

# Developing OBIS into a Tool to Provide Reliable Estimates of Population Indices for Marine Species from Research Trawl Surveys

Robert M. Branton<sup>1</sup> and Daniel Ricard<sup>2</sup>

<sup>1</sup>Bedford Institute of Oceanography  
1 Challenger Drive, Dartmouth, Nova Scotia, Canada B2Y 4A2  
email: BrantonB@mar.dfo-mpo.gc.ca

<sup>2</sup>Dalhousie University, Department of Biological Sciences,  
1355 Oxford Street, Halifax, Nova Scotia, Canada, B3H 4R2

## Abstract

Research trawl surveys from Canada's Department of Fisheries and Oceans (DFO) and the US National Oceanographic and Atmospheric Agency (NOAA) are used as a basis for developing a road-map to prepare research trawl surveys in general for public access via systems such as Ocean Biogeographic Information System (OBIS). Data quality issues associated with surveys includes validation of species names, treatment of zeros, data standardization techniques and provision of confidence limits. Suggestions to improve the OBIS system include support for summary statistics and length classes as well as addition of a gazetteer facility. The Bedford Institute of Oceanography's recently established OBIS provider service is also described.

Keywords: fisheries; trawl surveys; OBIS schema; data quality control.

## Introduction

Traditional research trawl surveys (Doubleday and Rivard, 1981) are species rich (100s) with analysis focused only on commercial species (~10s). Recently other species (*e.g.* mega-invertebrates) have been added to sampling protocols thus enabling investigation of ecosystem issues (DFO, 2003). Current expectations are that the Ocean Biogeographic Information System (OBIS) will provide a basis for interoperability of these data with other scientific disciplines (Grassle, 2000). Using Canada's Department of Fisheries and Oceans (DFO) and the US National Oceanographic and Atmospheric Agency (NOAA) research trawl surveys we provide a basic road-map for preparing research trawl survey data sets for public access via systems such as OBIS. This presentation focuses on: DFO/NOAA trawl surveys, trawl survey data quality issues, and ways to improve OBIS. A description of the new OBIS provider service located at the Bedford Institute of Oceanography (BIO) is also given.

## Methods

DFO and NOAA have been collecting data from standardized research trawl surveys on the east coast of North America since the 1960s and have considerable local expertise and software for preparing resource assessments from these data. The first significant effort to integrate these data for biogeographic studies was in 1995 as part of the East Coast North America Strategic Assessment Project (ECNASAP) (Brown *et al.*, 1996). Amongst other things, the ECNASAP project integrated basic survey catch data (numbers and weights) from 5 fisheries laboratory databases (Table 1), providing observations on 276 species from ~50,000 fishing sets for the period 1970-95. Although the surveys are ongoing, this dataset has not been updated since '95 and is presently only available on Compact Disk from project principals as a 300+ column flat file.

Table 1. The ECNASAP project integrated data from 5 fisheries laboratory databases for the period 1970-95.

Laboratory	Laboratory Location	OBIS Collection Code
North Atlantic Fisheries Centre	St. John's, Newfoundland, Canada	DFO-NFLD
Maurice Lamontagne Institute	Mont-Joli, Québec, Canada	DFO-NG
Bedford Institute of Oceanography	Dartmouth, Nova Scotia, Canada	DFO-SF
Gulf Fisheries Centre	Moncton, New Brunswick, Canada	DFO-SG
Northeast Fisheries Science Center	Falmouth, Massachusetts, USA	NMFS-NEFSC

In 2002, BIO's Scotian Shelf Summer survey data was temporarily placed directly on the OBIS portal as an interim measure until a permanent OBIS provider service could be installed at BIO. Since 2002, BIO has developed a relational database version of the ECNASAP data for local research as well as for serving to OBIS and the 'Gulf of Maine Ocean Data Partnership'. These efforts are expected to provide a basis for DFO and NOAA to develop publicly accessible and near real-time links to all of their ongoing surveys. Providing public access to other than the basic survey catch data presently contained in the ECNASAP dataset will require careful attention to trawl data quality issues and to extending limits of the present OBIS schema.

## Results

Following are the major issues to be considered when preparing data such as the DFO/NOAA and ECNASAP research trawl surveys for systems like OBIS.

### ***Species list validation***

OBIS uses the Species 2000 Catalogue of Life (CoL) annual checklist CD-ROM as its basis for validating species names. All OBIS providers are therefore recommended to

use this same CoL checklist to find the most current scientific names and hierarchies for species contained in their databases. The CoL checklist is available online from <http://www.sp2000.org/>. A preliminary comparison between the ECNASAP species list and the Integrated Taxonomic Information System (ITIS) database, which is part of the CoL database, indicates that approximately 50 of 276 species names in the ECNASAP database are either obsolete or incorrectly spelled. Local taxonomists and survey staff should review discrepancies in the local lists and make corrections where appropriate. They should also note which species are difficult to identify or not routinely sampled. The taxonomic hierarchy data are particularly useful as they can be used to prepare cumulative discovery curves (Costello, 1996) for each hierarchy level. This type of analysis can help identify new species appearing as a result of protocol changes from those appearing as a result of environmental changes.

### ***Taking care of zeros***

Fisheries surveys are primarily intended to provide observations (*e.g.* numbers and weights) for species captured in the sampling gear. Absence of a species is not recorded during surveys and hence not recorded in the trawl survey database. Fisheries analysts work with well documented pre-established stock area definitions and employ database queries and analysis programs that automatically generate ZERO (0) for the missing data, as opposed to NULL values. ZERO is interpreted as the absence of a surveyed species from a trawl, and would be included in calculations of averages; a NULL value would signify that no information is available, and would be omitted from any further analysis. Survey species lists must also include these established stock area definitions (*e.g.* lists of survey strata) thus providing a clear indication of when and where fisheries analysts are interpreting the missing data as ZERO.

### ***Adjusted v. standardized observations***

The probability of a particular organism being retained in a research trawl depends on many factors, not least of which are fishing vessel and gear used. Data contributors should provide distinct survey series names (*e.g.* OBIS-Collection Code) for each unique survey vessel, sampling gear, stratification plan, and season combination. Given that good data management practices dictate that data be stored as they were recorded, observed values (*e.g.* observed individual count and weight at length, sex and maturity) given to end-users or to systems such as OBIS should be automatically adjusted by sampling ratio (*i.e.* total/sample). Adjusted numbers-at-age from sampled materials (*e.g.* otoliths and scales) should be based on stock specific age-length keys (*e.g.* proportion at age for given length). Observations from sets where gear has been damaged although containing rare organisms should not be given to users expecting adjusted results. End users should be further aware that not all fisheries laboratories routinely standardize their observed values for distance towed (*e.g.* standard/observed) or species for catchability by gear (*e.g.* proportion caught at length). Databases should include sufficient metadata to clearly indicate how the data at hand have been adjusted and standardized.

### **Confidence limits**

The DFO and NOAA research trawl surveys discussed here all follow the same basic stratified random design. Relative indices such as 'average per standard tow' should include variance or standard error. Absolute estimates such as 'total biomass' and 'total abundance', if presented, should be peer reviewed and given with Internet links to citable publications (*e.g.* Canadian Science Advisory Secretariat - <http://www.dfo-mpo.gc.ca/CSAS/>).

### **Recommendations**

These recommendations are for improving the OBIS system as a whole. Extending the schema by providing support for area summaries (*e.g.* number of observations and variance) at the same time as allowing for more details (*e.g.* length, parent catalog number) is intended to broaden the range of information and products conveyed to the public.

#### **Add new keywords to existing schema concepts**

- Basis of Record – stratum average, stock estimate.
- Locality – stratum, ecozone, grid square, stock area ...
- Life Stage – maturity stage, age class.

#### **Add new schema concepts**

- Number of Samples and Sampling Units in Locality.
- Length Class of Observed Individuals.
- Variances or Error Estimates for Observed Individual Count and Weight.
- Parent Catalog Number for stomach contents and parasites.

#### **Add new schemas**

- Collection metadata – descriptions of vessels, gears ...
- Gazetteer – stratum, ecozone, grid square, stock areas

#### **Enhance end-user interface**

DFO Maritimes routinely provides a variety of publicly available survey based data products (Branton and Black, 2003). The OBIS portal should consider providing a range of mapping products including collection based multi-species mapping and reporting using expanding pie symbol maps (*e.g.* multiple species on one map). Observations for multiple species should be optionally given by row or column, with missing values being given as zeros or nulls. In addition to set by set catch data, OBIS should also provide summary statistics by stratum, ecozones, etc. Methods that enable species catchability standardization should also be investigated.

## Regional Scale DiGIR Services

The Global Biodiversity Information Facility (GBIF) and OBIS networks use the Distributed Generic Information Retrieval (DiGIR) protocol and the Darwin Core (DwC) schema including Darwin Core 2 (DwC2) and OBIS variants. DFO has established a regional-scale DiGIR server at BIO to enable near real-time posting of multiple datasets to the OBIS portal. BIO's DiGIR service sits within a specially controlled portion of the DFO firewall known as the Demilitarized Zone (DMZ) allowing limited connections with a controlled set of known partners. Inputs include small scale specialized databases (e.g. Atlantic Conservation Data Centre) provided via the File Transfer Protocol (FTP) and large institutional scale databases (e.g. DFO Maritime and North East Fisheries Science Center trawl surveys) via the Oracle SQL\*net protocol. The only output from the DiGIR provider service is XML-formatted data sent to OBIS portal's global cache. Data flow into and out of the DiGIR provider service is in the form of pre-scheduled transfers, with all of the public queries handled by the OBIS portal. The OBIS portal manages all movement of data to the GBIF portal.

## Conclusion

The suggested improvements provide a systematic basis for ongoing enhancement and extension of the OBIS schema and interface. Improved ability to integrate data from disparate sampling schemes would in turn provide a capacity to derive population/community indices of abundance, diversity, production, etc. around the world. Trophic cascade models using trawl survey and Continuous Plankton Recorder (CPR) data being developed for the Scotian Shelf (Choi *et al.*, 2004) could, for example, be tested in the North Sea.

## References

- Branton R.M. and J. Black. 2003. Summer Groundfish Survey Update for selected Scotia Fundy Groundfish Stocks. DFO Canadian Science Advisory Secretariat Research Document 2003/089. 60p.
- Brown S.K., R. Mahon, K.C.T. Zwanenburg, K.R. Buja, L.W. Clafin, R.N. O'Boyle, B. Atkinson, M. Sinclair, G. Howell and M.E. Monaco. 1996. East Coast of North America Groundfish: Initial Explorations of Biogeography and Species Assemblages. NOAA, Silver Springs, MD, DFO, Dartmouth, Nova Scotia. 100+ pages.
- Choi J.S., K. Frank, W.C. Leggett and K. Drinkwater. 2004. Transition to an alternate state in a continental shelf ecosystem. *Can J Fish Aquat Sci* 16(4):505-510.
- Costello M.J., C.S. Emblow and B.E. Picton. 1996. Long term trends in the discovery of marine species new to science which occur in Britain and Ireland. *J Mar Biol Ass U K* 76:255-257
- DFO. 2003. State of Eastern Scotian Shelf Ecosystem. DFO Canadian Science Advisory Secretariat Ecosystem Status Report 2003/004.
- Doubleday W.G. and D. Rivard. 1981. Bottom trawl surveys. *Can Spec Publ Fish Aquat Sci* 58: 273p.

Grassle J.F. 2000. The Ocean Biogeographic Informations System (OBIS): an on-line, worldwide atlas for accessing, modeling and mapping marine biological data in a multidimensional context. *Oceanography* 13(3):5-9.