

NSIS AND ATLANTIC ECOSYSTEMS INITIATIVE TO MAKE DATA ACCESSIBLE – ONE DATASET AT A TIME

GEOFF BROWN¹, MARY KENNEDY²,
ANDREW SHERIN³, WARD APPELTANS⁴, ERIC L. MILLS⁵,
LANCE LAVIOLETTE⁶

¹ *Dalhousie University, Halifax, NS*

² *OBIS Canada, Bedford Institute of Oceanography, Dartmouth, NS*

³ *COINAtlantic Secretariat, Dalhousie University, Halifax, NS*

⁴ *Ocean Biogeographic Information System (OBIS),
Intergovernmental Oceanographic Commission of UNESCO,
IOC Project Office for IODE, Oostende, Belgium*

⁵ *286 Kingsburg Road RR #1, Rose Bay, NS*

⁶ *22350 County Road 10 RR#1, Glen Robertson, ON*

There has been considerable discussion in recent years related to open science and improving the transparency and reproducibility of scientific research. Much of this discussion is centered on the need for open access to research data. Like other aspects of open access, some researchers have been slow to respond. This is due, in part, to a lack of digital tools and standards and possibly related to misconceptions around attribution and digital rights management. One bright spot though in the open data movement is in the vibrant community of biodiversity data producers.

Cornerstones of the biodiversity data community are the many natural history museums and scholars with a history of contributing observational data to build larger datasets of species occurrences. Within this community, most data producers are choosing to share data via a growing number of online repositories. The ROpenSci website (github.com/ropensci/spocc) lists ten major repositories including GBIF (Global Biodiversity Information Facility), Berkeley Ecoengine, iNaturalist, VertNet, BISON (Biodiversity Information Serving Our Nation), eBird, AntWeb, iDigBio, OBIS (Ocean Biogeographic Information System) and the Atlas of Living Australia.

* Author to whom correspondence should be addressed: Geoff Brown and Mary Kennedy.
E-mail: geoff.brown@dal.ca or mary.kennedy@dal.ca

Collectively these repositories contain millions of records ready for digital access, discovery and computational processing.

The move to online repositories is also promoted by the many journals that now recommend or require authors to deposit their data in a recognized data repository. As an example, the Canadian Journal of Fisheries and Aquatic Sciences includes in its instructions to authors the following text:

“For primary biodiversity data authors are strongly encouraged to place all species distribution records in a publicly accessible database such as the national Global Biodiversity Information Facility (GBIF) nodes (gbif.org) or data centres endorsed by GBIF, including BioFresh (www2.freshwaterbiodiversity.eu) for freshwater data and the Ocean Biogeographic Information System (OBIS, iobis.org/) for marine biodiversity data, which also holds supporting measurements taken alongside the species occurrence data.” (nrcresearchpress.com/page/cjfas/authors)

Unfortunately, while current online journals promote deposit in purpose built repositories, older print journal articles and their more recent digital surrogates have trapped a lot of occurrence data in a very analog-like digital state. That is to say, the tables and text of these articles aren’t really useful to researchers without significant manual intervention. The result is that we have a significant gap in the availability of historic species occurrence data that could support current and future research and analysis.

In Atlantic Canada, COINAtlantic (Coastal and Ocean Information Network Atlantic, (coinatlantic.ca) is in its second year of the Atlantic Ecosystems Initiative (AEI) project. The purpose is to improve the accessibility of marine species occurrence data including the ‘rescue’ of datasets and make them available to the public, other researchers and managers for reuse. For details, see Sherin *et al.* (2016).

As part of the AEI project, many articles published in the Proceedings of the Nova Scotian Institute of Science (NSIS) were identified as being of interest because they contain species distribution information that includes scientific name and location. The taxonomic and geographic scope of the AEI project covers any taxa observed or collected within the Atlantic region and datasets collected by Atlantic Canadian researchers in other locations.

The focus of the AEI data rescue activity thus far has been to extract content from published articles and create a set of standardized records that can then be shared with the larger biodiversity data community. These records must conform to the Darwin Core body of standards using controlled vocabularies and terms promoted by biological and oceanographic data managers. Like most descriptive metadata standards, Darwin core is intended to facilitate data sharing and discovery. The standard allows biodiversity data managers to move their species occurrence data into shared data repositories where it can be managed alongside similar datasets. Before Darwin Core existed, biodiversity data producers managed their data in silos where it remained largely undiscoverable and unusable by others. Darwin Core has been successful because it is both simple and flexible. Simplicity is achieved by having a limited number of available terms and flexibility is achieved by having no required terms. Wieczorek *et al.* (2012) provide an excellent background on the history and development of Darwin Core but the official terms documentation should be consulted for the most current set of terms (rs.tdwg.org/dwc/terms/).

The AEI rescued datasets are uploaded to the OBIS (Ocean Biogeographic Information System) Canada data repository (ipt.iobis.org/obiscanada/), where they are reviewed for compliance with the Darwin Core standard and then committed to the OBIS Canada repository. Once the data is approved in the OBIS Canada repository, it is made accessible to the public and can be harvested by larger biodiversity aggregators. OBIS Canada is a node in the larger OBIS network (OBIS, iobis.org/contact/). OBIS currently harvests data from the Canadian node once every four months but plans to provide more seamless access in the near future. This work represents a small but significant contribution to national and global initiatives. Not only are gaps in species distribution being filled through these activities but critical temporal coverage is being expanded through the mobilization of content previously published by the Nova Scotian Institute of Science.

The public may query the OBIS database and its collection of integrated datasets through search interfaces on their portal (iobis.org/). All datasets include complete EML metadata (Ecological Metadata Language) with references and links to original source publications. Individual datasets may also be downloaded directly from the IPT (Integrated Publishing Toolkit).

The history and mission of OBIS is very complementary to the work of the COINAtlantic, Atlantic Ecosystems Initiative. OBIS emanated from the Census of Marine Life (2000–2010 and was adopted as a project under IOC-UNESCO’s International Oceanographic Data and Information Exchange (IODE) programme in 2009 (iobis.org/about/). One of the recommended strategies of the Census of Marine Life program (CoML, coml.org/) was to obtain data from museums, oceanographic institutions, universities and commercial companies (Paterson *et al.*, 2000).

Based on this strategy, it is quite possible that the Census of Marine Life (and later OBIS) envisioned the reclamation of data from regional journals such as the NSIS Proceedings. This journal was first published in 1863 and has remained current in both content and format ever since. A complete digital archive of the NSIS Proceedings is maintained in the Dalhousie University institutional repository, DalSpace (dalspace.library.dal.ca/handle/10222/11192). Current issues are available to NSIS members immediately (and to the general public with a six month embargo period) at the NSIS Proceedings website (nsis.chebucto.org/; ojs.library.dal.ca/nsis). In both the archive and on the members’ site, a link is in place to make readers aware when there is supplementary species occurrence data available for the article or special publication.

The first NSIS publication identified during the AEI project was the Birds of Brier Island (dalspace.library.dal.ca/handle/10222/72244). This special publication was originally only available to the public in print for a modest fee and to NSIS members via the NSIS Proceedings website. As part of the data rescue project, a pdf version of this publication is now accessible online to everyone with links to the species occurrence data in the OBIS Canada repository. The first version of species occurrence data associated with this publication contains a small subset of the source material that is essentially a listing of taxa found on the island (ipt.iobis.org/obiscanada/resource?r=brierisland_birds). Observations of 355 bird species for the Brier Island area (www4.rncan.gc.ca/search-place-names/unique/CAFBS) were made accessible to OBIS and will fill gaps in species distribution. A second version of the dataset will include more detailed information such as temporal and spatial distributions enhancing the accessible content.

The ability to revise and refresh content shared with OBIS highlights the advantage of taking an iterative approach to data reclamation.

Depositing even a small amount of previously inaccessible data to a shared repository can have tangible benefits and serve to kick-start further initiatives that enrich data with even more detail. Making data accessible can be done in waves. Perhaps it is appropriate that as an example we chose a dataset from the Bay of Fundy area – this initiative and NSIS are part of a tidal wave of researchers who, in addition to communicating their research via traditional articles and publications, are now making their data accessible to fellow scholars to answer their own research questions. After all, this is the foundation of science research building upon the research that preceded it. The NSIS is keeping pace with other leaders in scholarly communication by embracing technology, open data and a general willingness to share but maintaining a strong connection to the traditional scholarly record.

Williams *et al.* (2011) stated that “biodiversity is essential for the health of the planet and for humans because it underpins ecosystem functions that provide a wide range of goods and services to human societies.” The continued collection of data on the diversity, distribution, and abundance of life in the ocean by OBIS enables ongoing monitoring of decline and/or recovery of biodiversity and ecosystem function against the baseline that was originally established by the Census of Marine Life. Rescuing older species observation data from publications and museum collections expands the geographic, temporal and taxonomic coverage of this initial baseline.

The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES, 2016) emphasized the significance of scenarios and models that forecast changes in ecosystem services. Such forecasts are relevant to public policy and play a vital role in creating more direct connections between ecosystem services and social, economic, and political systems. Rich biodiversity data is required to create and test these models.

The Organization for Economic Co-operation and Development (OECD) Declaration on Access to Research Data from Public Funding recognized that “... international exchange of data, information and knowledge contributes decisively to the advancement of scientific research and innovation; ... that open access to, and unrestricted use of, data promotes scientific progress and facilitates the training of researchers; ... that open access will maximize the value derived from public investments in data collection efforts;” and “... that the

substantial increase in computing capacity enables vast quantities of digital research data from public funding to be put to use for multiple research purposes ... thereby substantially increasing the scope and scale of research” (OECD, 2004). Implicit in the declaration is a recognition of the substantial benefits that science, the economy, and society at large could gain from the opportunities that expanded use of digital data resources like OBIS have to offer.

REFERENCES

- IPBES (Intergovernmental Platform on Biodiversity and Ecosystem Services).** (2016). Summary for policymakers of the methodological assessment of scenarios and models of biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Bonn, Germany.
- OECD (Organisation for Economic Co-operation and Development).** (2004). Declaration on Access to Research Data from Public Funding. acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157. Accessed Feb. 10, 2017.
- Paterson, G., Boxshall, G., Thomson, N. & Hussey, C.** (2000). Where Are All the Data? *Oceanography* 13(3), 21-24.
- Sherin, A., Kennedy, M., & Baccardax Westcott, A.** (2016). Research Data Transparency and Openness for the Proceedings of the Nova Scotian Institute of Science. *Proceedings of the Nova Scotian Institute of Science* 48(2), 169-171.
- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Doring, M., Giovanni, R., Robertson, T. & Vieglais, D.** (2012). Darwin Core: an evolving community-developed biodiversity data standard. *PLOS ONE* 7(1), e29715.
- Williams, M., Mannix, H., Yarincik, K., Miloslavich, P., & Crist, D.T.** (2011). Scientific Results to Support the Sustainable Use and Conservation of Marine Life: A Summary of the Census of Marine Life for Decision Makers. Census of Marine Life International Secretariat, Washington, DC.