Appendix A:

Supplemental Data Processing Steps for Section 7.0: Guide to Submitting Marine biological data to OBIS Canada IPT.

- Review the latest version of DwC terms: http://rs.tdwg.org/dwc/terms.Review list of fields in the source file and map to DwC terms and identify reformatting required.

- Review list of fields in the source file that were not mapped to DwC terms
  - Are these calculated values and redundant? If yes, then ignore.
  - Are measurements beyond species occurrence / abundance associated with the species occurrence (e.g. temperature)? If yes consider adding as measurements using DcW term 'MeasurementName' and measurement(s).
  - Are these measurements associated with the sampling location but not necessarily related to the species occurrence? If yes, consider referencing this data in the metadata and direct readers to the source file should they wish to access this extra information. Add note to the metadata: the original report contained information on various measurements (e.g. salinity).

- Identify DwC fields that can be populated with default values for the resource

- Extract distinct list of taxon names
  - Review the name list.
  - Do the names contain '?' or '*'. If yes determine the meaning of these symbols.
  - If necessary create a new table with the original name and then add columns parsing out various name modifiers.
    - Create column for Original name, EditedScientificname, authorship, lifestage, gender, size, modifier, qualifier, etc.
    - Review, clean and standardize content in the various new fields
  - Add a column for common or vernacular name if included.
    - If only common name provided then consult source or local experts and assign a scientific name.
  - Perform name QC procedures as recommended by OBIS. http://iobis.org/node/47
    - WoRMS Taxonmatch
    - If the name cannot be mapped then contact the OBIS Canada data management team

- Extract distinct list of locality place names
  - Review the place name list.

- If necessary create a new table with the original place name(s) and then add columns parsing out various area names.
  - Separate county, province and country names from the place name.
  - Identify if place name references a sampling station or a defined fishing or marine area or a place name or a specific point location.
  - Are numerous place or area names provided? If yes, are any of these redundant? Example is fishing area within a larger NAFO zone? If yes then it is not necessary to record the NAFO zone as this can be determined from the fishing area location information.
  - Are provided place/area names defined elsewhere? Is it possible to identify and link to a gazetteer containing the area definitions? Is it possible to assign a gazetteer identifier?
  - If area definitions are not contained in an accessible gazetteer and if these areas are in common use for a specific type of data then contact the OBIS Canada data management team.
- Create a new field called VerbatimLocality and populate with standardized formatting for given place name. Example one entry might say 'Salmon River, New Brunswick' and the next might say 'Salmon R., N.B.' Adopt a format and revise content.
- Review decimal latitude and decimal longitude coordinates.
  - Have calculated values resulted in too many decimal places?
  - Determine if any coordinate pairs are null. All records must be georeferenced.
  - Determine if any coordinates pairs = 0,0. This is likely an error.
  - Determine if coordinate pairs are signed correctly for the hemisphere (+N-S; -W+E).

- Extract distinct list of latitude/longitude coordinates.
  - Consider truncating the number of decimal places if there are hundreds of records
  - Plot lat/long and visually confirm that points fall within the expected geographic extent. Verify coordinates if points fall on land or outside sampling area.

- Review measurement/ facts
  - NOTE: it is possible to refresh IPT resources. Consider creating a version 1 of the resource without measurement IF additional time is required to consult experts re the assignment of DwC MeasurementNames and mapping to the data set's associated data.

2. Innovation and scientific research collaboration

In 2007 the Organization for Economic Cooperation and Development adopted a Declaration on Access to Research Data from Public Funding. The motivations for this declaration were to take full advantage of the new opportunities and benefits offered by Information Communication Technologies, to promote improved scientific and social return on the public investments in research data and improve access to, and sharing of data.

Accessibility to research data has become an important condition in:

- The good stewardship of the public investment in factual information;
- The creation of strong value chains of innovation; and
- The enhancement of value from international co-operation.

Furthermore, improved access to, and sharing of, data:

- Reinforces open scientific inquiry;
- Encourages diversity of analysis and opinion;
- Promotes new research;
- Makes possible the testing of new or alternative hypotheses and methods of analysis;
- Supports studies on data collection methods and measurement;
- Facilitates the education of new researchers; and
- Enables the exploration of topics not envisioned by the initial investigators;
- Permits the creation of new data sets when data from multiple sources are combined.

Appendix B:

Data Recovery Description – based upon Elsevier data rescue award criteria: www.elsevier.com/idra2015

A. General Project Information
   1. Date: project started in April of 2015 and ended March 31, 2016
   2. Name, institution and email address of project contact:
      ACZISC Association, 6414 Coburg Road, P.O. Box 15000, Halifax, NS
      The Atlantic Coastal Zone Information Steering Committee (ACZISC) was established in January 1992 to foster cooperation in Atlantic Canada with regard to Integrated Coastal and Ocean Management (ICOM), coastal mapping and geomatics. The mission of the ACZISC as expressed in its strategic plan is "to promote, facilitate and influence information management, policies and programs that enhance integrated coastal and ocean management in Atlantic Canada."
      The plan focused on three priorities:
      - Encouraging action on the implementation of Integrated Coastal and Ocean Management (ICOM) as a tool to realize environmental, economic and social sustainability;
      - Collaborative sharing of data and information; and
      - Encouraging the engagement of organizations in the ACZISC that is reflective of the diversity of the ICOM Community of Practice.

      The ACZISC addresses these priorities by delivering products such as regular meetings and occasional workshops, the monthly Coastal Update e-newsletter, the ICOM information hub at www.coinatlantic.ca, and the development of Coastal and Ocean Information Network (COIN) Atlantic tools to encourage the sharing and accessibility of data and information. It also leads or participates in projects with other organizations that advance the ACZISC priorities
   3. Name of Project:
      Atlantic Canada's Biological Data for Ecosystem Planning and Decision-making: Opening Access and Increasing Reuse
   4. Project URL:
      http://coinatlantic.ca/index.php/atlantic-marine-biological-data-partnership.
   5. Relevant papers related to the Project, with links, if possible
      - NSIS article
      - International Coastal Atlas Network newsletter article: http://www.iode.org/index.php?option=com_content&view=article&id=494:4-april-2016-ican-newsletter-march-2016&catid=23&Itemid=115.
      - ACCESS poster **Engaging in Biological Data Rescue** 12-14 May 2016
      - Coastal Zone Canada 2016 (12-14 June 2016) presentations **Collaborative efforts to train future scientists how to share marine and coastal species**

distribution information and **Opening Access and Increasing Reuse of Atlantic Canada's Biological Data** http://www.czca-azcc.org/html/home.html.

1. Name of Project team members (and email addresses); Institutions of Project team members; Details of any other Project partners
    - The Project lead, Andrew Sherin, Director, ACZISC Secretariat has over 35 years' experience managing people, money and projects completing his career in the federal public service in 2009 as a responsibility centre manager. He has extensive experience in data management and the application of GIS to coastal and estuarine issues. He is a member of the State of the Scotian Shelf Advisory Committee and the Advisory Committee for the National Climate Change Adaptation Coastal Assessment. The ACZISC Secretariat authored the Emerging Issues Theme paper for the State of the Scotian Shelf.
    - Alexi Baccardax Westcott, has been Project Officer for the ACZISC Secretariat since September 2010. Alexi obtained a BSc. in Environmental Science with an Emphasis in Ecology (2006) and a Master of Environmental Studies (2010) from Dalhousie University. A particular focus of her studies and volunteer work during graduate school was Community based Coastal Management in St. Margaret's Bay, Nova Scotia.
    - Mary Kennedy, retired biologist and data base manager and volunteer for DFO, has extensive experience as the regional node manager for OBIS Canada and DFO project lead for the Canadian Register of Marine Species.
    - Huntsman Marine Science Centre staff Dr. Gerhard Pohle (30 years) and Lou Van Guelpen (37 years) participated in the project. They are curators at the Huntsman's Atlantic Reference Centre (ARC), a research museum for Canadian Atlantic Marine life and a centre for biodiversity information and applied environmental research. Pohle and Van Guelpen brought substantial biodiversity and information strengths to the project. They have taken leadership roles in previous biodiversity data sharing initiatives.
    - Angela Douglas, Executive Director of the Southern Gulf of St. Lawrence Coalition on Sustainability, has over ten years' experience working for environmental NGOs in a project management capacity. Angela obtained a BSc. in Biology (2007) and an MSc in biomedical sciences (2010) from UPEI/AVC and has worked as executive director of the Southern Gulf Coalition since spring of 2015.
    - Dr. Bruce Hatcher, Director of the Bras D'Or Institute for Ecosystem Research at Cape Breton University annually conducts data collection in the Bras D'Or lakes valued at over $100,000 and holds 40 years of data including species population data. He has close ties with the Biosphere Reserve Association.
    - Project Partners:
        - The ACZISC Secretariat will provide project management and coordination, execute the communications plan, assist partners in the loading of their data to

OBIS in collaboration with Mary Kennedy, provide COINAtlantic tools, and author the State of Biological Data Accessibility;
- OBIS Project Office will provide data management and technical advice and provide the IT and data base infrastructure;
- the Nova Scotia Museum of Natural History, the Huntsman Marine Science Centre/Atlantic Reference Centre, the Southern Gulf of St. Lawrence Coalition on Sustainability and the Bras D'Or Institute for Ecosystem Research will identify, document and upload species occurrence data to OBIS and assist in the writing of the State of Biological Data Accessibility Report

B. Status

1. Did the project receive funding? If yes, include Agency or funding organization which funded the project (including funding numbers).
   - Funding ($70k) was received from Environment Canada under their Atlantic Ecosystems Initiatives. Planning and decision making: Eligible activity includes: building on existing or the development of new integrated strategies, plans, frameworks and action plans; integration of data/information; knowledge and science – geographic information system (GIS) mapping; assessment of baseline environmental conditions;
   - ACZISC received $20,000 from the Province of Nova Scotia in 2015-2016 which needs to be focussed on the priorities of the Nova Scotia Department of Fisheries and Aquaculture (NSDFA) and the Province. This project contributes to the mandate of NSDFA and 50 % of the funding will be used to support the project.
   - In addition, the data holding organizations have made commitments to provide staff time to prepare and upload their data to OBIS.
   - The in-kind support provided by the ACZISC includes office space and IT infrastructure provided to the ACZISC by Dalhousie University and volunteer support provided by Mary Kennedy (OBIS Canada expert). All participating data holding organizations have committed to providing in-kind support.
   - The ACZISC Secretariat received funding from the Canada Summer Jobs program ($13,911) from the program at four sites of the participating data holding organizations (ARC, NSM, Southern Gulf Coalition and CBU) and hired students to assist with the loading of data to OBIS

C. Description

1. Goal of the project

   "To gather, document, track, and upload estuarine biological data for Atlantic Canada to Ocean Biogeographic Information System (OBIS) with a focus on data that will support AEI program priorities in habitat and biodiversity and ecosystem based management in Atlantic Canada."

The purpose of this project was to establish a partnership of organizations holding estuarine biological data committed to sharing their data using OBIS. The partnership will be an ongoing collaborative to support future uses of the data to support AEI and other ecosystem objectives. It established a better understanding of what data has already been collected; made datasets accessible through the Ocean Biogeographic Information System (OBIS); to develop customized geospatial products and services to contribute to the State of Biological Data Accessibility Report.

- Within the context of this project the resulting Atlantic Marine Biological Data Partnership resulted in a number of new quality assured data sets uploaded to OBIS.
- A literature search project resulted in a list of more than 130 publications containing information of potential interest to OBIS Canada
- A database was designed to organize publication references and to track data rescue activities. This included the development of a vocabulary for dataset status.
- Training of students and volunteers in data rescue and data standardization activities.
- The project identified additional organizations who will be invited to join in the Atlantic Marine Biological Data Partnership in future years.
- The project developed a case study for Saint John Harbour demonstrating how to use OBIS content and other data for projects in ecosystem based management.
- A State of Biological Data Accessibility Report based upon a data accessibility assessment of baseline species occurrence information in estuaries of the Bay of Fundy and Bras D'Or Lake using existing estuarine biological data sets and existing data management infrastructure was prepared.

2. Size and scope of user community
   Georeferenced species occurrence information is of interest to ocean and coastal zone management and biodiversity groups. OBIS Canada content will be shared with the Global Biodiversity Information Facility (GBIF) Canadian node. CBIF describes size by how much data they serve = hundreds of thousands on CBIF and millions on Canadensys. CBIF is the national platform (i.e., all departments with a biodiversity mandate) and the GBIF node so that would encompass potentially all Canadian content. Scope is all of life but of course there is a bias to macro-life. The users are all Canadians but from a research perspective it is international.

3. Size and topic of the dataset
   Historical and current surveys observations and collected specimens from the Atlantic region of Canada with a focus on coastal and marine species distribution. Individual datasets may have contained as few as a couple of records while other aggregated datasets contained up to 500,000+ records.
   The focus of the AEI project was on habitat and biodiversity with emphasis on significant habitats such as estuaries and eelgrass beds.

4. What type of metadata was added to the dataset?
   Dataset description; contact info for contact, resource creator, metadata authors and associated parties; description of geographic, temporal and taxonomic extents; dataset citation and associated references; description of sampling methods and QC as well as data processing steps.
5. What standards were used if any
   i. The metadata meet the current OBIS standards, **Ecological Metadata Language** (EML)
   ii. Scientific names were mapped to LSID with preference to use of the World Register of Marine Species (WoRMS) as the standard.
   iii. Geographic place names when relevant were standardized to the Canadian Geonames Database (CGNDB) or to the Marine Regions gazetteer (MRG)
   iv. Dataset information, facts and measurements were mapped to Darwin Core (DwC) terms
   v. Measurement vocabularies – measurement names were standardized whenever possible to terms included in the BODC/Seadatanet parameter list and or the WoRMS species traits project.
6. Data annotation process
   i. The data annotation process involved manually converting analogue text and/or tabular data to digital spreadsheets which were subsequently standardized and reformatted into DwC data records. In addition, it may have been necessary to georeferenced sampling locations based on original maps and/or location place names as described in the original paper version of the dataset.
7. Description of the data annotation process
   i. Identify data of relevance to OBIS
   ii. Determine if data is currently digitized and accessible.
   iii. Contact the source to determine if dataset is the master copy; if shared and various versions exit and if more data exists.
   iv. Identify the rights holder and the license to be associated with the dataset
   v. Compile metadata associated with the dataset
   vi. Digitize the source data records
   vii. Georeferenced sampling locations following a defined protocol
   viii. Reformat, standardize and QC the data records
   ix. Create DwC dataset
   x. Upload to IPT (a publically accessible site) and share with iOBIS
   xi. Summarize project, identify successes, describe issues and lessons learned, make recommendations.
8. Example of data before the rescue process

Figure B1: (left) Grey literature collection at Cape Breton University, (right) Examples of analog data.

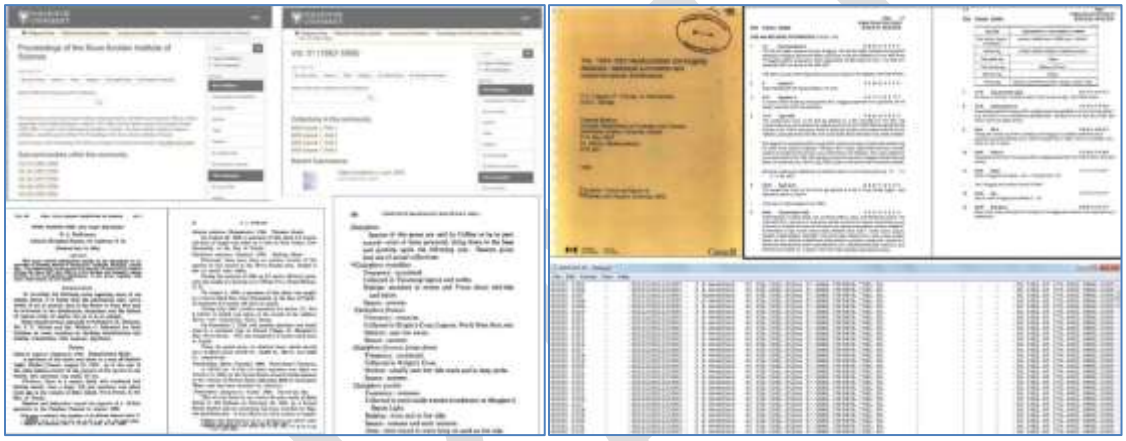
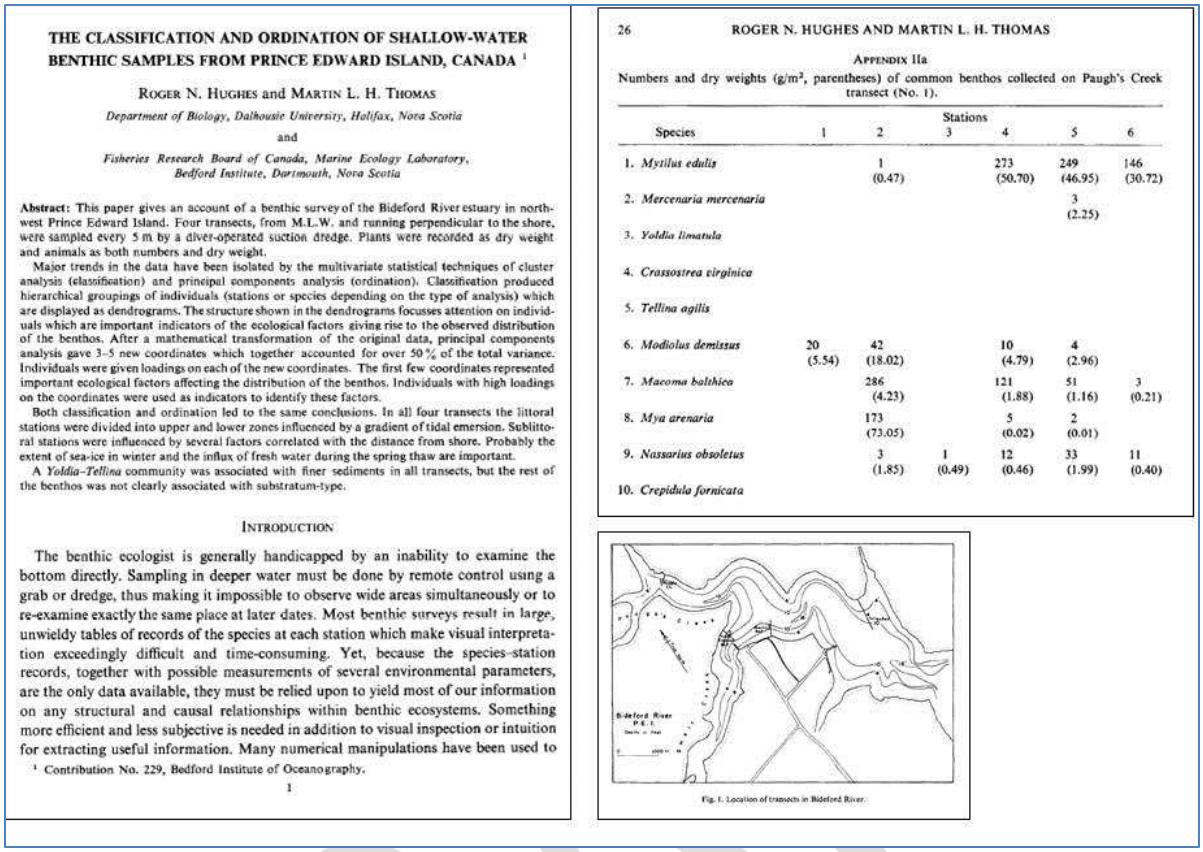


Figure B2: (left) An example from the Journal of the Nova Scotia Institute of Science, (right) An example from the Canadian Technical Report on Fisheries and Aquatic Sciences.



THE CLASSIFICATION AND ORDINATION OF SHALLOW-WATER BENTHIC SAMPLES FROM PRINCE EDWARD ISLAND, CANADA [1]

ROGER N. HUGHES and MARTIN L. H. THOMAS

Department of Biology, Dalhousie University, Halifax, Nova Scotia

and

Fisheries Research Board of Canada, Marine Ecology Laboratory, Bedford Institute, Dartmouth, Nova Scotia

Abstract: This paper gives an account of a benthic survey of the Bideford River estuary in north-west Prince Edward Island. Four transects, from M.L.W. and running perpendicular to the shore, were sampled every 5 m by a diver-operated suction dredge. Plants were recorded as dry weight and animals as both numbers and dry weight.

Major trends in the data have been isolated by the multivariate statistical techniques of cluster analysis (classification) and principal components analysis (ordination). Classification produced hierarchical groupings of individuals (stations or species depending on the type of analysis) which are displayed as dendrograms. The structure shown in the dendrograms focusses attention on individuals which are important indicators of the ecological factors giving rise to the observed distribution of the benthos. After a mathematical transformation of the original data, principal components analysis gave 3–5 new coordinates which together accounted for over 50 % of the total variance. Individuals were given loadings on each of the new coordinates. The first few coordinates represented important ecological factors affecting the distribution of the benthos. Individuals with high loadings on the coordinates were used as indicators to identify these factors.

Both classification and ordination led to the same conclusions. In all four transects the littoral stations were divided into upper and lower zones influenced by a gradient of tidal emersion. Sublittoral stations were influenced by several factors correlated with the distance from shore. Probably the extent of sea-ice in winter and the influx of fresh water during the spring thaw are important.

A *Yoldia–Tellina* community was associated with finer sediments in all transects, but the rest of the benthos was not clearly associated with substratum-type.

INTRODUCTION

The benthic ecologist is generally handicapped by an inability to examine the bottom directly. Sampling in deeper water must be done by remote control using a grab or dredge, thus making it impossible to observe wide areas simultaneously or to re-examine exactly the same place at later dates. Most benthic surveys result in large, unwieldy tables of records of the species at each station which make visual interpretation exceedingly difficult and time-consuming. Yet, because the species–station records, together with possible measurements of several environmental parameters, are the only data available, they must be relied upon to yield most of our information on any structural and causal relationships within benthic ecosystems. Something more efficient and less subjective is needed in addition to visual inspection or intuition for extracting useful information. Many numerical manipulations have been used to

[1] Contribution No. 229, Bedford Institute of Oceanography.

1

26 ROGER N. HUGHES AND MARTIN L. H. THOMAS

APPENDIX IIa

Numbers and dry weights (g/m², parentheses) of common benthos collected on Paugh's Creek transect (No. 1).

Species	Stations 1	2	3	4	5	6
1. *Mytilus edulis*		1 (0.47)		273 (50.70)	249 (46.95)	146 (30.72)
2. *Mercenaria mercenaria*					3 (2.25)	
3. *Yoldia limatula*						
4. *Crassostrea virginica*						
5. *Tellina agilis*						
6. *Modiolus demissus*	20 (5.54)	42 (18.02)		10 (4.79)	4 (2.96)	
7. *Macoma balthica*		286 (4.23)		121 (1.88)	51 (1.16)	3 (0.21)
8. *Mya arenaria*		173 (73.05)		5 (0.02)	2 (0.01)	
9. *Nassarius obsoletus*		3 (1.85)	1 (0.49)	12 (0.46)	33 (1.99)	11 (0.40)
10. *Crepidula fornicata*						

Fig. 1. Location of transects in Bideford River.

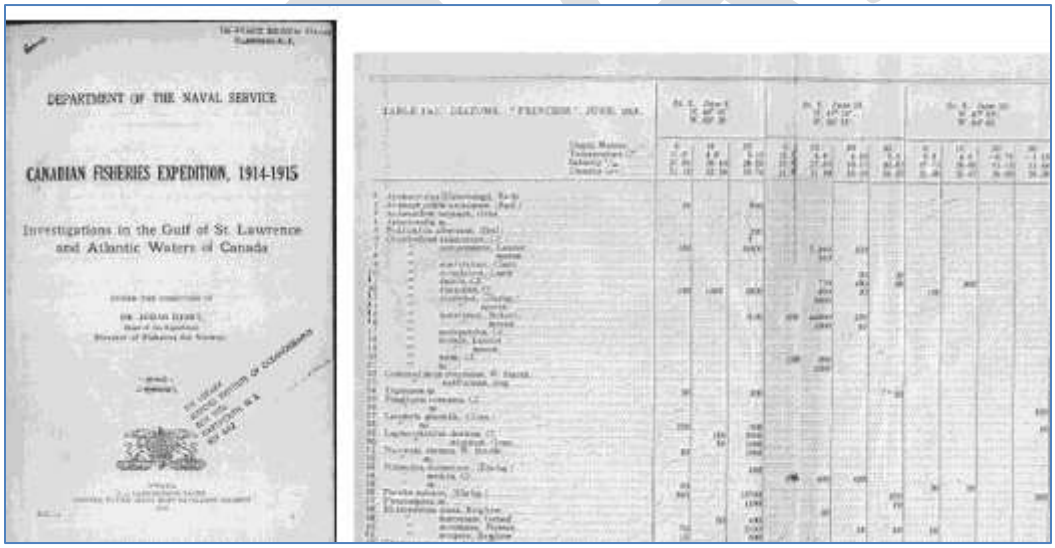


Figure B3: An early (1914-1915) example of data from a report for the Gulf of St. Lawrence.

Formats in the original journal articles were in a variety of format including in line text; tables (multiple formats), card formatted data files, maps, EXCEL files & Word tables.

9. Example of data after the rescue process

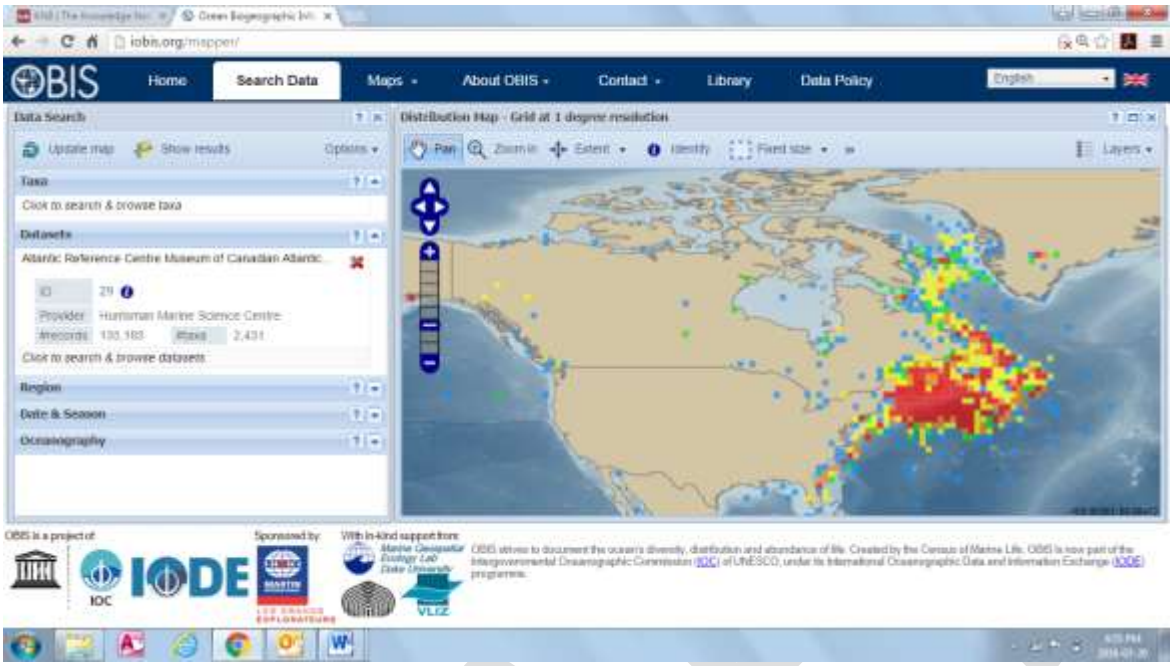


Figure B4: Heat map (colour coded by number of records in a 1 degree cell) generated by OBIS mapper application for all the data submitted to OBIS from the Atlantic Reference Centre, Huntsman Marine Science Centre, St. Andrew's NB.

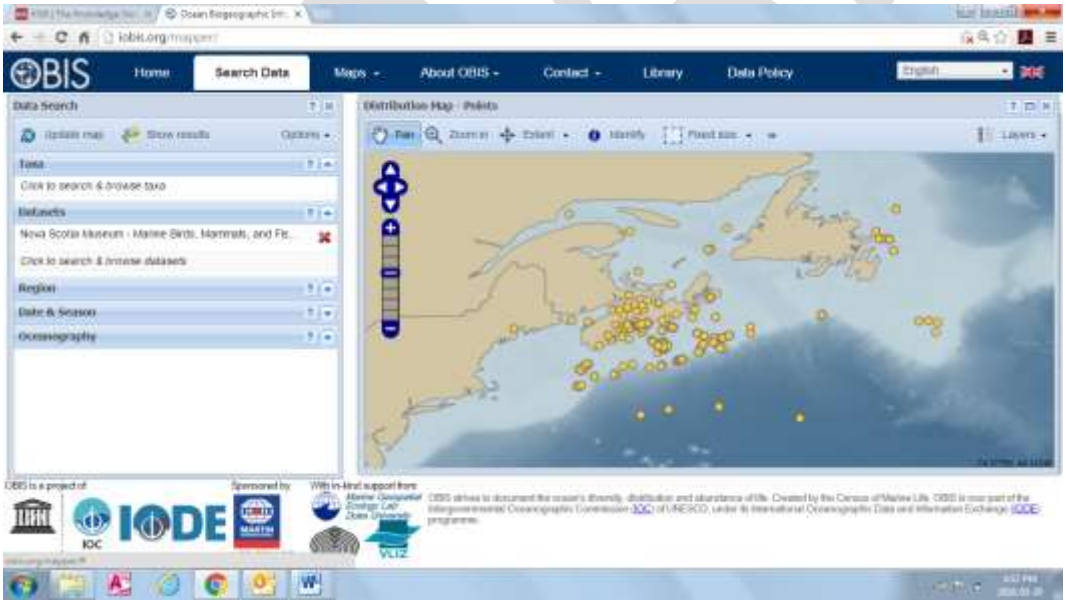


Figure B5: Marine birds, mammals, and fish records added or updated and made accessible through OBIS from the Nova Scotia Museum of Natural History (NSM) collections.

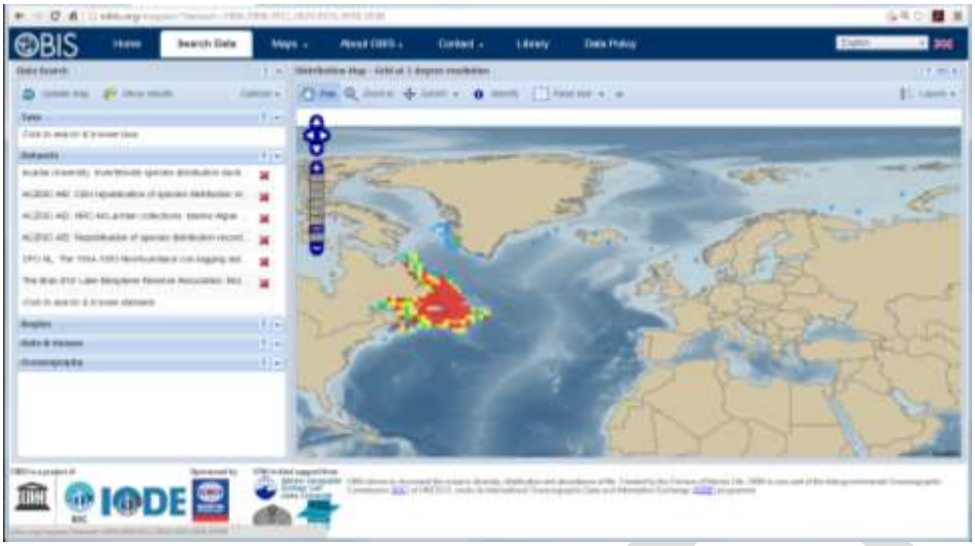


Figure B6: Heat map (colour coded by number of records in a 1 degree cell) generated by OBIS mapper application of the six new resources appended to OBIS as part of the Project as of January 2016. To generate this map follow this link: (http://iobis.org/mapper/?dataset=3906,3909,3911,3929,3935,3938,3938).

OBIS harvests content from the OBIS Canada IPT site on a regular schedule (approximately every 3-4 months). Users may extract content from the OBIS portal, from its geoserver or may register and obtain permission to run ad hoc queries against the OBIS postgres database.

Figure B7: An example of a metadata record publically accessible from the OBIS Canada IPT (data repository)

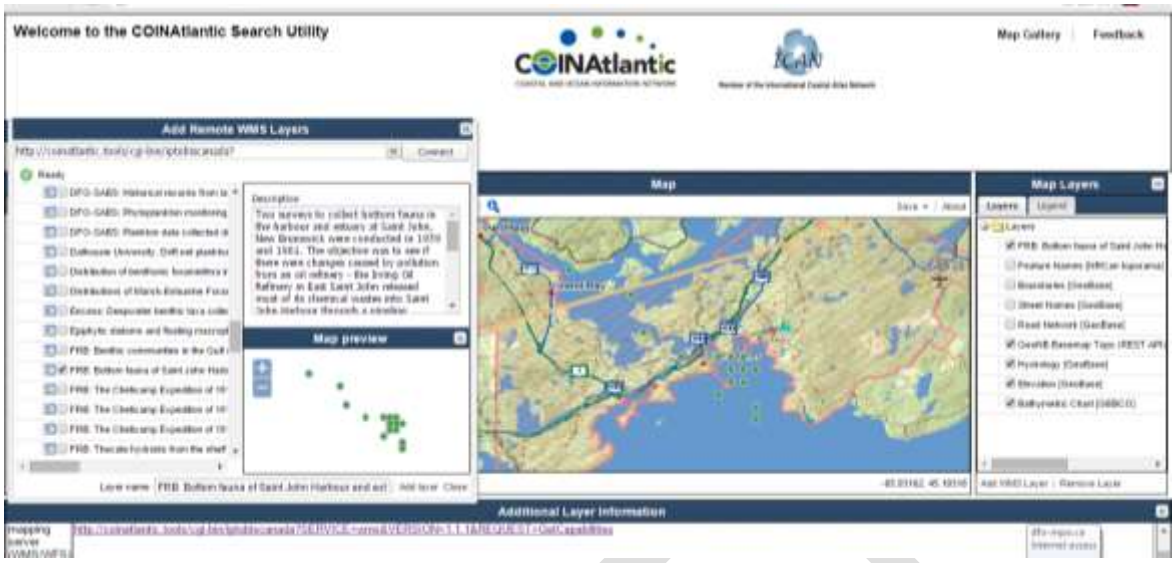


Figure B8: Saint John Harbour bottom fauna data set displayed in the COINAtlantic CSU using the web mapping service generated from content on the OBIS Canada IPT.

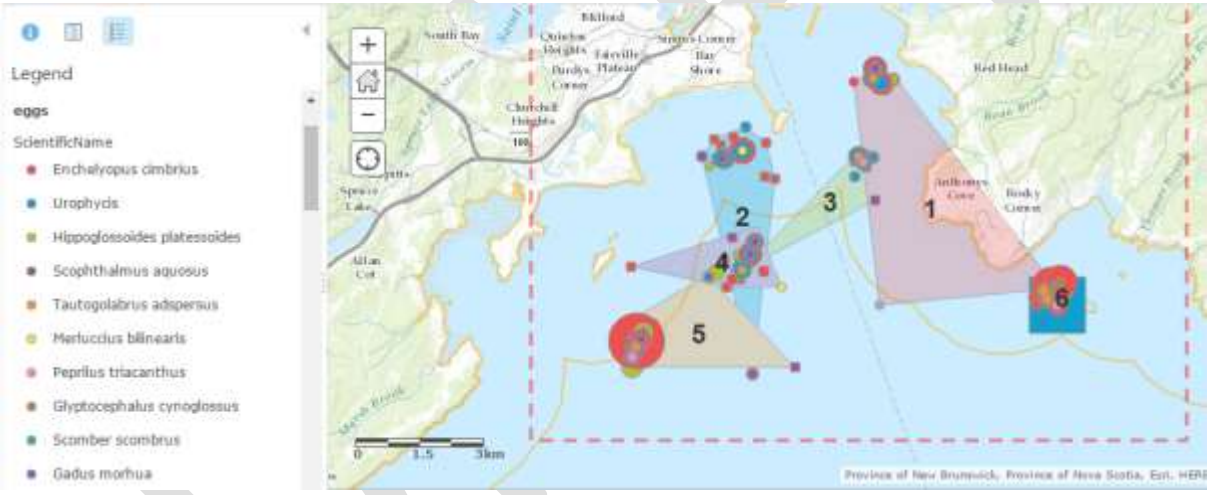


Figure B9: Example of an initial Story Map for Saint John Harbour showing species distributions and abundances (size of circles and squares) for eggs and larvae. Data set kindly provided by Atlantic Reference Centre (in press).

D. Motivation

1. A description of ways in which the data could be or is used after being made available electronically:
   private version can be shared on OBIS Canada IPT; once public version published on the IPT is accessible to all then accessed by public; once the public version has been harvested by iOBIS it is appended to the iOBIS dbase and content displayed on portal; accessible by public and can be used by others and ARCGIS products.
   i. access DwC package from IPT site and write scripts to process datasets and or create story maps – see arcgis online products
   ii. IPT resources accessible on ACZISC CSU. Can overlay layers

iii. download records from OBIS web portal
iv. run ad hoc queries directly against the OBIS database.

2. Data repository or website used to make the data available to the wider user community
   i. Discoverable through a CGG entry and a CSU search http://coinatlantic.tools/
   ii. Links on coinatlantic.ca website http://coinatlantic.ca/
   iii. Direct access on OBIS Canada IPT http://ipt.vliz.be/obiscanada/
   iv. iOBIS portal http://iobis.org/
   v. registered IPT resources will be accessible through GBIF http://www.gbif.org/
   vi. llinks to data repositories promoted by scientific journals
   vii. Accessible by request thru DFO archives

3. Please explain why you chose to rescue this data set, in particular
   Prior to this project the contributions to OBIS were focussed on a few data holders (e.g. DFO). The project wanted to fill the gaps in temporal, geographic and taxonomic extents of the data accessible through OBIS to include data sets from other data holders in Atlantic Canada. The project reached out to the community and formed partnerships to contribute these datasets including appropriate biological data from the Journal for the Nova Scotia Institute of Science, museums, academic institutions and non-governmental organizations.

4. Description of how was, is being used
   All of the data submitted to OBIS was acquired from existing research or held in existing collections. Historically users had to contact multiple sources to obtain data of interest and develop procedure to integrate content. With the submission of standardized content to OBIS and inclusion into an aggregated/standardized database the OBIS portal provides users with easy access to datasets from multiple sources in a standardized interoperable format. Content is refreshed on a regular basis which means that data is not orphaned but rather all edits and/or additions are accessible.

5. Plans to extend the project including long-term sustainability
   It is anticipated that the project will be extended for another three years with the long term objective of enabling biological data holders in control of uploading their own data. In addition, the project will train earlier career scientists in the value associated with sharing data in a standardized and well documented manner such as supported by OBIS so it will become a regular part of their future research projects.

6. Possible extensions to other fields, datasets, etc.
   The project will make new estuarine biological data sets accessible through OBIS as well as new related geospatial products and services to support a wide range of environmental objectives and focus on the objectives of the AEI program. It will

establish a mechanism to enable new data sets to become accessible in the future. The project will provide baseline assessments for selected areas based upon species occurrence data.

The OBIS schema is expanding to include not just presence and abundance information but also to include other biological and environmental measurements.

E. Other – please refer to the lessons learned section (8.2) in this report.