEAFC's convention area, while being considered homogeneous from the point of view of fishing methods, catch profile, and socio-economic structure; an example would be *High Seas Rockal and Hatton bank industrial bottom trawlers fishery*. Such type of definition for management purpose finds its most proxy application within the framework of the General Fisheries Commission for the Mediterranean under the Operational Units concept.

Even considering the scarce amount of publicly available knowledge, such information architecture contributes to demystify the subject, and already provides an interesting basis for management support. Further, it provides the required flexibility to evolve towards information reporting needs for advanced management systems, in support to the Ecosystem Approach to Fisheries Management.

This case study shows how critical information exchange is for sound fisheries management: information has to be contributed from various disciplines, national or regional institutions, and countries. Building a platform for sharing information is the key answer. FIRMS, a formal Information Partnership arrangement that FAO coordinates acting as the FIRMS Secretariat, is such a platform. FIRMS' objective is to provide information users with a better means to monitor the status and trends of world fishery resources and their management, based on authoritative information sources. Launched in February 2004, FIRMS currently includes 12 Regional Fishery Bodies and FAO. The FIRMS partnership concentrates its discussions on information sharing mechanisms, including information exchange protocols, streamlined workflow mechanisms, information standards, and quality assurance issues. These sharing mechanisms result in unique products: homogeneous Resources or Fisheries fact sheets, tagged with the agreed standard terms for status and trends description. These baseline products will provide the basis for the elaboration of synoptic products on status and trends of resources by species and regions.

In addition to the well established need to share information at regional level for managing shared resources, the case study herewith presented stresses that information also needs to be shared between disciplines in order to encompass the complexity of multifaceted fishery systems. Thus there is no viable answer to information needs in support to fisheries management without information exchange. With the FIRMS partnership backed by the FIGIS infrastructure, an enabling environment is being set-up for this information exchange purpose: the Resources and Fisheries concepts are comprehensively described in the Fisheries Metadata Element Set (FIMES – an XML schema), which is progressively consolidated as an international standard through test and usage by the FIRMS partners, and will subsequently be made available for use by national institutions. This schema, together with the complementary services of the fishery ontology under development within the NeOn project, will be major tools to ease systems' interoperability.

<u>Questions</u>

Question: Is data collection independent of fishery logbooks or are you demanding new information?

Answer: We are using existing information, and we are partnering with fisheries organisations who have data and will provide this for consolidation.

Panel Discussion

The moderator of this panel discussion, Bob Branton, asked the panel several opening questions: "Will XML ever be easy, and when does metadata become data or move into the realm of literature?"

Dennis Rutherford reversed the question to ask, "Where do I go so that I understand the idiosyncrasies of the data so that when I do analysis, I interpret the results correctly. Looking for answers."

Based on Tony Rees' West African experience, he would say that these issues should clearly be stored as metadata, but agrees that it is difficult to know the limits. For example, "Do you need to store information on gear for each record?" Changes should be captured at the metadata level.

Bob Branton noted that the Maritimes groundfish survey was split into two parts: one for groundfish and one for everything else. The question is how to determine when to split the database.

Tony Rees indicated that the deciding point should be whether the metadata that is being created needs to be machine readable. If only people are going to read it, then plain text can be used instead of XML.

A data user from the audience highlighted the importance of capturing data limitations in ways beyond just the metadata. She felt that it was needed in each record so that when she pulled out individual records, they made sense. It is also useful if protocols for data collection are available. Gear changes can be recorded in individual fields, but there may be higher level issues that would not be included in each record, e.g., management changes.

Marine metadata discussion distinguishes between discovery metadata versus six other types. This was considered to be a critical issue for legacy data. It is not incumbent on the data provider to provide all possible fields. It is incumbent on data users to communicate with provider to find out the necessary details.

A government participant indicated that there are times when, for security reasons, you cannot give people access to databases. It was noted that there is a difference between read accessible and write accessible. There are security protocols that can be followed.

Session 5: Data Use and Analysis
Session Chair: Dr. Edward Vanden Berghe, OBIS

<u>GeoCod: Towards a Better Understanding of Spatial and Temporal Changes of Four Key Species in the Northwest Atlantic Region</u>
Prepared by: R. Devillers, M.-J. Fortin, G. Rose, and M.-A. Mostafavi
Presented by: Rodolphe Devillers, Memorial University of Newfoundland, Canada

This presentation gave an overview of the Canadian project GeoCod that aims at better understanding spatial and temporal changes of 4 key species (cod, shrimp, crab, and capelin) in the Northwest Atlantic region. The GeoCod project is a 2-year initiative involving 3 Canadian universities and several Canadian and international partners from the government, industry, and NGO. The project aims at: 1) integrating existing environmental and fisheries data for the

Northwest Atlantic region, 2) analysing the data using spatial statistics techniques to better understand the relationships between the 4 fish species and their environment, as well as the relationships amongst the species, and 3) develop new tools to visualise and analyse fisheries data. Fisheries scientific surveys, as well as fisheries observer program data of the 4 Canadian Atlantic regions (Quebec, New Brunswick, Nova Scotia, and Newfoundland and Labrador), in addition to the USA data, was collected and integrated into a single database including about 1 million records. This exercise in itself has never been done as data were typically stored and analysed separately by the different regions. This exercise is very challenging as data are very heterogeneous regarding the technologies used to collect them (e.g., type of vessel and gears), the time of the year, etc. These data are currently being analysed, but preliminary results will be presented. In addition, GIS-based visualisation tools are being developed to allow fisheries scientists and decision-makers to easily view and analyse the data. These tools allow to seamlessly connect the different datasets, visualise interpolated surfaces of species abundances (e.g., catch per unit effort), and perform different analysis on the data to better understand spatial and temporal dynamics of the stocks.

Questions

Question: How do you take into account that the data is not collected randomly, i.e., directed/targeted surveys?

Answer: We are currently trying to work out how best to deal with different periods (time/season) for surveys toward standardising on this. We are trying not to play with the data too much.

Fish Stock Depletion Assessment System

Prepared by: Y. Jaques, M. Iglesias, C. Caracciolo, S. Kim, and J. Keizer
Presented by: Yves Jaques, Fisheries and Aquaculture Department, FAO

Fisheries scientists are faced with a rapidly increasing set of data and information systems of potential use to fish stock assessment. As collections grow, problems of relevance and recall in information retrieval are increasing steadily. In addition, scientists are being asked to make ever more comprehensive assessments that include not only traditional biological, catch, and effort data, but also a wide array of societal and ecosystem factors.

The Fish Stock Depletion Assessment System (FSDAS) is a decision-support tool that utilises fisheries ontologies, together with fisheries data sets, to create a queryable, browsable knowledge base. Jointly developed by the Fisheries and Aquaculture Department and the Knowledge and Communication Department of the Food and Agriculture Organisation of the United Nations, it is 1 of 2 proof-of-concept case studies for the NeOn project (http://www.neon-project.org), a 14.7 million Euro project involving 14 European partners and co-funded by the European Commission's Sixth Framework Program.

NeOn began in March 2006 and has a duration of 4 years. Its aim is to advance the state of the art in using ontologies for large-scale semantic applications in distributed organisations. It aims at improving the capability to handle multiple networked ontologies that exist in a particular context, are created collaboratively, may be highly dynamic, and are constantly evolving.

Since the beginning of the project, the case study has identified requirements including use cases and use scenarios, inventoried 140 fisheries information systems, and begun developing a set of modular fisheries domain ontologies in Ontology Web Language (OWL), such as species, water areas, and commodities.

Current work is focused on architectural design of the case study specific components not covered by the generic NeOn architecture, as well as a set of ontology learning experiments to generate relationships across ontologies.

The next year will see a continuation of ontology modeling for other fisheries domains, together with the deployment of the first version of the NeOn toolkit allowing for ontology life-cycle management, as well as connection of ontologies to existing fisheries databases and web services. This will form the beginnings of the actual knowledge base.

The case study will provide the first semantic framework for browsing, querying, and reasoning over fisheries data, including document-like objects, time series, and, hopefully, GIS and sensor data, as well. Utilising NeOn's combination of open-source licensing and pluggable Eclipse-based architecture, FSDAS should provide the fisheries community with a flexible and powerful decision support tool for fish stock assessment.

<u>Questions</u>

Question: On issues of global food and environment, is any industry (commercial) influence received?

Answer: The only commercial involvement is of software providers. As for data from membership, there are country level (i.e., political) interests more than commercial.

Question: Have you given any thought to using a reasoner?

Answer: Yes, we are using two different reasoners so as to be able to compare/play against each other. Both will be included in the first release.

Question: Does the underlying database consist of five-degree squares or has it been minimised?

Answer: It is dependent on the data. We are starting with time series stats from FAO (Tuna statistic are in five-degree squares), but others may not be. The system will include instruction for how to add database data to it so will again be dependent on the data loaded.

<u>Who Needs Public High-Quality Online Data Anyways? Foundations for a Sustainable Management of the Oceans and Beyond</u>

Prepared and presented by: Falk Huettmann, EWHALE laboratory, Institute of Arctic Biology, USA

Research making use of public online data has been labeled as inferior ‘Internet Science’ removed from fieldwork and true biology. Instead, this presentation reviews many high profile studies that make use of such internet information freely available at the ‘finger tips’ of the global village citizen. Such data are crucially needed for the betterment of ocean management, and the lack, fragmentation, delay, politisation, and strategic disturbance of such approaches (e.g., formats, technology, funding) harms a sustainable ocean management. It is shown that high-quality data dealing with research design, detectabilities and sampling, consistent survey protocols, taxonomies, metadata and fast online publication are crucial for a science-based management to address invasive species, ecological impacts, harvest and climate change for instance. Approaches using raw data, as well as latest predictive modeling in space and time, in order to obtain a powerful synthesis from the various online data sources available, are presented. Finally, a new education system and graduation requirements that embrace concepts like the ones above, showing new ways to a global adaptive management, government, economy and sustainable society are presented.

Questions

Comment: ICES has data policies that allow access to their data. The issue may be that they do not own the data held and so they have to go back to the scientist who owns it for permission.

Response: Geo-referenced data is needed, but we currently have no access to the ICES data.

Comment: In relation to your DIF comparisons to other metadata standards. GCMD gives control for discovery, which makes it easier for different countries. It is really hard to get the data, so this initial step needs to be addressed. Need some kind of system.

Comment: Metadata standard needs to be moved up in priority.

AquaMaps

Prepared by: K. Kaschner, J.S. Ready, E. Agbayani, J. Rius, K. Kesner-Reyes, P.D. Eastwood, A.B. South, S.O. Kullander, T. Rees, C.H. Close, R. Watson, D. Pauly, and R. Froese
Presented by: Jonathan Ready, Swedish Museum of Natural History, Sweden

AquaMaps (http://www.aquamaps.org) are computer-generated predictions of natural occurrence of marine species, based on the environmental tolerance of a given species with respect to depth, salinity, temperature, primary productivity, and its association with sea ice or coastal areas. These 'environmental envelopes' are matched against an authority file which contains respective information for the Oceans of the World. Independent knowledge such as distribution by FAO areas or bounding boxes are used to avoid mapping species in areas that contain suitable habitat, but are not occupied by the species. Maps show the colour-coded likelihood of a species to occur in a half-degree cell, with about 50 km side length near the equator. Experts are able to review, modify and approve maps.

Environmental envelopes are created in part (FAO areas, bounding boxes, depth ranges) from respective information in species databases such as FishBase and, in part, from occurrence records available from OBIS or GBIF.

Questions

Question: In temperate systems, variability is much higher. Since minimum and maximum depths are usually provided, why would we settle for mean salinity and temperature ranges? Why not use actual minimum and maximum values?

Answer: Due to current limitations, we did not additionally design Aquamaps to look at all probabilities at this time. Development of more accurate parameters is intended in future. We can add in more fields for variability (we can add month) and then tweak the system to produce seasonal maps.

Question: It is unclear how the species have been added. Does this rely on users logging in and making queries then making additions, or are they taken from Fishbase? How can parameters be refined?

Answer: Most has been bulk crunching from Fishbase, OBIS, and GBIF. Three hundred or so species have been specifically reviewed from partnerships/projects. Currently, you can download data for analysis, but it has been used more for quality control.

Question: With data integration, there is much detail and data lost. With your modeling, how accurate are these assessments?

Answer: We have been working on this very hard over the last few years. We will have a publication soon.

An ACON-based API for OBIS and its Application to Quantify the Dynamics of Four Scotia-Fundy Groundfish Populations
Prepared by: D. Ricard, B. Branton, and D. Tittensor
Presented by: Daniel Ricard, Dalhousie University, Canada

The dynamics of four groundfish populations in the Scotia-Fundy region are examined using scientific trawl survey data. These stratified random surveys are conducted by DFO and are available on OBIS. The dataset is accessed through an ACON-based application programming interface (API) that facilitates the download of the dataset as a whole. Using strata statistics, temporal changes in the abundance of four groundfish species, Atlantic cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*), pollock (*Pollachius virens*), and silver hake (*Merluccius bilinearis*), are computed using stratified random estimates and are compared to published stock assessment results. The API circumvents the limitations imposed by OBIS for the maximum number of records that can be downloaded, and also provides access to entire datasets without having to navigate through the website's graphical user interface. The stock estimates derived from OBIS data compare favourably with published stock assessment results, indicating the reliability of the OBIS version of this dataset and the great potential of the system as a global repository for survey data following similar sampling design.

Questions

Question: Is it possible to get the ACON system and R code installed on my machine?

Answer: Once this is more mature, we are planning to release it as a package in R. You can go to ACON site to download ACON; however, we are using ACON as it is set up on OBIS. The source code is not available, but binary is. Access to the scripts used on OBIS would be an OBIS decision. Actual C code is owned by the Canadian government.

Question: Have you considered using GBIF's current database web services access?

Answer: OBIS data is harvested, so GBIF has it. GBIF does not have all OBIS schema fields. Will consider looking into this. At present, we are using a local optimised copy of the OBIS cache, so this is not reflective how it would work from directly from OBIS. This should only require a change to where the ACON call is being sent, i.e., BIO changes to iOBIS once iOBIS ACON is updated.

Colombian Caribbean Marine Biodiversity Mapping for Conservation Planning

Prepared and presented by: Martha Patricia Vides Casado, Marine and Coastal Research Institute, Colombia

Perturbations resulting from human activity in marine systems have resulted in habitat change and species lost. As such perturbations may reduce diversity in marine environments, site conservation networks are increasingly seen as refuges for conserving biodiversity. Because all the marine areas contain biodiversity but not all of them can be targeted for action, a selection of sites has to be accomplished. An adequate way of mapping, measuring, and analysing marine biodiversity supporting different stakeholders' priorities and interests identifying potential conservation areas in the Continental Caribbean coast of Colombia is the focus of this research. Due to the impossibility to map biodiversity as such, macrobenthic species assemblages, benthic habitats, and ecoregions are used as surrogates for biodiversity. Three main phases are identified in the research: 1) a data management phase, in which a preparation of museum collection data of marine biodiversity of the Caribbean Colombian slope was made, 2) the biodiversity assessment phase deals with the measurement of biodiversity itself throughout classical community analysis and the mapping of macrobenthic species assemblages predicted distribution, and 3) the final phase, site identification, deals with the selection of representative sites or a sum of planning units accounting for the representation of high biodiversity sites. This selection is based solely on specific biological and environmental features. Cluster and ordination analysis based on presence/absence macrobenthic species showed a clear separation of 9 species assemblages regulated by depth and ecoregions distribution. Three classified data layers became the proxy for conservation features in the site selection: benthic habitats (coral reefs and sea grasses), ecoregions, and 9 macrobenthic species assemblages' distribution. Eleven solutions were produced in the MARXAN software searching for the 2 best comprehensive representations of the conservation features within the available planning units. The best solution over 2 scenarios: the 'No reserve' and 'Seed PU' were selected. For both scenarios, 4895 planning units and 20 conservation features were used. Based solely on biological criterion, the southern region of Islas del Rosario National Park, Santa Marta, and Guajira regions were consistently selected as a representative sites of biodiversity. The study advocate a systematic and objective framework to site selection based solely on biodiversity contents in a regional scheme template.

Panel Discussion

Jonathan Ready was asked if environmental parameters for fish would be usable for invertebrates. John felt that they would be to some extent. Sediment variables would be particularly useful because using 0.5 degree.

Rodolphe Devillers was asked about the sampling background issue, i.e., would using a spatial design mean that randomness is not an issue. Rodolphe indicated that it was not an issue if one kept samples as a whole and not subsets. This was considered to be an important question. Research design can be ignored if the correct statistical methods and right prediction methods are applied. Assessments are done with assessment data, so the results can be used to compare with predictions; thus, research design could be ignored. Dan Ricard felt this would be correct if one were creating a distribution map - then the sampling design could be ignored. To compute time series, two tows that are close together versus two tows that are far apart makes a difference. Samples are not random as needed for overall abundance over time. To take data without knowing the sampling design and just computing the mean abundance over the year is wrong.

The Panel was asked whether they were users of OBIS data, and, if so, how they found the quality of the information contained within OBIS. If they have not used OBIS, would they now that they had heard more about it?

Rodolphe Devillers responded saying that OBIS often has coordinates aggregated, so that only the time/location where a set was pulled up is included, and does not necessarily include the true location of the sample, e.g., only the end point of a half-hour trawl set. Edward Vanden Berghe noted that start and end points are fields in the OBIS data exchange Schema, and could be included in these cases.

Dan Ricard responded by saying that presence/absence information is sometimes appropriate. There are occasionally issues with developing time series, but at the end of the day, the lack of zeroes is simple to address if your model is a relational model rather than a single, flat file. Regionalisation is a good thing, so that each region can see or display their data for discovery purposes. You can always ask for more detailed data from the regional nodes.

Edward Vanden Berghe responded to this saying that deep normalisation, in principle, is possible. It would be possible to describe data by strata and then generate zeroes as ACON is doing. However, there are issues with bandwidth, so there would be a need to have some restrictions. It may be better to enable users to generate zeroes at their end, i.e., get data from OBIS, as well as method/software to generate zeros where appropriate.

Dan Ricard suggested that OBIS is making an environment so one can know that data is available, and then contact the provider for full records via data end use agreement. However, Rodolphe has already explained some difficulties and time delays with doing this. Edward suggested that it would be better to work together to make the full data more easily available. Data providers could coordinate so as to be willing/able to provide their data in a common format.

Rodolphe Devillers indicated that he did not care so much about standardisation when using a single dataset/survey design. Standardisation may be more of an issue if/when merging data from multiple countries. There is core information that is needed, such as documented standardisation, if any. It would also be nice to have accuracy of data indicated in the metadata (i.e., plus minus 1 m, 100 km). Edward noted that there are precisions fields, which data providers are unfortunately not often using.

Someone noted that there had been a lot of discussion about data, but not much about services. It would be useful if OBIS provided web services.

Dan Ricard suggested the use of bug tracking and call numbering on OBIS. For example, when a user reports an error or bug, OBIS could provide a tracking number and keep track of user requests. Edward suggested that this would be very simple to do and offered to look into this as soon as possible.

A conference participant noted that much of the discussion had been about mapping of fish populations. Missing molecular information about how different the populations are from each other (not just diversity, but difference). It was suggested that links should be provided between survey data and genetics data. Techniques are under development, e.g., mapping tree genetics to overlay on distribution so may find/see possible issues. OBIS does not have this capability, but it could create links to systems that contain genetics information.

The Panel was then asked about data coming into OBIS. What do we need to do to convince scientists and others to provide more data?

Martha Vides responded by saying that she was more than happy as a researcher to do so. However, often there are impediments at the institutional level. It should only be matter of time as others continue to attend these meetings and see the potential. Interest will grow.

Different institutions deal with data differently. At BIO, there is a person in charge of managing data. However, academics working by themselves may only have a computer, some data, and no mandate on documentation, maintenance, or release of data.

How do we get everyone to create metadata? Get your peers who have done so in the past to help you. If data is all in Excel, then it takes a lot of mouse clicks to get it into OBIS. Dalhousie University (Canada) has moved to a relational database for this purpose, and metadata is being created. Academics should increase amount of computer knowledge (web, metadata, etc.). At the COML meeting two years ago, people complained about free riders. Scientists' job descriptions should include a clause that says they must provide their data with documentation, including metadata.

Information on species names, distribution, and diversity in Australia has been entered by staff over a period of 30 years. If someone can build a better web service, then Tony Rees would be willing to provide his masterlist. However, unless he has a local copy, he cannot do his work. This list is made available to those who ask, but is not just posted to the web.

John Payne noted that it is a big job to produce metadata. He asked whether there was some method for automatic production. There have been attempts to create metadata generators, but at the end of the day someone has to sit down and write the text, and that someone should know the data. There is semi-automatic software to extract keywords, etc.

Session 6: New Data Systems
Session Chair: Mark J. Costello, OBIS and the University of Auckland, New Zealand

ICoMM, the International Census of Marine Microbes: Unveiling the Ocean's Hidden Majority
Prepared by: L. Amaral-Zettler, M.L. Sogin, J.W. de Leeuw, D.J. Patterson, S. Schouten, L.J. Stal, G.J. Herndl, S. Huse, S. Bordenstein, and P. Neal
Presented by: Linda Amaral-Zettler, Marine Biological Laboratory, Woods Hole, USA

The International Census of Marine Microbes (ICoMM), an ocean realm project of the Census of Marine Life Program, seeks to determine what is known, what is unknown but knowable, and what may never be known about the biodiversity of microorganisms in the world's ocean. ICoMM is a joint venture between The Royal Netherlands Institute for Sea Research and the Marine Biological Laboratory in Woods Hole. The ICoMM Secretariat at Woods Hole hosts the website (http://icomm.mbl.edu) and the distributed database network MICROBIS. It has sponsored meetings for 4 primary working groups (Benthic, Open Ocean and Coastal Systems, Technology, and Informatics and Data Management) and its Scientific Advisory Council. Through a series of workshops that engage the international community of marine microbiologists, ICoMM is forging a large-scale strategic plan to characterise microbial diversity in the sea through molecular approaches (high-throughput ribosomal tag pyrosequencing, genomics, lipidomics, etc.) and to capture of existing legacy data (morphological, etc.), and build a cyberinfrastructure to index and organise the emerging body of information. The community of microbial oceanographers represented within ICoMM (>100 and growing) recognises the enormity of the task at hand, both in terms of total volume of the oceans (estimated to be 1-$4x10^{18}$ m^3) and marine sediments, with a potential population of more than 1030 microbial cells. Although a complete census is most likely beyond grasp, the scientific return will be considerable if the information is integrated with contextual information that can be informative about the interplay between microbial mediated activities and oceanic processes.

Questions

Question: Things have come a long way in biology since the development of operational taxonomic units, where the concept of species does not apply. How does this fit in OBIS?

Answer: There are still ongoing discussions about similarity cut-offs. We want to use DNA proxies for identifying taxa. It is a legitimate concern. Hope to be doing more in future.

Question: How are we changing the environment? You mentioned legacy data; to what extent do you have this?

Answer: The data does not go back very far. There is some sequencing data from 10-20 years ago. We are thinking of going back to archives for legacy samples.

Pacific Coral Reef Data Integration for Ecosystem-Based Management
Prepared by: R. Brainard, M. Parke, R. Moffitt, and M. Moews
Presented by: Michael Parke, NOAA Pacific Islands Fisheries Science Center, USA

Though coral reefs are the most biologically diverse and complex of all marine ecosystems, they are also among the most threatened. Significant declines in key indicators of reef ecosystem health suggest a degradation of coral reefs globally in response to natural and anthropogenic stressors. With predicted increases in coral bleaching and decreased calcification rates due to global warming and ocean acidification, there are concerns that much reef diversity could be lost before it is even documented. As a result, scientists and managers will be left with a limited understanding of undisturbed reef communities on which to base future decisions. To effectively manage these ecosystems, it is critical to both obtain baseline biodiversity data and make existing data more accessible and useful. Since 2000, the NOAA Pacific Islands Fisheries Science Center's Coral Reef Ecosystem Division has conducted long term biennial monitoring of fish, corals, macro-invertebrates, and algae, in the context of their benthic and oceanographic environments at 55 USA-affiliated islands, atolls, and reefs across the Central, Western, and South Pacific. In 2005, the Coral Reef Ecosystem Division partnered with the Census of Marine Life's Census of Coral Reef Ecosystems (CReefs) project to focus more on biodiversity and to expand globally. This ecosystem-based approach will facilitate integrated analyses of biodiversity across diverse gradients of habitats and biological, physical, and chemical environments. The Coral Reef Ecosystem Division has built an internal geodatabase for management of the multiple data types gathered during these efforts and the subsequent generation of synthesised information. The Coral Reef Ecosystem Division and CReefs are working together to improve dissemination and integration of these data with other coral reef data for effective and evolving global marine ecosystem management. With recent funding from NOAA's Pacific Region Integrated Data Enterprise, the Coral Reef Ecosystem Division and CReefs will more effectively distribute biodiversity data via the Pacific Basin Information Node and OBIS, which consist of databases with data search and delivery capabilities that can provide geographically and temporally explicit selections of coral reef biological data in the Pacific region and globally. By combining coral reef biodiversity data sets, these efforts will fill gaps, provide insight into species ranges and connectivity, supplement baseline information with the discovery of new species and records, and serve as a foundation for the ecosystem-based management goal of conserving biodiversity.

Questions

Question: You mentioned bringing down ArcIMS Java Viewer due to limitations. What were the limitations?

Answer: It will not display across 180 degrees and can only provide images. Once user selected the desired image, we would (internally) have to go somewhere else to provide the actual data.

Question: You mentioned coral reef monitoring reports. How often are these reports provided (monthly, etc.)?

Answer: We just finished the first one after six years, which has been presented to American Soma. There are data quality control problems with so many scientists calibrating between different sampling processes and observations.

Question: Would there be one report for each region?

Answer: There would be one for every time we do a cruise. In the future, there is supposed to be one every third year.

The Northeast Bentho-pelagic Observatory to Support Fisheries and Ecosystem Management

Prepared by: S. M. Gallager, L. Mayer, P. Auster, R. Taylor, N. Vine, M. Fogarty, and D. Hart
Presented by: Scott Gallager, Woods Hole Oceanographic Institution, USA

The Northeast Bentho-pelagic Observatory is designed to produce unique data products for fisheries and marine protected area managers and to foster development of Ecosystem Approaches to Management (EAM). Key taxa, benthic community structure, species diversity, seafloor habitat characteristics, and coincident water column properties are being observed and quantified, with repeated measurements in multiple, sentinel sites on time scales of weeks to years. At sentinel sites along the USA northeast coast, that have both high fisheries and conservation value, how communities respond to system change (climate events, fishing activity, position of oceanographic features [fronts], etc.) are also being quantified. This requires fusion of disparate, synoptically acquired data sets, including high-resolution acoustic bathymetry and backscatter (on scales of metres to kilometres), stereo optical imagery (on scales of millimetres to metres), water column plankton distributions (microns to millimetres), and the development of image bioinformatic tools for classifying targets and substrates. Integrated data products are being developed using advanced visualisation tools so key fishery target species and non-target community responses to regulatory practices can be observed and quantified at multiple, relevant space and time scales, in relation to variations in seafloor habitat and boundary layer conditions. Data products will be incorporated into predictive community dynamics models for use in fisheries and sanctuary management. Data products will be of direct utility to fishery and conservation scientists, fishery and sanctuary managers, and environmental policy makers and evaluated for their impact on management practices through socio-economic modeling.

Questions

Question: Besides mosaic image, have you considered using photo telemetry to get size and distance?

Answer: Yes, stereo photo telemetry will be in use on the January cruise. This is useful as bathymetry is changing, and you can use the z axis to help with height.

PLANKTON*NET: A Versatile Online Database System for Marine Biodiversity Information

Prepared by: A. Kraberg, A. Ardelean, A. Macario, B. Onken, D. Vaulot, F. Jouenne, A. Amorim, T. Moita, J. Young, and K.H. Wiltshire
Presented by: Alexandra Kraberg, Alfred Wegener Institute for Polar and Marine Research, Biological Station Helgoland, Germany

Marine biology is changing profoundly. More and more biodiversity related information is becoming available in digital form and the task of disseminating, searching for, and analysing this information is increasingly carried out online.

The online database system PLANKTON*NET, which at present consists of 4 individual nodes (planktonnet.awi.de, planktonnet.sb-roscoff.fr, plankton-net.fc.ul.pt, and plankton.ocean.org.il), is designed to facilitate the quick retrieval of information on a score of marine and freshwater plankton taxa. It currently contains over 5000 images (including 1000 observation records). Originally conceived mainly as an image database, a major redesign of the PLANKTON*NET data model and software in 2007, has produced an extremely flexible database system that can accommodate not only plankton images and glossaries but also taxonomic information (authorities, taxonomic references, in addition to descriptions), biogeographic information, and dynamic links to external resources, all of which can be viewed simultaneously via the taxon sheets generated for each taxon in the database. This information can be searched and browsed using a variety of search functions (e.g., taxon or place name).

In recognition of the vast amount of digitised marine biodiversity information becoming available, data are not only entered into PLANKTON*NET by the core partners, but also by a growing number of external contributors. However, to maintain a degree of quality control PLANKTON*NET operates a rights system, and only users that were assigned contributor status by one of the administrators can enter data. Every time data are entered the relevant administrator is notified and can immediately check the relevant record. Despite the greater scope of the database system, image records are still one of its most important assets. They can be entered as pure image records or as an observation record including biogeographical information. The records in the database are usually entered as part of an image set dealing with a particular topic, e.g., with data from a particular region, a long term dataset, or a collection assembled by a particular researcher. Customised image sets can also be created from existing records using a simple drag and drop facility, for instance, to produce collections of images for teaching purposes.

In addition, a portal has also now been designed which allows the harvesting of metadata from the four existing and any future nodes via the Open Archives Initiative-Protocol for Metadata Harvesting. By combining strictly taxonomic information with information on the distribution of marine plankton a comprehensive system has been produced with a wide range of applications from studying the taxonomy and distribution of harmful algae or invasive species to the production of comprehensive species lists for a number of geographic areas. In this presentation, the latest PLANKTON*NET improvements and tools and their use for marine biodiversity research were presented.

Questions

Question: Do all the records have the same level of details versus discovery level, and how are you dealing with different levels of metadata?

Answer: We are only performing very simple searches on a few fields. It is very simple at the moment.

Question: You stated that you can import any taxon system. If you imported several, how do they correspond or link to each other?

Answer: Users need to specify what they want to use, and they must have the software to use or deal with it.

Caspian Interactive Map Service

Prepared by: I. May, V. Myroshnychenko, and I. Mitrofanov
Presented by: Igor Mitrofanov, United Nations Environment Programme (UNEP) World Conservation Monitoring Centre

The Caspian Interactive Map Service has been developed by the UNEP World Conservation Monitoring Centre, in collaboration with the Caspian Environment Programme and the International Petroleum Industry Environmental Conservation Association, as a response to the need for a tool that will facilitate appropriate environmental planning, as well as emergency response to mitigate the impacts of environmental damage.

The Caspian Sea IMapS is an interactive mapping tool, accessible over the Internet, delivering environmental information in the form that is usable by a variety of different people from field based environmental practitioners to high level policy makers. IMapS is particularly focused on biodiversity issues, combining information on the vulnerable key species, habitats distribution, protected areas, and the ecological conditions and environmental sensitivity of the Caspian coasts. The IMapS allows for an integrated understanding of key environmental information across the Caspian Sea giving users a greater understanding of the information across the region rather than just within individual countries. The Species Query Tool, linked to the Caspian Biodiversity Database, allows users to view a list of individual species that occur in the current map extent.

The IMapS holds information collected from a variety of sources such as local and national governments, non-governmental and international organisations, and private companies on a diverse range of environmental subjects such as hallmark species, habitats, threats, and socio-economic resources. With IMapS, end users can create customised maps online to meet their individual information requirements. Combining biodiversity with environmental and industrial data in a map-based view creates a powerful early warning resource for the decision makers, helping them understand how planned activities and emergencies can impact sensitive ecosystems.

<u>Questions</u>

Question: Do you have a good buy-in from the five countries involved, i.e., are they truly multilateral meetings? Also, who is using the tool?

Answer: Only the environmental component is finished, so it is possible to use but it is unknown if it is being used.

GBIF and Ocean Biodiversity
Prepared and presented by: Éamonn Ó Tuama, GBIF Secretariat

Since its inception in 2001, the Global Biodiversity Information Facility (GBIF) has worked towards its goal of making the world's biodiversity data freely available and universally accessible to all. It achieves this by establishing and promoting an international network of participant nodes, maintaining a central indexing service for data served by those nodes, and building a data portal and various web services as access points to the virtual aggregated data store. Ease of access and interoperability of distributed databases are a fundamental requirement for data sharing and, to this end, GBIF promotes the development of community endorsed standards by working closely with Biodiversity Information Standards/TDWG.

Throughout its first 5-year development phase, GBIF has concentrated on primary species-occurrence data, that is, the specimens housed in museum and other collections around the world, as well as observational records, and through partnership with marine data providers and aggregators, in particular, OBIS, a substantial portion of the 120,000,000+ records now served by the GBIF data portal are marine.

GBIF is now in its second phase of development where the emphasis is on enriching the data portal and the variety of web services it offers, and making it easier for users to find and use data. Tools for visualisation/mapping, quality control, and downloading data in a variety of formats for further modeling/analysis are available, or under development. In addition, the integration of new data types into the portal, e.g., geospatial and environmental data, species checklists, and images is underway.

This presentation provided an overview of the new GBIF data portal, highlighting its contribution to extending the range of information products for marine biodiversity data and, thereby, becoming an essential resource for all users of such data, whether the general public, policy makers, or scientists.

<u>Questions</u>

Question: You mentioned trying to support sustainability. Do you have an example of this?

Answer: We have demo projects, e.g., species response to climate change online system. This queries GBIF, then goes to the Global Earth Observation System of Systems registry to obtain climate data, and then uses prediction models (OPEN modeler), e.g., Alaskan butterfly migration. It needs a better tracking mechanism for users.

Comment: OBIS was leading in marine data three years ago. Since all OBIS data is now going to GBIF, it is unclear that we need all the OBIS nodes. However, OBIS can do certain things that GBIF can not and vice versa. We need to look into which system does what better to reduce workloads. We don’t have the answer to this now, but we need to look into it again.

Response: GBIF can not be the portal for all. It will provide a template portal using universal standards.

Comment: OBIS is planning some developments, but instead it could build on what GBIF already has. For the sustainability of both our organisations, rather than separately investigating what marine data in GBIF is not in OBIS, we can build OBIS as the portal for one stop marine data. There are some fields that would be of interest to GBIF that are/would not be to OBIS. We need to work on overlaps and perhaps should be addressing sponsors and funders together rather than separately.

The Encyclopedia of Life

Prepared and presented by: David Patterson, Marine Biological Laboratory, Woods Hole, USA

The Encyclopedia of Life will build a web site for every species. This very large project is a communal endeavour and is coordinated by a growing collective of cornerstone institutions, currently comprised of Smithsonian Institution, Field Museum, Harvard University, Missouri Botanical Gardens, the Biodiversity Heritage Library, and the Marine Biological Laboratory, Woods Hole. EOL will apply a taxonomically and semantically intelligent informatics infrastructure to link selected aspects of the biology of any organism from any location accessible via the internet. Taxonomic intelligence brings together data about the same organism, even if the information has been labelled with different names. The semantics will allow the content to be used beyond the Encyclopedia of Life context. The project will use Wiki and Web 2.0 style approaches to engage the community, but at the same time will involve experts to ensure that content can be subject to peer review. One component of the infrastructure is a WorkBench, a dynamic yet virtual compilation of open source software modules that can be used to link, organise, annotate, manipulate, analyse, or visualise the available data. Through this WorkBench, users will be able to deposit information for use in Encyclopedia of Life pages, improve and enhance the underlying tools and architecture. Encyclopedia of Life will incorporate the output of the Biodiversity Heritage Library, will seek to integrate with the Barcode of Life, and to work collaboratively with the WoRMS project to promote the emergence of an 'Encyclopedia of Marine Life'.

Questions

Question: What about providing additional pages for many levels?

Answer: The class level Tree of Life project will help classify above genus levels.

Question: Is it possible that as new descriptions come in that they would immediately appear in the Encyclopedia of Life?

Answer: It would be a great idea that as papers were published; they would immediately become available on EOL.

Question: Do you have a handle on the effort required to do this?

Answer: I am going back on Thursday, and we are 48 hours to a week away from initial completion. There are many parts that still need work. We also need to engage the public as it needs to be a community effort.

Session 7: Integrating Different Types of Data

Session Chair: Bob Keeley, DFO National Headquarters, Canada

Large-scale Marine Biodiversity Data Integration in Europe

Prepared and presented by: Ward Appeltans, Flanders Marine Institute, Belgium

MarBEF, a European network on marine biodiversity and ecosystem functioning, has created two unique and large integrated databases: one on soft-bottom macrobenthos (MacroBen database) and one on meiobenthos (Manuela database). An integrated database on hard-bottom macrobenthos and pelagic biota is currently being made. The benefits of these kinds of integrative projects are enormous. The analyses on these data with such a large geographical, temporal, and taxonomical scale will bring new insights in biogeographical patterns, which were impossible to obtain on an individual basis. This presentation showcased how a network of so many different institutes came together and agreed to share their data within a partnership approach. The difficulties dealt with during the data integration were highlighted and the first scientific results from the data analyses were presented.

Integrated Management of Ocean Observing, Biogeographic, and Tracking Data

Prepared by: B. Branton, D. Ricard, and F. Huettmann
Presented by: Bob Branton, DFO Maritimes Region, Canada

In the simplest of terms, OBIS (Zhang and Grassle 2003) consists of a portal at Rutgers University (www.iobis.org) and a diverse network of approximately 25 regional and thematic nodes. The OBIS schema is an extension of Darwin Core standard used by GBIF and others. It has 5 required fields, namely: 1) CollectionCode, 2) CatalogNumber, 3) ScientificName, 4) Latitude, and 5) Longitude. The portal provides text boxes and clickable maps to permit searching of the central database by species name and by location. Results appear as species lists linked to species information pages. They are in turn linked to download and mapping options and to information pages on a variety of other Internet sites. OBIS is effectively a data warehouse system (Inmon 1996) that is evolving in a variety of ways (Figure 6). The DiGIR protocol was originally intended to be used by the nodes to map local databases to the OBIS schema and to respond to query requests from the portal. In practice, node managers generally create SQL 'views' of their local databases, which the OBIS portal manager occasionally crawls and copies into a central database or cache. Similarly, providers were expected to give information about the data to OBIS via the DiGIR configuration files. Since these fields were optional, the information was not generally given. As it turned out, many providers were publishing discovery metadata via other means such as GCMD or with the International Polar Year, so the OBIS portal was modified to instead go directly to GCMD. Coming soon on the OBIS portal are new mapping products such as missing catches, survey effort, and track maps (Black 2007). These maps depend on highly recommended and optional database fields such as: BasisOfRecord, RecordURL, FieldNumber, Locality, SampleSize, RelationshipType, and RelatedCatalogItem. Efforts are now underway to encourage relevant providers to start supplying these important fields. For OBIS, this means employing stricter quality control to ensure that certain key fields are filled depending on the content in BasisOfRecord. Fisheries type data, for example, would be required to include FieldNumber and tracking data to include RelatedCatalogItem. Efforts are also underway to auto-fill missing taxonomic hierarchy information (Kennedy and Bajona 2007), and to use ACON as an application program interface (Ricard and Branton 2007).

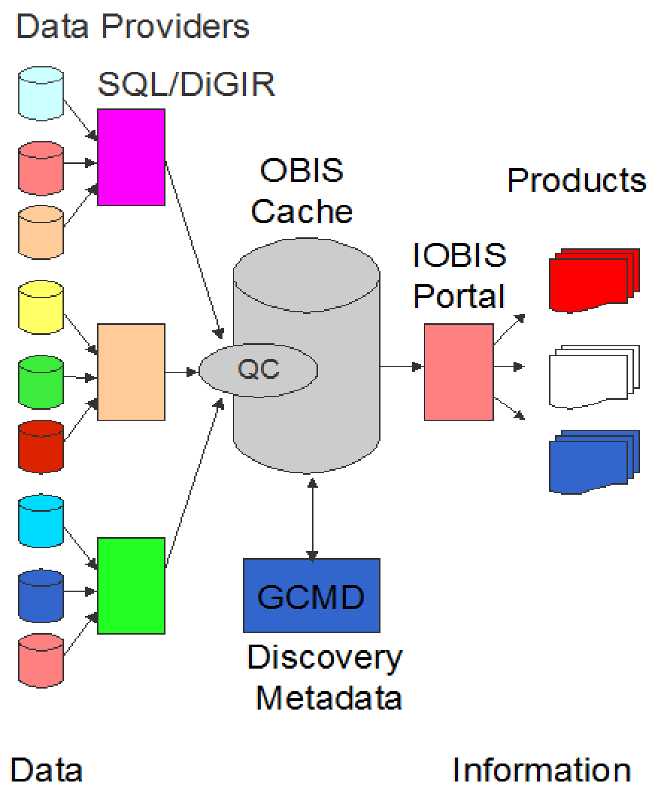


Figure 6. OBIS viewed as a data warehouse system.

The Ocean Tracking Network (OTN) is about the acquisition, management, and integration of 3 kinds of data: 1) acoustic telemetry – POST, 2) archival tags – TOPP, and 3) oceanographic sensor data. A comparison of OBIS and OTN data is given in Table 1. Although OBIS has gone far with a lowest common denominator approach to standards and the GCMD discovery metadata has improved usability of the OBIS portal, it actually has proved very difficult to go back to providers to fill in details to satisfy demands of increasingly sophisticated analytical techniques and users. OTN expects to take a metadata first approach, starting with preparation of detailed Ecological Metadata Language (EML) (Jones et al. 2006) descriptions for its various inputs. OTN data managers are expecting to also use the POST and Electronic Tagging Data Repository (Ancheta 2006) as a basis for OTN's future development.

Table 1. Comparison of the Ocean Biogeographic Information System (OBIS) and the Ocean Tracking Network (OTN).

OBIS	OTN
Global scale.	Global scale.
Many providers.	Many providers.
Many data types.	Three major data types.
Ranging in complexity	All are complex
Often summarised	Never summarised
Small to large in size	All large in size
All data is public	Some data is private
Many taxa	Few taxa

After identifying the OTN user community and conducting a systematic needs assessment survey, OTN will be pursuing a data warehouse style architecture upon which it will iteratively develop open application interfaces and products. From day 1, OTN expects to keep online standards based data and metadata at the very core of the project (Figure 7). Time must be taken at the beginning of the project to learn how to effectively use standards based systems to provide high-quality online data for the global village, and not reinvent any wheels.

Past ... Future ...

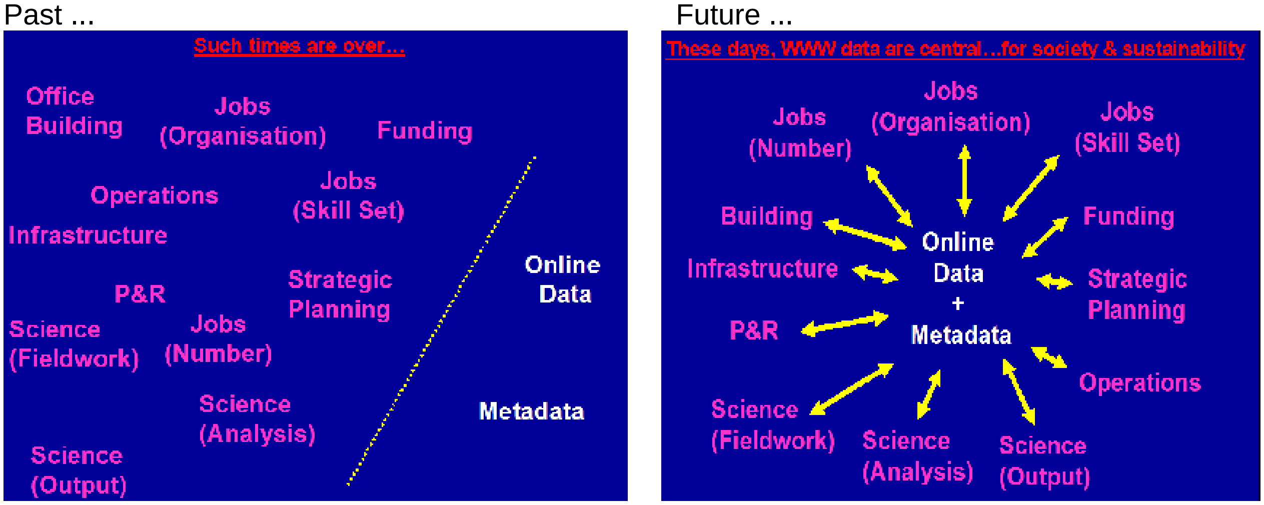


Figure 7. Centralised standards based online data and metadata.

Ancheta, J. 2006. Electronic Tagging Data Repository, Website at University of Hawaii, School of Ocean & Earth Science & Technology (SOEST). https://www.soest.hawaii.edu/tag-data/.

Black, J. 2007. Visualization of Passively Tracked Marine Species using a Web-based Mapper and the OBIS Schema, Oral Presentation: Ocean Biogeographic Informatics Conference 2007. http://marinebiodiversity.ca/OBI07/sessions/visualization-tools/black.

Inmon, W.H. 1996. Building the Data Warehouse, 2nd Edition: Wiley and Son Inc. http://en.wikipedia.org/wiki/Data_warehousing.

Jones, M.B., M.P. Schildhauer, O.J. Reichman, and S. Bowers. 2006. The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere, Annu. Rev. Ecol. Evol. Syst. 2006. 37:519–44. http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.37.091305.110031.

Kennedy, M., and L. Bajona. 2007. Is the Integrated Taxonomic Information System (ITIS) the Comprehensive Solution to Sharing and Linking Marine Biological Data by Organism Name? Poster: Ocean Biogeographic Informatics Conference 2007. http://marinebiodiversity.ca/OBI07/posters/kennedy/.

Ricard, D., and R. Branton. 2007. An ACON-based API for OBIS and its Application to Quantify the Dynamics of Four Scotia-Fundy Groundfish Populations, Oral Presentation: Poster: Ocean Biogeographic Informatics Conference 2007. http://marinebiodiversity.ca/OBI07/sessions/data-use-and-analysis/oral-ricard/?searchterm=ricard.

Zhang, Y., and J.F. Grassle. 2003. A Portal for the Ocean Biogeographic Information System. Oceanologica Acta. 25(5): 193-197. http://en.wikipedia.org/wiki/OBIS.

The Oceanlife System for Integrating Marine Biodiversity Information

Prepared by: K. Stocks, C. Condit, X. Qian, P. Brewin, L. Situ, and A. Gupta
Presented by: Chris Condit, San Diego Supercomputer Center, USA

Understanding and predicting the distribution of genes, individuals, species, and ecosystems through space and time requires the integration of data of multiple types from multiple sources. Point data on observed locations and abundances of species may come from a local survey or from a collection like OBIS. Information on environmental conditions might come from in-situ instruments, satellite imagery, compiled climatologies, or hydrographic models. Taxonomic relationships between species may come from a resource like the Catalogue of Life or from the literature, and existing food web models might provide trophic relationships. Examining complex ecological questions, such as determining the effects of coastal upwelling on zooplankton communities and subsequent fisheries yields, ideally requires the integration of all of these data sources. The more extensive a study becomes, the more the diversity and complexity of the information required increases; and, consequently, the task of putting together the data to enable proper analysis also becomes more complex.

This presentation demonstrated a possible data integration system that copes with data types listed above. A central component of the system is a type-extensible mediator, which was originally designed to integrate data from relational databases, but has been extended to handle spatiotemporal and hierarchical information coming from any registered data source over the internet. The spatial information can be both vector (land boundary, exclusive economic zone) and raster (e.g., satellite data, bathymetry) information. The hierarchical information includes, for example, taxonomy and habitat classifications.

Questions

Question: There seems to be a problem determining warehousing versus data mediation. Do you think mediation is best?

Answer: Not particularly. It depends on usage and updating frequency.

Integrating Biological, Physical, and Chemical Data using a General Data Collection System for Marine Survey Data

Prepared by: K. Deneudt, R. Brackez, and F. Hernandez
Presented by: Klaas Deneudt, Flanders Marine Institute, Belgium

There is an increasing need for measuring biodiversity and quantification of the rate of its decrease. Recently, various international initiatives aim to map marine biodiversity by compiling datasets that reach beyond the scientific objectives of one individual scientist or research group. These compilations rely on data management systems that can store marine biological data in its original context and that can feed their data into the international dataset. Marine biological data are often gathered during ecological surveys and since environmental conditions largely determine the presence of species, the biological data rarely go without detailed measurements of the corresponding abiotic conditions. Storing marine biological data in its original context therefore implies integration of biological, physical, and chemical data.

In conclusion, a general data collection system for marine survey data is sufficiently generic to integrate different data types coming from various sources, takes into account internationally accepted vocabularies, and includes functionalities for transfer to international data compilations. With the Integrated Marine Environmental Readings and Samples (IMERS) database, the Flanders Marine Institute has set up a system that aims to meet these requirements. The use of IMERS in various national and international projects will be demonstrated.

Dynamic Approach to Integrating Oceanographic and Biogeographical Data in the Gulf of Maine

Prepared by: R. Franks Jr., D. Briggs, B. MacLeod, L. Incze, and N. Wolff
Presented by: Richard Franks, Department of Computer Science and the Aquatic Systems Group, University of Southern Maine, USA

Research into the diversity of species that populate the Gulf of Maine and how they live, thrive, and interact is urgently needed in order to maintain, manage, and preserve the Gulf of Maine ecosystem. Working with the Census of Marine Life's Gulf of Maine Area Program, the primary goal has been to develop a general, dynamic system that facilitates the integration of physical and biological data, each having spatial and temporal components. These ongoing efforts will culminate in a computational mechanism allowing researchers to explore relationships between physical and biological data to expose interesting scientific questions concerning these relationships.

The user interface design provides a flexible mechanism for the researcher to specify which of the biological and physical data values are to be joined. The system permits data values over temporal intervals to be aggregated and summarised over a range of temporal granularities and in a variety of ways. It also accommodates the investigation of causal linkages that unfold over time by letting the scientist join data across a time lag. For example, a large spring rainfall may affect the water salinity in April, which, in turn, may affect the distribution of species in the summer. With this tool, the researcher can also compose data values from different attributes according to his/her own functional specifications, and map them into a revealing display, all through a convenient, web-based interface.

The issue of incorporating data from other sources having other formats has been regarded as a critical feature of the system's utility. As a result, an extensible mechanism for allowing different types of sources of physical data to be integrated into the system was developed. The system is based on a collection of drivers, all implementing the same data translator interface. This allows multiple sources of physical data to be integrated into the system, including simulation model output data and DODS data.

Questions

Question: Are all data served as layers?

Answer: No, they are generated on the fly based on a query. A lot of people are using GIS, so we felt that many users would be comfortable with it.

Integrated Approaches to Acquisition, Analysis, and Discovery of Video Imagery from Marine Benthic Habitats
Prepared by: P. Lawton, M. Strong, J. Sameoto, and R. Benjamin
Presented by: Peter Lawton, DFO Maritimes Region, Canada

The use of georeferenced digital video imagery for analysis of marine benthic habitat physical structure (both biogenic and abiogenic elements), species occurrences, and biological community attributes, is expanding rapidly. In this presentation, examples were used from recent studies engaged in within the Gulf of Maine Biodiversity Discovery Corridor, to describe:

1) Real-time georeferenced data capture during benthic video deployment,
2) Development of relational database models for benthic habitat classification, and
3) Use of video processing software and output of analysed video to geographic information systems.

Questions

Question: What scale is the classification being done at?

Answer: This is a meso-scale survey classification. Others are doing frame by frame analysis so that individual records can later be associated to the classification.

On the Importance of Managing Biomolecular Sample Information Alongside other Environmental Measurements
Prepared and presented by: Gwenaelle Moncoiffe, British Oceanographic Data Centre, UK

The rapid growth in environmental genomics and biodiversity studies has resulted in an increasing number of samples being collected at sea for genetic sequencing and other types of molecular analyses. These analyses generally lead to large and complex datasets, which are not easily managed alongside more traditional oceanographic and other environmental measurements. As a result, molecular data are often managed separately, and not always in a way that allows investigators to easily retrieve associated environmental parameters.

Both the environmental and the molecular biology communities have seen very rapid progress in the development of data management systems and web-based tools, services, and networking initiatives. However, with regards to basic data management and data discovery needs, the two communities have yet to share common practices, learn from each other's experience, and collaborate more effectively.

While the kind of bioinformatics tools needed to support environmental genomic research can be extremely complex and beyond the range of expertise of a traditional environmental data centre, there are a number of basic steps that could be followed to ensure that molecular data do not become detached from other data collected at sea.

This presentation reviewed these steps and discussed areas of data management practices where greater collaboration would be required in order to make the integration of molecular and oceanographic data possible.

Questions

Someone commented that an important database management step is required to develop the initial storage in which to create cruise reports. If we are starting to link the DNA database to cruise reports, there needs to be a distinction between the environmental data collected at the same time versus other data from the same place and time, but that is collected at a different resolution.

Question: Which method did you use for acquiring environmental data?

Answer: We are going back for molecular data and asking for missing environment information.

Towards Global Data Synthesis: A Predictive Model for the Circumpolar Arctic using High-Quality Public Online 'Presence Only' Data Sources for Zooplankton
Prepared by: I. Rutzen, R. Hopcroft, and F. Huettmann
Presented by: Imme Rutzen, Institute of Wildlife Research, Hannover, Germany

Copepods are the most dominant contributors to the biomass of the Arctic and Antarctic oceans: they are of primary importance in marine ecosystems. Feeding on phytoplankton or other zooplankton, zooplankton presents a direct prey-link with fish (e.g., herring), birds (e.g., little auks), as well as mammals (e.g., bowhead whales). Efforts are being made to generate a first digital distribution atlas, covering 4 copepod species (*Calanus hyperboreus, Calanus glacialis, Metridia longa,* and *Metridia pacifica*) in the Arctic Ocean as example species. Public online databases such as the CoML's Arctic Ocean Diversity project (ArcOD) and National Marine Fisheries Service data were used and extended with recently collected unpublished datasets. A predictive GIS model was designed using ecological niche modeling with ArcGIS 9.2. Environmental data (e.g., water temperature layers, salinity, nutrients, bathymetry, ice cover, distance to coastline, settlements, and hydrology) was taken from the World Ocean Database (http://www.nodc.noaa.gov/OC5/WOD05/pr_wod05), US National Snow and Ice Data Center (http://www.nsidc.org), and US National Oceanic and Atmospheric Administration (http://www.noaa.gov). This model elaborates on a future global warming scenario, allowing prediction of how a warming environment might impact the distribution of the 4 study species. In order to derive the best possible distribution maps for the chosen species, the latest statistical modeling algorithms, such as TreeNet and RandomForest, were used. Predictions are assessed for quality performance using alternative datasets and related error metrics. This project serves as an example for determining how useful current public online datasets are, and for illustrating their advantages, as well as their failings (e.g., when Metadata are missing or of poor quality).

TDWG's Social and Technical Architectures for Interoperability in Biodiversity Informatics

Prepared by: L. Belbin, S.D. Blum, D. Hobern, R. Hyam, and R.S. Pereira
Presented by: Éamonn Ó Tuama, GBIF Secretariat

With support from the Gordon and Betty Moore Foundation and the Global Biodiversity Information Facility (GBIF), the Taxonomic Databases Working Group (TDWG) has recently completed an overhaul of its infrastructure for developing standards and promoting interoperability across the biodiversity information domain. The project has achieved 5 key improvements: 1) the modernisation of TDWG's standards development process, 2) the establishment of modern technologies for remote collaboration, 3) the adoption of a framework for persistent and globally unique identifiers, 4) a protocol for retrieving structured data from distributed information resources, and 5) an ontology that will enable smaller communities of practice to develop conceptual data models that serve their needs, while maintaining an ability to share data across the entire scope of biodiversity science.

The new standards development process uses interest groups to explore solutions to an information sharing problem and task groups to create particular work products on a schedule. Every interest group and task group must publish and maintain a 'charter' that describes the group's scope and purpose, identifies its communication forums and key personnel, and provides a timeline for events and milestones. TDWG's Executive Committee and Technical Architecture Group review every charter to ensure technical and political harmony across subgroups and their activities. Once a task group has produced a standard specification, the procedure for reviewing and ratifying a standard includes both expert review and a thirty day period for open comment. All new TDWG standards will be archived in an open repository.

The new TDWG communications infrastructure includes a web site supported by a content management system and a Wiki system that enables subgroups to develop draft documents collaboratively and informally. Traditional mailing lists enable subscribers to monitor and participate in broader discussions, which naturally have a more sporadic pace of interchange. Finally, an electronic journal system supports a formal review process and is being used for both traditional publishing, as well as a standards development process. A single sign-on infrastructure supports most of these facilities and now has more than 650 registered users.

Persistent and Globally Unique Identifiers (GUID) are critical for creating information interoperability because they enable reference of particular items within authoritative information resources. This promotes consistency and enables the partition of responsibility for information across a larger domain. After 2 workshops and 18 months of discussion, TDWG has adopted Life Sciences Identifiers (LSID) as the technology for globally unique identifiers wherever another GUID system has not been established already, such as Digital Object Identifiers in the publishing industry.

The TDWG Access Protocol for Information Retrieval (TAPIR) provides a mechanism for interrogating information repositories in the biodiversity domain and replaces the DiGIR and BioCASe protocols. TAPIR providers have been developed in Python, PHP Hypertext Preprocessor, and Microsoft.NET; and a Java version should be available within the year.

Finally, the TDWG ontology identifies the upper level information concepts that expect to be shared across the biodiversity domain. The ontology is expressed in Resource Description Framework in anticipation of the emergence of semantic web technologies.

Panel Discussion

OBIS was asked whether they had any plans to allow providers to provide data not via DiGIR but with web feature services (WFS).

Bob Branton indicated that we have a little experience here at BIO with WFS. For example, in order to combine the Maritimes Regional Cetacean Sightings Database with current information, the data were mapped to the OBIS schema and converted to a WFS. There is no way to easily select the data, though a solution has been found. OBIS Canada wants to provide all collections through WFS as an experiment. Hopefully, these efforts would be considered by OBIS.

Edward Vanden Berghe mentioned that OBIS is a little unhappy with DiGIR right now. At the time, it was ‘state of the art’. WFS were not well known then, but he thinks that we should move forward to WFS and then bring in physical oceanography. He just did not know when. There is/was a Sloan funded pilot project to put WFS in front of OBIS data, but he did not know how well it will work for providing the data.

A question was asked regarding duplication of data appearing in many data sets. POST would like to provide data at different resolutions, and this may create duplicates. How is OBIS going to deal with this? In OBIS, work is taking place with Melanie Meaux and Phoebe Zhang to work out discovery metadata. The metadata contains much more information than the OBIS records. The OBIS view of data is a product, which is a child record from a single parent. POST would have single a parent metadata record with multiple child records.

At the 2004 meeting in Hamburg, a declaration was made on the importance of free and open data sharing. Open source software can be very helpful in achieving this. Do you think that proprietary software is an issue, and are there any constraints in using open source software (as proprietary software companies would like us to believe)?

In an academic environment, low cost solutions are often a necessity. Open source products can be more flexible and thus more effective than proprietary software, such as Oracle™ and ESRI, and many academics are making a conscience choice to start using open source products. From a developers point of view, open source is much better. Also, it is not limited to what the software comes with. In open source, you can make changes or add tasks as needed. There are no risks in combining open source and proprietary information, as long as specifications fit all needs and requirements. For desktop applications, proprietary systems can be better, but for web-based applications, open source can be better.

In Central America, if you are cash strapped and choosing between contracting and/or training personnel, they train the personnel. Two years ago, OBIS asked Pat Halpin to look into this question and his conclusion was similar, i.e., for the desk top, proprietary software is better, but for internet/web development, open source is better. Many others have found the same. With desktop applications, there are many institutions with long histories of using the proprietary systems, so they will keep these for easy communication/sharing.

The Panel was asked to reflect on the use of metadata, including issues with how data are being used. Many data providers express concern that their data are being used incorrectly. We say it will be misused if you do not provide the information in the metadata. It is very difficult to develop an all-in-one system. It is better to link to existing data and add needed metadata.

In a different session, it was mentioned that data should include a description of its quality (resolution) so other future users can determine whether/how to use. They should have fields at the record level that will show quality and/or values to be used in analysis.

It was agreed that both metadata and individual record fields are needed, so the persons responsible for the data resolution reduction are provided the values.

The Panel felt that this was something that needs to be dealt with and that is very specific for biological information (physical data is generally not so complex). We need to know who identified a specimen and how they did so (book, page), but it is going to be very difficult to include all this information in each record.

In speaking about including instruments, vessel, etc., the information is there (in cruise reports), but it is not linked, so when the scientists go to analyze the data, it is lost. We need to set certain procedures to ensure these details stay with the data.

Data providers are concerned. Instead of talking about data distribution and sharing, we should call it submitting paper/journal/publication. All these are covered in the procedures used for publishing papers. This way, data providers would get credit in a way that can be used in a resume.

SYNTHESIS AND NEXT STEPS

Synthesis

A synthesis of the conference was provided by Marc Taconet of the FAO Fisheries and Aquaculture Department (Figure 8).

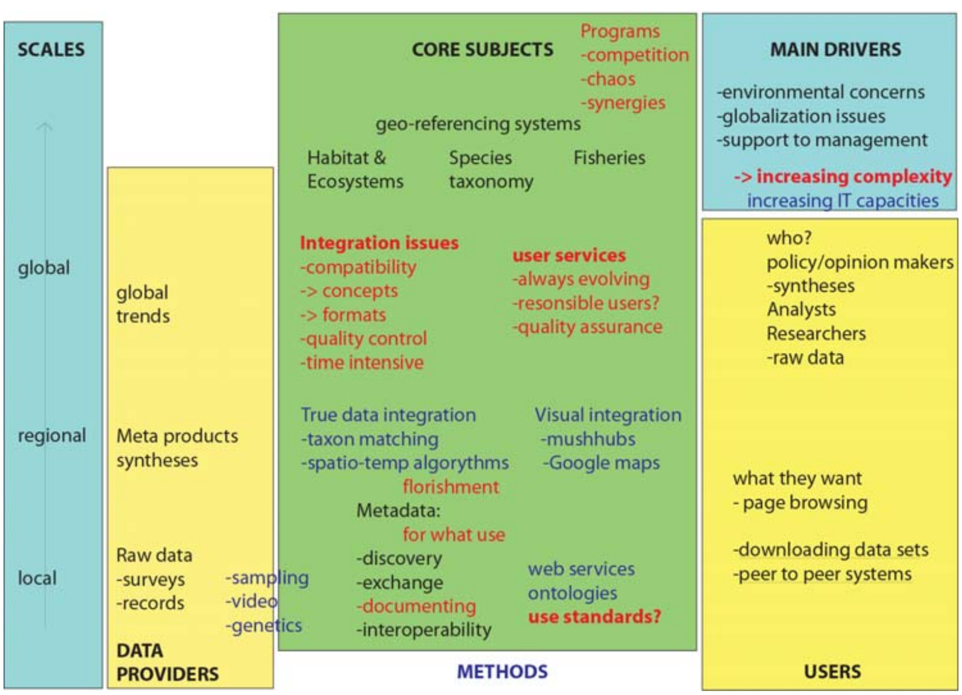


Figure 8. Graph synthesising the various topics of discussion at the OBI '07 conference.

Next Steps

A discussion on next steps was led by Dr. Edward Vanden Berghe and Dr. Mark Costello.

Marc Taconet's diagram is a good representation of what the meeting was about. There are disparate datasets and a number of projects going on. Those that started to look at how the data was being used in management (GeoCOD, pac science center) are particularly interesting. They show how science is being used within management. It would be nice to have some statistics on who is using the data and to what use -- not just the number of hits, but what types of data are being used and what types of management decisions are being made based on these data extractions.

OBIS does have some statistics, but it is not known yet what impact the data use is having on decision making. OBIS did do a user needs assessment (last spring) on who is using OBIS and why (English only). The next phase of this is to release the survey in multiple languages.

There are still issues around the ethics of data mining and how to recognise people who contribute raw data that they worked so hard to collect. Some of this data is being published responsibly but some is not.

Data providers are the keystone that will build the evolving system. People are using their own money to get their data together to post to OBIS. There was interest in having OBI put out an idea similar to that generated at Hamburg regarding this paradigm shift. Scientists need to get credit for the information that they provide. It used to be papers that people published; now scientists have to be credited for providing data. A numbering system could be developed so that a number is provided to put into a resume that will show who and how many times the data have been used. This requires a f paradigm shift, i.e., in doing scientific work.

Another example is the Argo program where participants must make their data available immediately. Argo has 3000 floats in the water, and 85 – 90% of the data is available to the public within 48 hours. Even though they spent their money, they gained by having access to the entire data set, which they would not have been able to do on their own.

We all know the process for paper publications, and we need a similar process for data publications. This does not mean much if data is rubbish. It is only valuable if data is being referenced in publications. Following Genbank, you can only publish a sequence if provided on internet access, then it becomes easy to see how often sequence published in other papers.

Nature and Science have requirement that all data must be provided. Taxonomists also have requirements.

We need similar process/system for data processing. OBIS includes citation information for each dataset, and this should be used. We are actively working on updating this with complete information.

As an OBIS provider, my citation is going to be my database reference.

We need to involve publishers and reviews in this.

How can we ensure that all data providers receive their due when their data are used? What about citations for the maps from obis? Any reference for these and data combined products?

OBIS-SEAMAP did send out requests and tried to involve editors, but there was not a full response. Bob Branton is trying to come up with proper dataset citations. This is a first step. Eventually we hope to see thousands of citations.

What about large aggregations (climate, biological, etc.), which could involve thousands of records? This would be very hard to cite; perhaps citing the integrator/system rather than trying to cite hundred/thousands of records. However, if there is some real dataset that caused you to come to a conclusions/recommendation, then that data provider should be cited.

Are there fora to make better synergy for the taxon item/standard, single managing point? We all hope that the Catalogue of Life will very soon be the top authority for/to taxon. GBIF sees it the same. Different groups should specialise in their areas and share. There is never going to be one place that does it all. We should work on the infrastructure so that sharing can be done.

The main difference in the content of this conference from the previous one in Hamburg are the discussions on use of web feature services, semantic ontologies, and, to a lesser extent, mentions of EML. Tools such as Google Earth (©2007 Google) did not exist then.

Tracking data is needed in OBIS, but there are many issues that need to be addressed beforehand, i.e., how to keep track of the different datasets being used in aggregations, names issues, differing methods, etc. We are facing more complex data with requirements to include information on the technology used, the quality of the instruments, software developments, and tracking the algorithms for working out things like longitude, which is being tweaked all the time. Something is needed that shows the algorithm used at the record level. We should generate new versions of the datasets as new algorithms are used.

The OGC working group is developing digital writing standards, and we should look at what they are doing. OGC web feature services need a standard under it, as sharing is not automatic. They all have an underlying standard. Need to state what common standard we are going to use under web feature services (GBIF de facto standard? as already have WFS), geography mark-up language. We know that CHS is doing similar stuff (many of us using OGC standard, we do not have a standard and are still deciding on metadata standard). None of us have the time/resources to bring in everyone when doing stuff.

Next Meeting

Would any group like to consider hosting next OBI, perhaps in 2010?

Suggestions included the next IODE meeting, the date of which had not yet been determined, though it was likely to be in November 2009. Alternatively, it could be done in conjunction with the next group of experts meeting, perhaps as separate/side section during March-April 2009.

Other possibilities include: The International Conference on Marine Data and Information Systems (IMDIS) 2008 in Athens during March 31-April 2, 2008.

April-May 2008 will be the OTN Launch. They are working toward an OBI structure in OTN. A Global Data Management meeting is supposed to be held in October 2008.

Administration

Questions: Do we need a new mechanism for continuing this discussion between meetings? Would it be useful to create an ad-hoc organising group or email list? Should we have a web site?

Suggestions included an online content Plone site, which could be used for online group document, scheduling meetings, and posting materials for others to use. Peter Lawton of the Canadian Centre for Marine Biodiversity mentioned that the Canadian Centre for Marine Biodiversity already has a committee and is using a Plone site. The Centre for Marine Biodiversity could conduct a survey to ask if people think Plone site usage a good idea and how people feel about maintaining this site into the future. A national biodiversity group is forming, so it may be able to maintain the OBI ’07 site.

Another suggestion was a bulletin board for posting questions and making conference announcements. The join capability has been removed from the OBI ’07 website, so it not there right now, but it could be added to the Centre for Marine Biodiversity website.

Is there interest in slightly more than an email list to maintain the energy from this conference? Would OBI be interested in having statements coming out from editorial board along with suggested bulletin board (vision statement, publications, upcoming meetings, etc.)? It was suggested that most people would not use the site unless they received an email saying to go look at the site. This suggestion could be added to the survey.

Closing Remarks

Closing remarks were provided by Fred Grassle, Director of the Institute of Coastal and Marine Sciences, Rutgers University, State University of New Jersey. He commented that OBI was recognised as a new science at the last meeting. He stressed that the really important thing about OBI is the community, especially in the area of software development. The marine diversity policy side has always been a problem of ‘out of sight out of mind,’ but The Census of Marine Life is starting to change the way people think about the oceans. “We are starting to reverse the idea that the ocean is ‘the deep unknown desert.’ The figure that Marc Taconet used to synthesise the meeting was quite helpful and is much appreciated. Thank-you all!”

POSTER SESSION

A poster session was held on the second day of the conference (3 October 2007). A list of the poster titles is provided below.

biodivDP - A Biodiversity Information System with Applicability in Ocean Science
A. Ardelean

Is the Integrated Taxonomic Information System (ITIS) the Comprehensive Solution to Sharing and Linking Biological Data by Organism Name?
L. Bajona and M. Kennedy

Ocean Biogeographic Information System Metadata Automation Framework
R. Branton, P. Zhang, M. Meaux, L. Bajona, and D. Broughton

Distribution of Meiobenthos with Special Reference to Nematode Assemblages of Continental Slope Sediments, Northeast Coast of India
C. Annapurna and L.M.G. Cooper

Testing Hypotheses of Degradation in Mediterranean Sea Food Webs using Stochastic and Mass-balance Modeling
M. Coll, H. Lotze, and T. Romanuk

SCAR-MarBIN: Disseminating Antarctic Marine Biodiversity Information
B. Danis

Planning of Marine Tourism Development at Teluk Lada Area, Banten with Geographic Information System Approach.
M.J. Elly

AlgaeBase - A Comprehensive Information Resource for the World of Algae
M.D. Guiry, D.J. Garbary, and F. Rindi

Integration of the Colombian Marine Biodiversity Information System (Sibm) to OBIS
A. Gracia, G.R. Navas, N. Santodomingo, J. Bohórquez, M. Díaz-Ruíz, L.M. Mejía-Ladino, and E. Montoya-Cadavid

The Design and Implementation of the CMarZ Database and Species Pages
R.C. Groman, N. Copley, A. Bucklin, and P.H. Wiebe

The Dynamic Atlas of the Gulf of Maine
J. Hodsdon, B. MacLeod, D. Briggs, N. Wolff, and L. Incze

Data Needs for Modeling the Maintenance of Ecosystems Services on Small Coastal Banks
L. Incze, P. Auster, and N. Wolff

Data Management in DFO Science
R. Keeley and R. Eisner

Fisheries and Aquatic Resources Node NBII
M. Kelly, B. Haupt, A. Ostroff, R. Baxter, and J. Spayd

A Federated Marine Biodiversity Information System: A Semantics-based Approach for Interoperability and Analysis
Z. Kemp

An OAI Framework for Biodiversity and Contextual Content: PlanktonNet as Pilot Study
A. Macario and B. Onken

The New European On-line Journal 'Aquatic Invasions': Services for Marine Biodiversity Related Information Systems
V. Panov and S. Gollasch

ChEssBase - A Database for Deep-water Chemosynthetic Species
D. Perry, M. Baker, and E. Ramirez-Llodra

History of Marine Animal Populations
P. Holm, A. Rosenberg, B. MacKenzie, B. Poulsen, and A.H. Marboe

Data Holdings, Global Analyses, and Future Plans for the Future of Animal Populations (FMAP) Project
D. Ricard, Z. Zahorodny, H.K. Lotze, I. Jonsen, and B. Worm

Changing the Focus: Improvements in Species Identification
K. Rutherford

Official and Large Collections - The Constraints to Make Known Unpublished Information on Biodiversity
F. Silveira

A Fuzzy Logic Expert System to Estimate Intrinsic Vulnerabilities of Marine Fishes to Fishing
W.L. Cheung, T.J. Pitcher, and D. Pauly

Developing Integrated Geospatial Tools for Coastal and Ocean Managers in the Transboundary Gulf of Maine Region
C.M. Tilburg, D. Keeley, S. Russell-Robinson, T. Shyka, and T. Gale

Marine Data Archive Centres in the UK – Unlocking Data Resources Through Partnership
J. Parr, H. Tyler-Walters, D. Lear, and R. Seeley

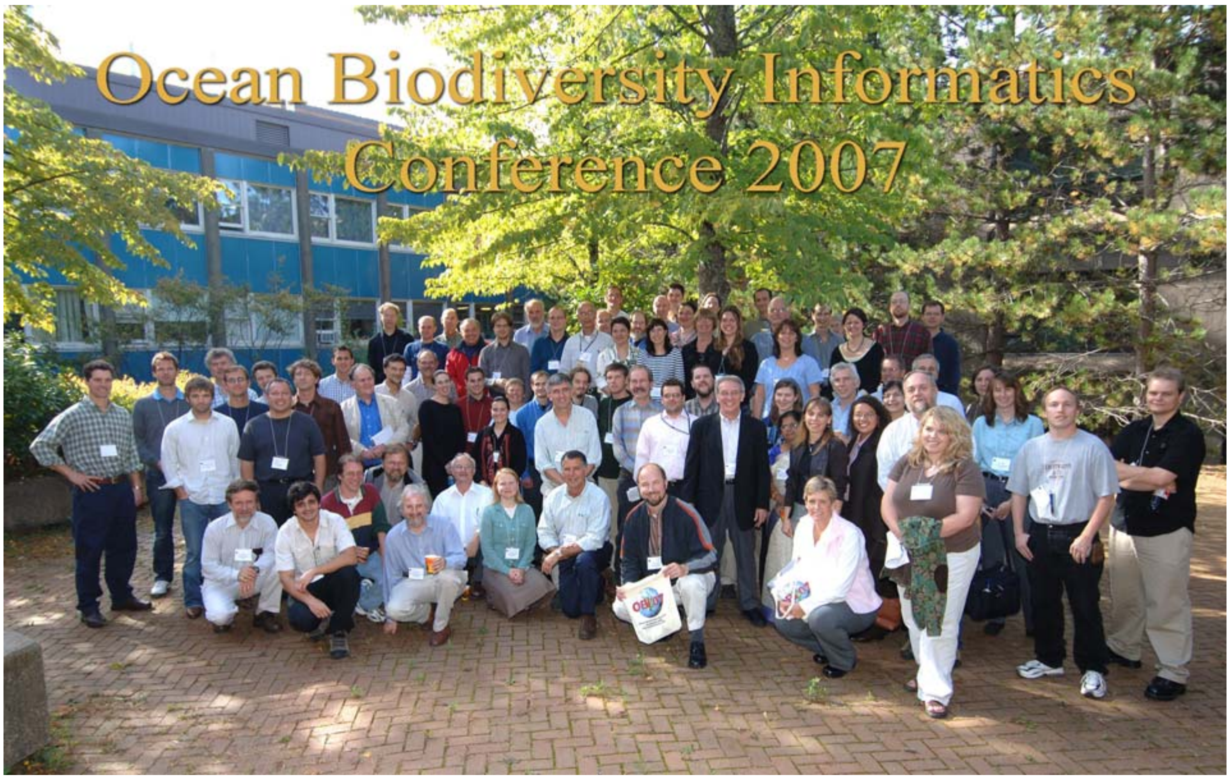
Oceanographic Data Mapper, 'A GIS for the Rest of Us'
P. Upson, D. Gregory, and R. Branton

The Atlantic Reference Centre – A Microcosm of Ocean Biodiversity Informatics Standards
L. Van Guelpen

Development of an Optical Image Database and Processing Tools for Automated Classification of Benthic Habitat and Enumeration OF Scallop Stocks
A. York, R. Taylor, N. Vine, S. Lerner, D. Hart, L. Prashad, and S. Gallager

ACKNOWLEDGEMENTS

The organising committee, comprised of Bob Branton, Victoria Clayton, Richard Eisner, Dan Ricard, Lenore Bajona, and Tana Worcester, thank the presenters and participants for their time and effort. There were also many others who were essential for the execution of the workshop and are also thanked. Thank-you also to DFO Maritimes Science Branch and the Bedford Institute of Oceanography.



APPENDICES

Appendix 1. Table of Abbreviations

ABCD – Access to Biological Collection Data
ACON – A CONtouring application
AJAX – Asynchronous JavaScript and XML
ARC – Atlantic Reference Centre
ArcOD – Arctic Ocean Diversity
BIO – Bedford Institute of Oceanography
CAML – Census of Antarctic Marine Life
CeDAMar – Census of Diversity of Abyssal Marine Life
CenSeam – Census of Seamounts
CGDI – Canadian Geospatial Data Infrastructure
ChEss – Chemosynthetic Ecosystems
CHS – Canadian Hydrographic Service
CMarZ – Census of Marine Zooplankton
CMB – Centre for Marine Biodiversity
CoMargE – Continental Margins Ecosystems
CoML – Census of Marine Life
CReefs – Coral Reef Ecosystems
CSIRO – Commonwealth Scientific and Industrial Research Organisation
DFO – Department of Fisheries and Oceans
DIF – Directory Interchange Format
DiGIR – Distributed Generic Information Retrieval
DRDC – Defence Research and Development Canada
EKME – Electronic Knowledge Management Environment
EML – Ecological Metadata Language
ERMS – European Register of Marine Species
FAO – Food and Agriculture Organisation
FGDC – Federal Geographic Data Committee
FIGIS – Fisheries Global Information System
FIRMS – Fishery Resources Monitoring System
FMAP – Future of Marine Animal Populations
FSDAS – Fish Stock Depletion Assessment System
GBIF – Global Biodiversity Information Facility
GCMD – Global Change Master Directory
GIS – Geographic Information System
GMBIS – Gulf of Maine Biogeographic Information System
GOMA – Gulf of Maine Area Census
GoMODP – Gulf of Maine Ocean Data Partnership
GoMOOS – Gulf of Maine Ocean Observing System
GUID – Globally Unique Identifiers
HMAP – History of Marine Animal Populations
ICES – International Council for the Exploration of the Sea
ICoMM – International Census of Marine Microbes
IMERS – Integrated Marine Environmental Readings and Samples
IODE – International Oceanographic Data and Information Exchange
IRMNG – Interim Register of Marine and Nonmarine Genera
ISO – International Organization for Standardization
ITIS – Integrated Taxonomic Information System

KML – Keyhole Markup Language
LSID – Life Sciences Identifiers
MAR-ECO – Mid-Atlantic Ridge Ecosystems
NAFO – Northwest Atlantic Fisheries Organization
NaGISA – Natural Geography in Shore Areas
NASA – National Aeronautics and Space Administration
NEAFC – Northeast Atlantic Fisheries Commission
NOAA – National Oceanic and Atmospheric Administration
NRC – Natural Resources Canada
OBI – Ocean Biodiversity Informatics
OBIS – Ocean Biogeographic Information System
OBIS-SEAMAP - OBIS-Spatial Ecological Analysis of Megavertebrate Populations
OGC – Open Geospatial Consortium
OPeNDAP – Open-source Project for a Network Data Access Protocol
OTN – Ocean Tracking Network
OWL – Ontology Web Language
POST – Pacific Ocean Shelf Tracking
ROPOS – Remotely Operated Platform for Ocean Science
SCIDAT – Science Dataset Inventory
SDI – Spatial Data Infrastructure
SDSC – San Diego Supercomputer Center
SOA – Service-Oriented Architecture
SOAP – Simple Open Access Protocol
SQL – Structured Query Language
TAPIR – TDWG Access Protocol for Information Retrieval
TDWG – Taxonomic Databases Working Group
TOPP – Tagging of Pacific Predators
UNEP – United Nations Environment Programme
VLIZ – Flanders Marine Institute
WMS – Web Map Service
WoRMS – World Register of Marine Species
XML - Extensible Markup Language

Appendix 2. List of Participants

NAME	AFFILIATIONS	EMAIL ADDRESS
Adamek, Adrienne	University of Southern Maine	aadamek@usm.maine.edu
Allison, Dicky	Woods Hole Oceanographic Institution	dallison@whoi.edu
Amaral Zettler, Linda	Marine Biological Laboratory	amaral@mbl.edu
Appeltans, Ward	Flanders Marine Institute (VLIZ)	ward.appeltans@vliz.be
Ardelean, Adorian	AquaEcology	ardelean@aquqecology.de
Baidya, Haragobinda	Minority Self Empowerment Foundation (MSEF)	msef_taposh@yahoo.com
Bajona, Lenore	Fisheries and Oceans Canada	bajonal@mar.dfo-mpo.gc.ca
Bassirou, Ndiaye	Centre for Marine Biodiversity (CMB)	bass.ndiaye@laposte.net
Black, Jerry	Fisheries and Oceans Canada	blackj@mar.dfo-mpo.gc.ca
Bond, Shelley	Fisheries and Oceans Canada	bonds@mar.dfo-mpo.gc.ca
Branton, Robert	Fisheries and Oceans Canada	bob.branton@gmail.com
Briggs, David	University of Southern Maine(Computer Science Department)	briggs@usm.maine.edu
Broughton, Derek	DFO contractor	derek@pointerstop.ca
Butt, Kenneth	Lotek Wireless Inc	kbutt@lotek.com
Chandrabhotla, Annapurna	Registered presenter , Andhra University, India	cannapurna@rediffmail.com
Clement, Pierre	Fisheries and Oceans Canada	clementp@mar.dfo-mpo.gc.ca
Coll Monton, Marta	Dalhousie University, Biology	martacoll@dal.ca
Condit, Christopher	San Diego Supercomputer Center	condit@sdsc.edu
Cooper, J. Andrew	Fisheries and Oceans Canada	coopera@mar.dfo-mpo.gc.ca
Costello, Mark	OBIS and University of Auckland	m.costello@auckland.ac.nz
Currier, Paul	New Hampshire Dept. of Env. Services	pcurrier@nh.gov
Danis, Bruno	Royal Belgian Institute of Natural Sciences	Bruno.danis@scarmarbin.be
Deneudt, Klaas	Flanders Marine Institute (VLIZ)	klaas.deneudt@vliz.be
Devillers, Rodolphe	Dept. of Geography, Memorial University of Newfoundland	rdeville@mun.ca
Eisner, Richard	Fisheries and Oceans Canada	eisnerr@mar-dfo-mpo.gc.ca
Elly, Muhamad Jafar	Bedford Institute of Oceanography (BIO)	bungjeff2003@yahoo.com
Fatoki, Adeolu		Olasunkanmi_Fatoki@yahoo.com
Franks, Richard	University of Southern Maine	richardfranksjr@gmail.com
Gallager, Scott	Woods Hole Oceanographic	sgallager@whoi.edu
Garbary, David	St. Francis Xavier University	dgarbary@gmail.com
González Taboada, Fernando	Área de Ecología, Universidad de Oviedo	fgtaboada@gmail.com

NAME	AFFILIATIONS	EMAIL ADDRESS
Gracia, Adriana	Museo De Historia Natural Marina De ColomBIA – INVEMAR	agracia@invemar.org.co
Grassle, J. Fred	IMCS, Rutgers University	grassle@marine.rutgers.edu
Groman, Bob	Woods Hole Oceanographic Institution	rgroman@whoi.edu
Güntsch, Anton	Dept. of Biodiversity Informatics, Botanic Garden & Botanical Museum Berlin-Dahlem	a.guentsch@bgbm.org
Hackett, Jennifer	Fisheries and Oceans Canada	hackettj@mar.dfo-mo.gc.ca
Halpin, Patrick	Marine Geospatial Ecology Lab, Duke University	phalpin@duke.edu
Haupt, Bernd	Penn State University	bjhaupt@psu.edu
Hodsdon, Joseph	University of Southern Maine	joe.hodsdon@gmail.com
Holborow, James	Department of Conservation, New Zealand	jholborow@doc.govt.nz
Holdsworth, Neil	ICES	neilh@ices.dk
Huettmann, Falk	ArcOD, EWHALE lab Uni of Alaska-Fairbanks	fffh@uaf.edu
Hum, Flo	Fisheries and Oceans Canada	humf@mar.dfo-mpo.gc.ca
Incze, Lewis	University of Southern Maine	lincze@usm.maine.edu
Isenor, Anthony	DRDC Atlantic	anthony.isenor@drdc-rddc.gc.ca
Jaques, Yves	FAO of the UN, Fisheries and Aquaculture Department	yves.jaques@fao.org
Kabbara, Nijad	Marine Research Centre - NCSR	nkabbara@cnrs.edu.lb
Keeley, Bob	Fisheries and Oceans Canada	keeleyR@dfo-mpo.gc.ca
Kelly, Maurie	Penn State University	mck4@psu.edu
Kemp, Zarine	University of Kent, UK	Z.Kemp@kent.ac.uk
Kenchington, Ellen	Fisheries and Oceans Canada	Kenchingtone@mar.dfo-mpo.gc.ca
Kennedy, Mary	Fisheries and Oceans Canada, BIO, IODE-GEBICH	kennedym@dfo-mpo.gc.ca
KIRN, Sarah	Census of Marine Life Gulf of ME	sarah@gmri.org
Kostylev, Vladimir	Natural Resources Canada	vkostyle@nrcan.gc.ca
Kraberg, Alexandra	AWI, Biological Station Helgoland	Alexandra.Kraberg@awi.de
Kumar, Dr. S. Senthil	Laboratory for Conservation of Endangered Species, CCMB Annexe-1	sensubbiah@gmail.com
Lambert, Timothy	Fisheries and Oceans Canada	lambertt@mar.dfo-mpo.gc.ca
Lawton, Peter	Fisheries and Oceans Canada, Centre for Marine Biodiversity	lawtonp@mar.dfo-mpo.gc.ca
Macario, Ana	Computing and Data Centre, Alfred Wegener Institute for Polar and Marine Research	Ana.Macario@awi.de
MacLeod, Bruce	University of Southern Maine	macleod@usm.maine.edu
Maillet, Gary	Fisheries and Oceans Canada	mailletg@dfo-mpo.gc.ca
Marcil, Julie	Fisheries and Oceans Canada	jmarcil@dal.ca

NAME	AFFILIATIONS	EMAIL ADDRESS
Meaux, Melanie	NASA GCMD	mmeaux@gcmd.nasa.gov
Meredyk, Shawn	Centre for Marine Biodiversity (CMB)	shawn.meredyk@unb.ca
Mitrofanov, Igor	CEP	igor.mitrofanov@videotron.ca
Mitrofanov, Igor	Caspian Environment Programme	Mitrofanov_igor@hotmail.com
Mohamed, Musab	Linkoping Univesity-Sweden	musos096@student.liu.se
Moncoiffe, Gwenaelle	British Oceanographic Data Centre (BODC)	gmon@bodc.ac.uk
Neal, Phillip	ICoMM and Marine Biological Laboratory at Woods Hole, MA	pneal@mbl.edu
Nowlan, Robert	Fisheries and Oceans Canada	nowlanr@dfo-mpo.gc.ca
Nozères, Claude	Fisheries and Oceans Canada	NozeresC@dfo-mpo.gc.ca
Ó Tuama, Éamonn	Global Biodiversity Information Facility Secretariat	eotuama@gbif.org
Olalekan Muideen, Gbadamosi		lakastabadmus@yahoo.com
Parke, Michael	NOAA Pacific Islands Fisheries Science Center Coral Reef Ecosystem Division	michael.parke@noaa.gov
Patterson, David	Marine Biological Laboratory, Woods Hole	dpatterson@mbl.edu
Payne, John	Pacific Ocean Shelf Tracking (POST)	jcpayne@u.washington.edu
Perry, Daniel	National Oceanography Centre, Southampton	dvp@noc.soton.ac.uk
Ready, Jonathan	Swedish Museum of Natural History	jonathan.ready@nrm.se
Rees, Tony	Commonwealth Scientific and Industrial Research Organisation, Australia	Tony.Rees@csiro.au
Ricard, Daniel	Dalhousie University	ricardd@mathstat.dal.ca
Rigby, P. Robin	Kyoto University	R310x@hotmail.com
Romaine, Steve	Fisheries and Oceans Canada	romaines@pac.dfo-mpo.gc.ca
Rozo Garzón, Daniel Mauricio	Marine and Coastal Research Institute-INVEMAR, Colombia	drozo@invemar.org.co
Rutherford, Dennis	Fisheries and Oceans Canada	rutherfordd@pac.dfo-mpo.gc.ca
Rutherford, Kate	Fisheries and Oceans Canada	rutherfordka@pac.dfo-mpo.gc.ca
Rutzen, Imme	Institute for Wildlife Research, University of Veterinary Medicine Hannover Foundation	imme.rutzen@web.de
Salih, Musab	Linkoping University	musos096@student.liu.se
Sampson, Daniel	Massachusetts Office of Coastal Zone Management	daniel.sampson@state.ma.us
Schmidt, Allison	Dalhousie University, Department of Biology	aschmidt@dal.ca
Shyka, Tom	Gulf of Maine Ocean Observing System (GoMOOS)	tshyka@gomoos.org
Silveira, Fabio	OBIS; University of São Paulo	fldsilve@usp.br
Soule, Deb	New Hampshire Department of Environmental Services	dsoule@des.state.nh.us
Steenbeek, Jeroen	UBC Fisheries Centre	j.steenbeek@fisheries.ubc.ca
Stephen, Rosamma		rosammastephen@yahoo.co.in
Taconet, Marc	Food and Agriculture Organisation (FAO)	marc.taconet@fao.org

NAME	AFFILIATIONS	EMAIL ADDRESS
Terry, Johnston	Fisheries and Oceans Canada, University of New Brunswick	johnstontl@mar.dfo-mpo.gc.ca
Thana, Kohila	Fisheries and Oceans Canada	thanak@mar.dfo-mpo.gc.ca
Tilburg, Christine	Gulf of Maine Council on the Marine Environment	ctilburg@securespeed.us
Tittensor, Derek	Dalhousie University	derekt@mathstat.dal.ca
Tyler-Walters, Harvey	Marine Biological Association of the UK	htw@mba.ac.uk
Upson, Patrick	Fisheries and Oceans Canada	upsonp@mar.dfo-mpo.gc.ca
Van Guelpen, Lou	Huntsman Marine Science Centre	arc@mar.dfo-mpo.gc.ca
Vanden Berghe, Edward	OBIS	evberghe@iobis.org
Vanhoorne, Bart	Flanders Marine Institute (VLIZ)	bartv@vliz.be
Vides Casado, Martha Patricia	Marine and Coastal Research Institute – INVEMAR, Colombia	mvides@invemar.org.co
Wolff, Nicholas	University Southern Maine	nwolff@usm.maine.edu
Worcester, Tana	Fisheries and Oceans Canada	worcestert@mar.dfo-mpo.gc.ca
York, Amber	Woods Hole Oceanographic Institution (WHOI)	adyork@whoi.edu
Young Morse, Riley	Gulf of Maine Ocean Observing System (GoMOOS)	riley@gomoos.org

Appendix 3. Schedule of Events

	Tue. Oct. 2	**Wed. Oct. 3**	**Thu. Oct. 4**
08:00-08:30	Registration (8:00-10:00)	meet and greet	meet and greet
08:30-10:00		5) Data Use and Analysis	7) Integrating Different Types of Data
	Opening (9:00)		
10:00-10:30	Break	Break	Break
10:30-12:00	1) Visualisation Tools	5) Data Use and Analysis (cont'd) 6) New Data Systems	7) Integrating Different Types of Data (cont'd)
12:00-13:30	Lunch	Lunch	Lunch
13:30-15:00	2) Habitat and Ecosystems 3) Species Names Management and Tools	Posters / Demonstrations	Review and Synthesis
15:00-15:30	Break	Break	Break
15:30-17:00	4) Metadata Developments	6) New data Systems cont'd	Closing (16:00)
	BIO Reception (17:00-19:00)	Pier 21 Banquet (19:00-20:00)	