

**Washington Workshop Minutes
Natural History Museum
September.3th - 5th, 2002**

ATTENDANCE:

September.3rd-5th: Ed Wiley, Karen Stocks, Phoebe Zhang, Fred Grassle, Pamela Brodie, Daphne Fautin, Phillip Lee, Gary Rosenberg, Ann Bucklin, Kirsten Yarnick, Ron O'Dor, Jon Norenburg, David Hyrenbach, Bob Buddemeier, Vardis Tsontos, Mark Costello (Chair), Angela Martin (recorder), Debbie Steinberg (absent 3rd).

Introduction: Mark Costello, OBIS chair

The vision of OBIS is for marine biological data to be accessible rapidly and freely through the world wide web. Future expectations of contributing data to on-line databases will be considered as standard practice including archiving type specimens in museum collections. It is argued that such data will stand the test of time better than most 'analysis and opinion' based publications in journals. To achieve this demand, extraordinary collective efforts and cooperation amongst taxonomists, ecologists, oceanographers, database experts, information managers and brokers must be done.

At present, an impressive range of taxon information and software tools that make data handling and analysis easier is available in databases, some on-line. This week will focus on what databases are available to participate in OBIS and what funding, new or existing is needed to provide the databases on-line.

In November there will be a similar meeting in Brussels for European based researchers.

Development of the OBIS portal/website: Phoebe Zhang

The goal of the OBIS portal is to have a way to gather information for the user using a species based geo-referenced information center with overlaying environmental data. The portal was originally developed by Karen Stocks, Phoebe Zhang, and Fred Grassle (1999-2000) through funding from Sloan. In 2000, the OBIS portal consisted of 8 NOPP funded OBIS projects. In 2001, OBIS received a NSF grant for a dynamic portal data searcher over a distributed network of databases, built on a set of standards. The structure is a call interface for data exchange between projects and to the OBIS portal. There are four ways to search the portal. They are 1-NODC search interface: (enter species and geographic region and choose databases); 2-OBIS dynamic mapper: take data from OBIS portal and overlay on bathymetry and temp (Tony Rees and Jeremy Bartley); 3-taxonomic name search, species 2000; 4-biodiversity modeling services (WHYWHERE).

Future plans: NODC, SOC, HMAP, provide more biological and environmental data, mapping service (resizing, projection), Name service: taxonomic history, biodiversity modeling services, set of bioinformatics tools, and genetic information serviced through GenBank.

The FishNet distributed biodiversity information system: Ed Wiley

The setup of the FishNet portal takes advantage of datasets of marine fish distributions (including freshwater), which is presently at 5-8 million records. The portal tool used is Species Analyst. An example as to the quality of the Species Analyst tool is when given a name to search, it takes 59 seconds to search 21 collections finding 719 species. This search can be downloaded to excel. Also available online is an atlas of where species are found. Future plans are to connect to environmental data.

A special feature on the portal is GARP, which is an ecological niche model that can make inferences to distributions (ecological variables). GARP can predict marine species in natural habitat and where they will show up. Problem is GARP only uses positives. Survey data uses 0 to show presence and absence. This should be added to the model.

Development of a dynamic biogeographic information system: A pilot application for the Gulf of Maine: Vardis Tsontos

GMBIS is a prototype for assisting OBIS in the creation of reusable GIS tools. The databases used in this project were provided from other institutions (Gulf of Maine data). EASy (environmental analysis system), is a four dimensional GIS containing lat, long, depth and time. Data connectivity (local/remote), PC desktop and web enabled GIS application is provided through Netviewer. The thematic approach taken was to demonstrate utility of an information system in the context of scientific biogeographic storylines. It provides a form of synthesis. Data includes salinity, temp, various physical info, satellite imagery, and more. EASy provides data integration, assimilations, and visualization that can be used on your own data.

Presently, EASy is available to collaborators of OBIS (not commercially available). GMBIS has provided a good testing ground of reusable tools that are of value to OBIS.

Biogeoinformatics of Hexacorallia: Daphne Fautin

The geographic extent of the project consists of Hexacorallia of the world. Here, the biological component was discussed (Taxonomic breath – entire subclass). The website began as oracle, then changed to coldfusion (webinterface), and presently is arc ims (GIS function). The geo-referenced database numbers in the several thousands and is growing (961 genera, 7602 binomens and trinomens, 2612 original descriptions, 2712 species list, 1648 type specimen lots, 4404 images). Black corals and hard corals are being worked on and other taxa are near or at completion. Also, there is now a linkage to hydrozoans and fossil corals. Databases are also in place for nematocysts, symbionts and associates. Secondary data from literature is available and is verified by D. Fautin. The Smithsonian museum is also linked on-line (interested in databasing yet most specimens are poorly identified if at all which makes the task very overwhelming for them in quality control).

- **CephBase-A global web portal for information on cephalopods: Phillip Lee**

CephBase is a species based searchable website yet geographical searches can be done through OBIS. There are presently 5182 searchable references. Choices for searches include: species search, image search, reference database, predator database,

prey search, and a video database. References can be downloaded to a .pdf file. Also, there is a CIAC beak database (version 1.0). All information is linked to Kansas University. At present, NSF funded CephBase to continue to enter data and make the interface available and useable to individuals such as High School students. Future hopes are to link to GenBank and paleobiology information to look to the past in how things have changed through time. People who want to contribute distribution and abundance data, CephBase will take care of the database preparation and linked to them. Quality control is still an issue.

A biotic database of indo-pacific marine molluscs: Gary Rosenberg

At present, the project is 3 months away from finishing data entry. All entries were entered over the internet (based at Biolink in Philadelphia). Each individual entering data is working on certain families (therefore no overlap of all 4 places). The database captures all names of species (type, synonym, common among spelling, etc.). There is a folder system for geographic regions where regions are named to help in placement. Presently, the project is working on ecology, biology of taxa, references, types, distribution, and images. The site does have GIS ability yet point data is not in the system yet (OBIS portal for mapping). Quality control has been done on database. Future goals are illustrations (30,000 species), worldwide marine taxon, distribution.maps (GIS) bringing it in house, linking more specimen databases, and including land and freshwater. The project is 2 weeks away from announcing webpage.

Euphausiids: an OBIS tool for uniform standards of species identification: Ann Bucklin

The website consists of DNA sequence data and morphological taxonomy (200 species of copepods on-line (DNA) and 86 species of Euphausiids). The website is linked out to a variety of other databases (GenBank, OBIS, GLOBEC). Presently, the project is preparing a curriculum for schools (protocols to be modified).

Standardization: Database design, JGOFS system, hierarchical free software and easy access to other databases. There is a hierarchical view from copepod to species (number) and DNA sequence line. Accession number and station number (GLOBEC) are not there yet. Refined species DNA sequence shows species may be different therefor morphology taken into account (DNA sequence, morphology, correct naming).

Seamounts On-Line: A biogeographic information system for seamounts: Karen Stocks

Seamounts are elevations in the oceans floor (1000m). As technology grows, so does the amount of seamounts. They are important to fisheries and tend to be over-fished (management needed). Seamounts are a good case study for marine biodiversity since they are a center for speciation, refugia for relict populations, and stepping stones for dispersal. The database consists of data from peer reviewed, non peer and electronic datasets. Presently on-line there are 2500 species, 8500 records, and 80 seamounts. The format is ACCESS on the computer and updated to oracle. An added feature is data from surveys containing present/absence information.

Looking at a seamount will let you know each sample taken there. Future goals are to add more seamounts (only have 15-20% now), continue literature, .hurl images, access to R/V PPSHIRSTOOV holdings, current sample efforts, and new data types such as synonymies, range data, life history/reproduction.

Spatial Ecological Analysis of Megavertebrate Populations: David Hyrenbach

The project started May 2002 (three year funding). The objective is to enhance the understanding of the distribution and ecology of marine mammals, birds and turtles through modeling and testing hypothesis. Future products will be a geo-referenced digital database, relational database, history biology information, GIS web based for display, query and exporting geo-referenced data, and an outreach program (education). Present focus is on geo-referencing at sea distributional data, shore based observation of breeding/hauling, and individual movement of animals (telemetry). Types of sea data are sighting data (locality/time), survey meta-data, and environmental data. Types of sea data categories are primary (vessel based and aerial surveys), secondary (primary and fisheries dependent), and tertiary (sighting of fishery observers).

Census of Marine Fish (CMF): Definitive List of Species and On-line Biodiversity Database: Stan Blum

Current funding is going to increasing references to data. The project consists of a catalogue of fishes that is accessed through FishBase. It is involved in doing quality control for some institutions and museums to correct problematic data. New funding has been granted to cover 100 people (specialists) who are reviewing the CMF and to add any information that is needed (purpose is to collect all the data of scientists that will be retiring so data is not lost in the future).

Handout was given (annotated checklist for fishes).

Diel, Seasonal and Interannual Patterns in Zooplankton and Micronekton Species Composition in the Subtropical Atlantic: Debbie Steinberg

This was a 2 year project ending February 2003. The project consisted of the census of species at the Bermuda Atlantic time-series study (BATS). Funding does not come from BATS for species section of project yet sampling is. Monthly sampling has been done since 1988 (temp., sal., nutrients, rate/sediment flux). It is an open ocean site with 150 sps of calonoids alone. Zooplankton and phytoplankton in samples are used for biomass only, not species composition. Phytoplankton is being stored yet not being studied. It is possible for future work that specimens could be preserved in alcohol for DNA analysis. Level of confidence and standardization of data are based on three levels that are subjective. Information of physical data is available on website along with all other BATS data available at <http://www.bbsr.edu/users/ctd/nisklist.html>. Zooplankton data is at: <http://www.bbsr.edu/users/ctd/zooplankton.html>. Satellite data and cruises are available to fill in monthly gaps. There is an issue of changing raw data to integrated data. Future goals consist of images/diagnostics of species to be done.

A Case Study of the Development of an Integrated Biogeographic Resource: Bob Buddemeier

This is a bioinformatics project (LOICZ lobal database for coastal zone biochemistry, PEET anemone taxonomic database). The purpose of the project is to connect and expand the database of the two projects (linked dynamically interactive databases, tools and servers, others tapping into data). Issues that will be addressed is the importance of real, right now, applications and utility; importance of a user community; importance of spin-offs and spin-ons, community building; and importance of local institution understanding and support. The talk focused on georeferencing, ARC-IMS running georeferenced. Presently, the website has a multiple prototype map interface where eventually one or 2 of the best will be chosen (presently using different map software). The OBIS portal should have only two or three collaborations to avoid adding too much to the portal. A new feature on the Hexacorallia website is living and fossil scleractinians. Both are running in oracle so can search individually (working on combining data pages). Future Goals are to create a more user friendly environment variables interface, taxon cart so multiple taxa can be searched, seawifs (works in 1/2 degree grid that is fast and cost effective) variables plus coupled GCM projections to databases, geographic input filter for batch biodata searches, clean up /standardize display and entry tool, refine links with other sites and projects and interoperability, and have historic data in reference to present data. A handout was circulated.

Nemertes: A Portal to the Nemertean Universe: Jon Norenburg

The purpose of the project is to build a better knowledge basis and curation of information on nemerteans. Started work on literature databases and nomenclature and presently working on images yet they are not on the website. There are 150 sps of nemerteans on the website (Zoogene for id's). The website is Macserver where problems are now arising (also issue of firewall). There is no point data at present due to lack of funding in search and data entry. 46 sps of nemerteans have been sent to ZooGene. We estimate the number of nemerteans species at 1200. Now after working out synonymies, which will drop the number originally viewed (4000). OBIS portal is a good way to show data as a teaching tool.

A Relational Database of the Aplacophora: Amielie Scheltema

Presently, the database is on Mac (filemaker pro). It is searchable from the web page (eight files such as taxon, geography). Work will be continuing on the database without funding. The database and webpage are moving to Sweden with Christopher Schander who is taking over Amielie position. There is cruise data in the database (153 records). This collection is important to capture since many id species are there. There are no images on-line, only drawings.

Hydrozoan Taxonomy: Steve Cairns

The grant for the project covers three families, 26 of hydroids. The scheme is to adopt the same web plan as Daphne F. The database will also include molecular and mineralogy. The database is mostly secondary in nature yet primary in terms of some

collections and molecular information. The database is small yet robust. Presently are beginning the morphology and mineralogy data. Symbioses are included.

Future of Marine Animal Populations: Ransom Myers

The website is <http://www.fish.dal.ca/welcome.html>. All information entered into the database is checked by Ransom. Therefore quality control is taken care of upon entry. The purpose of the project is statistical design and consulting (advice on design). The objectives are interpolation and overlap, using tools of epidemiology to find common frame, biogeographic modeling (pelagic fish), and movement models. Focus is on data model interfacing and standard data exchange formats.

One workshop joint with NCES has been done in connection with this project.

Progress with the Census of Marine Life; Ron D'Or

CoML is an international research program merging changes in diversity and distributions. Elements of CoML are HMAP, FMAP (modeling), SCOR (technology working group), and 7 initial field projects. All are connected to the OBIS. Some great sources of data are industry, oil, and fisheries. All data has the eligibility to be part of OBIS.

CoML started in 1997, the steering committee met in 1999. Phase one started in 2000 with time lines for each project as follows: HMAP (2000-2010), OBIS (2000-2015), FMAP (2002-2015). Phase two runs from 2005-2010, phase three from 2010-2015. The CoML IC committee are committed to OBIS and want to know what is needed to continue OBIS (where do you want to go). Implementation of the International committee creates a way to look to various countries to see where data is and what funds are needed for projects. CoML does not want to be strictly limited to either conservation or industries but to be a connector of the two.

Seven initial projects were: NAGISA, POST, TOPP, GoM (GoMAP, GoMOOS, GMBIS), MAR-ECO, CEDAMAR, Chess (Note handout for details on each project in detail). The POST project uses animal movement to get a global snapshot (256,000 codes, tagged 25 species). TOPP project looks at 20 species for real time data linked by radio and satellite information. Bioprobes sample the ocean animals (ex. elephant seals), when diving to give better understanding to physical oceanographers.

The CoML plan of ten years will have created projects through funding where OBIS will be the finalized output. Global field projects visions in the future are inshore, coastal, ridge, deep sea vents and other ocean habitats that have not been done. CoML will support projects where all habitats will be represented which can be inferred to other similar areas. GoM and ECO-MAR do not use technologies. With this in mind that CoML cannot drive technologies, SCOR was created. They will be interested in feedback from this meeting. Noted in discussion that ZooGene has developed technology that can be used that could be fine tuned to species data (Census can't drive DNA of every project yet can understand the use). Biological sensors would be a great legacy to CoML.

It is important to note that CoML projects were created to get OBIS into the water where they can learn to use the information collected. Funding was only given to the census. Therefore outreach and technology are needed for OBS to remain. OBIS is

the fundamental tool that people will appreciate where everyone may want to do something different yet all is provided. The databases are high quality projects in themselves that provide the information system where products are needed to modify technology and outreach. Rutgers is committed in maintaining the website which is not a collections of databases but of collaborations.

Discussion

• OBIS tools

- Suggested that the OBIS portal needs to explain in more detail as to what each tool available does. This could be corrected by adding a write-up on each one available on the portal (Fred G.).
- It was suggested by R. Myers that standards are needed in the development of software.
- Are GMBIS data tools available on OBIS website? The answer is yes. The GMBIS tools are also able to prepare data for analysis (noting presence/absence approach).
- FMAP will come up with away to deal with fisheries data yet it must be searchable with OBIS. It was stated that FMAP should create a system that communicates with other tools already available through OBIS. (V. Tsonos).
- One view is there needs to be better communication between marine, terrestrial and freshwater in relation to OBIS. OBIS needs to take into consideration that DIGGER will replace Species Analyst. Therefore, efforts should be on creating one software (S. Blum).

• Gazetteer

- A Gazetteer should be created on the OBIS website. A listing of cruises will help fill in the gaps that may be present to some databases. Also, this can address the issue of quality control.
- Some cruises available is the Challenge (Daphne F.) and Pilsburg (Ed W.).
- A problem to note is that Gazetteers are an rfp file that could be expensive. Efforts should be made in contacting the NOAA library for their information (NODC)(Fred G.).

• Portal

- Suggested that OBIS have a link at the front page of the portal as to how to become a member. It should also include techniques, what is critical for data expectance, and who you can talk too. Everyone must know the expanded benefits of OBIS (Ann B.).
- What should be on the portal: species, lat/long and salinity (Mark C.)? The marine community relies heavily on lats. and longs. B. Buddemier feels OBIS needs to prioritize what is put on the server. They cannot remain focused solely on a biology geo-referenced based program. A template an acceptable database to OBIS should be made available on the portal for users to upload. This can then be sent to the OBIS portal (P. Zhang.).
- In relation to project websites, self-evaluation recommendations and guidance needs to be done in order to be compatible to the OBIS portal (P. Zhang).
- Generic data needs to be addressed (GenBank, EASy) (Fred G.).

• Quality control

- Should OBIS be responsible in relating quality control to databases (Daphne F.)?
- Is it correct in saying that museum data has implied quality control since the specimen was available (Fred G.)?
- Question of what would OBIS like done with data FishNet will serve in relation to quality control issues. Who collected it is not important to everyone. (Ed W.)
- **Meta data standards**
 - Metadata protocol: what policy will OBIS meet (K. Stocks)?
 - Species 2000 asks a given set of questions in relation to data provided. Should OBIS go that way (M. Costello)? These questions could serve as a maintenance function. OBIS needs to identify the standards and have the members follow them. Yet, the question is at what level of meta-standard or confidence? This should be left for the technical working group.
 - When decided, the OBIS data policy should be posted on website (A. Bucklin). GBIF post their policy on their website. With a policy present, OBIS is providing users with knowledge of data quality, numerical confidence and buyer beware. One problem that arises is in terms of OBIS not owning the data served therefore no direct control over quality status (ex. Museum data).
 - With standards, there is a breakdown of information based not on strength vs. weakness but survey vs. type (P. Lee).
 - It was mentioned that there should be a creation of a dictionary for standardization of data fields.
 - Should OBIS have a precision field? This could be used as a search tool (G. Rosenberg). This could be something the tech group could look at.
 - Question of creating accession numbers in relation to citing information that is referencing site material. Looks good for promoting site (A. Bucklin).
 - Need to address the problem of different maps. Convergence is needed to get large groups of species represented geographically (G. Rosenberg).
 - Should the fields be expanded beyond what is presently served? Should geo-referenced and depth be optional? GBIF has standards for biological data that specifies fields needed. Fields need to be broad to accommodate everyone so people do not have to look over many sites to find what they want. Also, OBIS data fields should have a commonality in relation to their fields that correspond with fisheries and museum data. (A. Bucklin).
 - Biological data should be separate from environmental. When environmental data is attached to a specimen does OBIS want it?
 - With standards in place, someone will be needed to look at each dataset coming in to decide relevance and quality. Also, emerging data may be lost since it may not meet the criteria (B. Buddemier). A solution could be OBIS offering raw and quality controlled data (P. Zhang). Should OBIS look at non-OBIS datasets (A. Bucklin)?
 - Noted that databases need to contain both geographic and taxonomic data. A list of species is not useful (D. Fautin). Yet at what state are different

databases in? The goal of CoML is to cover all species of life. Realistically, we should be looking at capturing a biodiversity snapshot.

- How will State and Local politically make policy based on OBIS data form a long term goal? (E. Wiley). A question arises as to whether a meta-data committee should be formed.
- Overall, OBIS can encourage standards but not impose since it is on the responsibility of the projects.
- **TDWG**
 - Question of OBIS being involved in the development of TDWG. The draft of TDWG is on their website (biological collections working group). Note that TDWG is primarily museum based where flags will be present (DIGIR to flag survey, museums, etc.). TDWG is controlled by GBIF (S. Blum).
 - How much is OGC involved in TDWG (P. Zhang)? Concern is with all geographic data then biological data. TDWG will bring the two together accommodating including OGC.
 - **ACTION: Everyone is to e-mail Mark C. their policies that will be generically put together and critiqued by TDWG.**
- **Data capture**
 - There is a budget for data rescue. Requests should be directed to Ron D'Or.
 - Russian data: Bill Eschemeyer is having problems getting a hold of Russian data. MAR-ECO and HMAP do have Russian collections but are having problems with the connections.
 - Access is needed to GOOS since environmental data is time consuming.
- **Archiving**
 - Is there a need for archiving since some partners are no longer funded by OBIS? These groups may be at risk of OBIS losing their data. A data policy should be in place. FishBase could serve as a template since a cd already exists. (Fred G.)
- **Polygons/Mapping**
 - Should a system of polygons be adopted? This will help in the geo-referenced data. All participants agreed on the important of polygons.
 - NatureServe is developing polygons. Should protocol be similar? (Fred G.)
 - Was suggested that the polygon software would have to serve on the central server to avoid a polygon on each site since most databases do not and will not have the capability to do so (Stan B.).
 - Polygon and map searches are available on the OBIS server (digital maps). Political maps, fishing maps, etc. all have points that show up on maps throughout the database. They are expensive to buy and will bring in new OBIS members who are interested in access to them. (Daphne F.)
- **Outreach**
 - Outreach: CephBase has received a NSF grant for outreach. They are wanting suggestions as to how to approach high school students. This raises the question of funding towards creating an outreach program for promoting OBIS. Teachers need help in understanding and interpret the data that OBIS provides (address national science standards).

- Question arose as to whether outreach needs to be done for PEET programs. Agreed OBIS should look into contacting all PEET PI's.
- Does OBIS need an educational science outreach committee? (Fred G.)
- **What should OBIS provide: Future goals**
 - Does OBIS want to own the database? What is in the future for expansion?
 - Will OBIS provide the environmental overlap? OBIS should look into expanding the connections to environmental data that is provided from the source, not from OBIS (Ed W.). Data should be accessible through the portal, not served by OBIS (searches databases). V. Tsontos notes that GOM environmental data is searched thru the server but is not matched to species/biological data.
 - Is it safe to say that OBIS future goals is to focus away from a museums and look towards more of a survey approach. In response F. Grassle, stated that that was the case.
 - Funding is needed to link groups in the synthesis and modeling phase (Debbie S.). OBIS should fund pilot projects that show what OBIS can provide usable information to policy makers. Analysis needs to be done that demonstrate the utility of the data OBIS provides (Ed W.). In response, D. Fautin stated that funding is available in the states to test the system (8 million). Yet what project should be chosen? Funding needs to go to concepts that are believable. Possibly in case studies or boundary decisions. Maybe Seamounts? Also, it was noted that focus should be on creating a project that shows the datasets in use together. OBIS needs to look at groups that can come together and provide a conservation paper (Karen S.). An example could be tables of predator/prey associated with FishNet looking at food web modeling. (Phillip L.)
 - It was expressed that OBIS is working towards generalizable concern. Managers are asking for data that will help them in decision making in the future. We need to use the data we have now to predict trends in the future (Bob B.).
- **Five Year Plan: in the works by Fred Grassle**
 - A question is how central will OBIS be? Funding is offered to keep data static yet not dynamic. What funding is available for projects to remain dynamic (Phillip L.)?
 - For the five year plan, F. Grassle needs to know what funding is needed by each project to bring them to completion. Also, costs of maintenance should be included (Upgrades need to be addressed every three or four years). The draft proposal will need to be ready for Sloan by December. It has become apparent that there are databases out there that could contribute to OBIS yet to not have the funding for maintenance of the site. We need to look at the different projects as a community to create a database that is easy to fund (each represents a base to start with).
 - So far, some projects have responded with an estimate on costs: FishNet maintenance: 75,000 a year for upkeep (5 year – 75,000 plus 65 institutions, 50 million specimens, 20 million lots) (75% marine geo-referenced, 10% freshwater geo-referenced). Therefore, 75,000 a year plus 100,000 year for 65

institutions; Pacific molluscs maintenance is 500,000 thousand; CephBase maintenance is 32,000 a year to keep going the computer server and one half of a computer person's salary; FMAP maintenance is two million a year in order to bring in fisheries data (it is the cost of convincing them to do it).

- **CoML Press release**

- CoML is working on a press release for November that proposes to bring all parties of OBIS together. Each project will be contacted and PI's will be brought together at the press release gathering. Ron is cautioned by members to not oversell OBIS at this point since some feel it is not ready.
- The value of OBIS should be presented in peer-reviewed literature. This could be in connection with the CoML press release date.