

# COINAtlantic Workshop

ACCESS conference

May 17, 2017

McGill's Geographic Information Centre (GIC), 5<sup>th</sup> floor of Burside Hall

Mary Kennedy (OBIS Canada) and Andrew Sherin (COINAtlantic)

[obiscanada2017@gmail.com](mailto:obiscanada2017@gmail.com) or [a.sherin@dal.ca](mailto:a.sherin@dal.ca)



# Data Management for Research with Special Focus on Species Occurrence Data

**Objective:** The aim of this workshop is to promote best practices in data management to facilitate data accessibility and integration with habitat measurements and to connect the Atlantic Canadian coastal and estuarine biological research community to the Ocean Biogeographic Information System (OBIS).

**Relevance:** Any researcher, student, citizen scientist, community group, staff or volunteer who has collected or plans to collect data (or works with an existing dataset) that includes georeferenced observations and wants to ensure their data is eventually accessible for the benefit of the wider scientific community and support national and global biodiversity objectives such as those set by the United Nations Convention on Biological Diversity. Although the focus is on biodiversity the data management procedures described are applicable to any coastal and estuarine dataset.

*Participants were encouraged to bring some of their own data to be used in the hands on sessions.*

# Workshop Schedule

- 09:00-09:10 Introductions
- 09:10-09:30 Module 1: Making research data accessible
- 09:30-09:45 Module 2: Introduction to OBIS
- 09:45-10:00 Module 3: Introduction to the standards used by OBIS
- 10:15-10:30 Break
- 10:30-11:00 Module 4: Map your dataset content to Darwin Core terms
- 11:00-11:30 Module 5: Clean and reformat dataset content
- 11:30-12:00 Module 6: Standardizing species lists
- 12:00-13:00 Lunch on your own
- 13:00-15:00 Module 7: Georeferencing observations
- 15:00-15:15 Break
- 15:15-16:00 Module 8: Metadata (data required to properly interpret a dataset and facilitate reuse)
- 16:00-16:30 Module 9: 'Data processing' of your datasets
- 16:30-17:00 Discussion

# Introductions

**Relevance:** Any researcher, student, citizen scientist, community group, staff or volunteer who has collected or plans to collect data (or works with an existing dataset) that includes georeferenced observations and wants to ensure their data is eventually accessible for the benefit of the wider scientific community and support national and global biodiversity objectives such as those set by the United Nations Convention on Biological Diversity. Although the focus is on biodiversity the data management procedures described are applicable to any coastal and estuarine dataset.

Round table – pre-workshop survey  
Who are you? (affiliation and role)  
Do you collect or use data?  
Did you bring a dataset?  
Do you have or know of many more datasets?

*Participants were encouraged to bring some of their own data to be used in the hands on sessions*

*Roles: Students, professors, project leaders and participants, data managers, publishers, ecosystem managers, ...*

*Needs & types of data:*

*Research project data*

*Thesis data*

*Info to incorporate into lectures*

*Info to know BEFORE project or thesis starts*

# Introductions

Round table introductions



Question: Who are you????

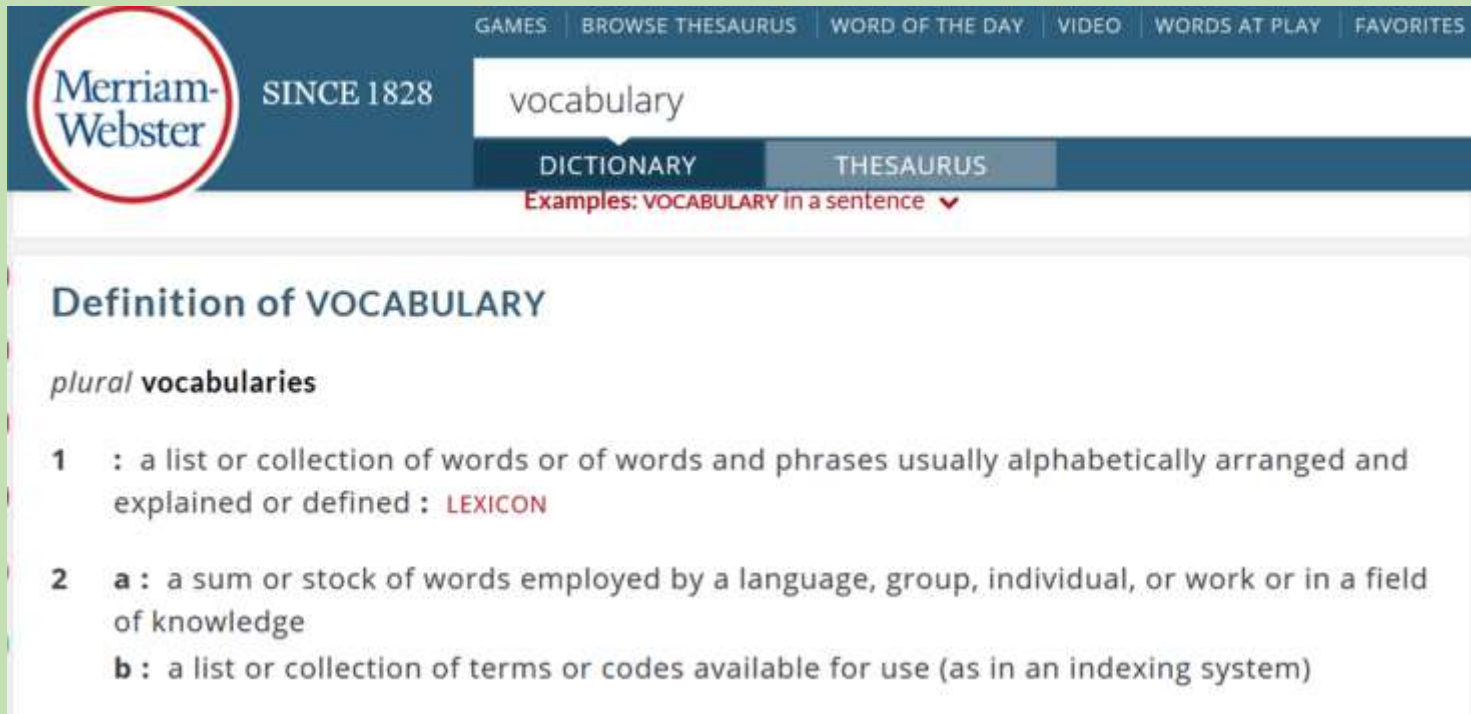
# Introductions



Full Name	Affiliation	City	Province	Status
Gail Chmura	McGill	Montreal	Quebec	I am not a student
Heather Hunt	UNB-SJ	Saint John	NB	I am not a student
Dr. Timothy Rawlings	CBU	Sydney	NS	I am not a student
Matthew Penney	CBU	Sydney	Nova Scotia	Undergraduate Student
Allen Beck	UNB	Fredericton	New Brunswick	Graduate Student

# Introductions

More introductions....



The screenshot shows the Merriam-Webster website interface. At the top, there is a navigation bar with links for GAMES, BROWSE THESAURUS, WORD OF THE DAY, VIDEO, WORDS AT PLAY, and FAVORITES. The Merriam-Webster logo and 'SINCE 1828' are on the left. A search bar contains the word 'vocabulary'. Below the search bar, there are tabs for 'DICTIONARY' and 'THESAURUS'. A dropdown menu shows 'Examples: VOCABULARY in a sentence'. The main content area is titled 'Definition of VOCABULARY' and lists two definitions:

**Definition of VOCABULARY**

*plural vocabularies*

- 1** : a list or collection of words or of words and phrases usually alphabetically arranged and explained or defined : **LEXICON**
- 2** **a** : a sum or stock of words employed by a language, group, individual, or work or in a field of knowledge  
**b** : a list or collection of terms or codes available for use (as in an indexing system)



# Module 1 – Making research data accessible

## Basic outline

- Why share (research) data?
  - Excuses made in the past...
  - Lots of papers discussing the issue..
- Not my data ... questions about who holds the rights to the data
- It is your data or your groups data
  - Do you know what your data policy is?
  - Do you have a data policy?
- Are you willing to share but need training and/or assistance?
- Are you sharing data with other groups or using a data repository?



# Module 1 – Making research data accessible

## Excuses used in the past

- People will **copy my work** from the web and plagiarise it
- Where can one **publish** data? (journals will not publish primary raw data)
- It is **my data**, why should I make it available?
- The data I used was not my own and I did not get **permission** to publish it.
- If I release data then I may be **scooped**.
- I have **not finished analysing** the data. I may do further analysis on the data.
- Somebody will use my data and **benefit from such use**, and worse still, they may be a commercial organisation or consultant.
- The publisher may **profit**.
- I fear that the data will be used for an **incorrect purpose**.
- I do not have the **skills** to publish data on the internet
- IPR related to data and databases **differs between countries**
- I will not get **due recognition** for creating the data.
- Other reasons

*From Mark Costello*



Insert slides from CZC pre

# Module 1 – Making research data accessible

Reading list...

- <https://academic.oup.com/bioscience/article/59/5/418/297578/Motivating-Online-Publication-of-Data>
- <https://www.elsevier.com/authors-update/story/access-to-research/encouraging-and-supporting-researchers-to-share-research-data>

Add cover of pa



**Motivating Online Publication of Data** 

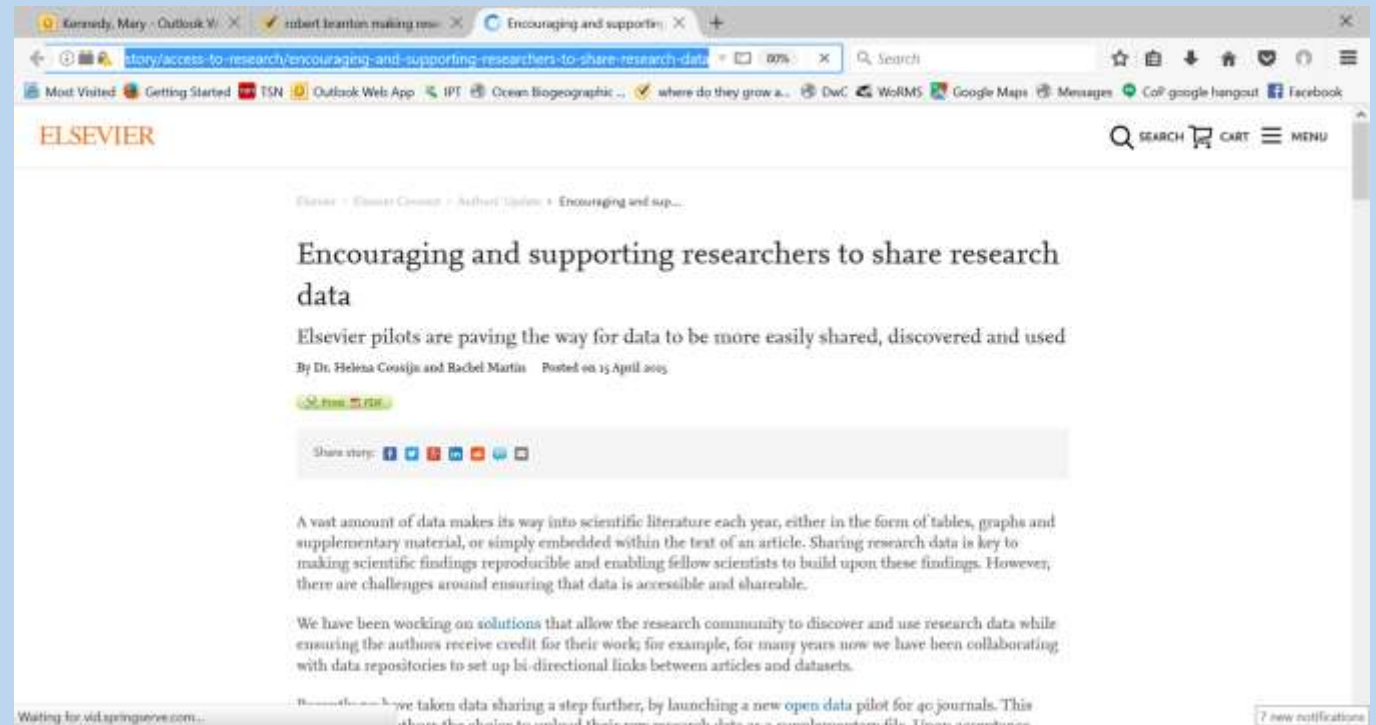
Mark J. Costello

BioScience (2009) 59 (5): 418-427. DOI: <https://doi.org/10.1525/bio.2009.59.5.9>  
Published: 01 May 2009

**Abstract**

Despite policies and calls for scientists to make data available, this is not happening for most environmental- and biodiversity-related data because scientists' concerns about these efforts have not been answered and initiatives to motivate scientists to comply have been inadequate. Many of the issues regarding data availability can be addressed if the principles of "publication" rather than "sharing" are applied. However, online data publication systems also need to develop mechanisms for data citation and indices of data access comparable to those for citation systems in print journals.

**Keywords:** data availability, online publication, environment, citation indices, biodiversity informatics




ELSEVIER

Encouraging and supporting researchers to share research data

Elsevier pilots are paving the way for data to be more easily shared, discovered and used

By Dr. Helena Coussijn and Rachel Martin | Posted on 15 April 2009

Share story: 

A vast amount of data makes its way into scientific literature each year, either in the form of tables, graphs and supplementary material, or simply embedded within the text of an article. Sharing research data is key to making scientific findings reproducible and enabling fellow scientists to build upon these findings. However, there are challenges around ensuring that data is accessible and shareable.

We have been working on solutions that allow the research community to discover and use research data while ensuring the authors receive credit for their work; for example, for many years now we have been collaborating with data repositories to set up bi-directional links between articles and datasets.

...we taken data sharing a step further, by launching a new open data pilot for 40 journals. This offers the choice to upload their raw research data as a supplementary file. Upon acceptance

7 new notifications

# Module 1 – Making research data accessible



If we all agree that research data should be accessible then the questions become:

- How, where and in what format?
- Are there guidelines that can be followed?
- Are there groups that can assist and/or provide training?
- What does this cost as I have no funding?

# Module 1 – Making research data accessible



The screenshot shows the Canadensys website header with a red background and white text. The navigation menu includes 'about', 'digitization', 'data publication', and 'activities'. The main content area is titled 'Norms for data use and publication' and lists several key principles: 'Give credit where credit is due', 'Be responsible', 'Share knowledge', 'Respect the data license', 'Data publication conditions', 'Preferred citations', and 'Acknowledging funding agency'. A paragraph explains that these norms are not a legal document but help build a community around biodiversity data. A section titled 'Give credit where credit is due' provides more detail on citation practices.

<http://www.canadensys.net/about/norm>

S

## Be responsible

Use the data responsibly. The data are published to allow anyone to better study and understand the world around us, so please do not use the data in any way that is unlawful, harmful or misleading. Understand that the data are subject to change, errors and sampling bias. Protect the reputation of the data publisher and clearly indicate any changes you may have made to the data.

## Share knowledge

Let us know if you have used the data. It helps our participants to showcase their efforts and it helps you reach a wider audience. You can [contact us](#) or share your work via [our public forum](#). Inform the data publisher(s) if you have comments about the data, notice errors, or want more information. Their contact details are included in the dataset metadata and on [our repository](#).

## Respect the data license

Understand and respect the data licence or waiver under which the data are published. It is indicated in the *"license"* field of every record and in the dataset metadata. To help you [make greater use of the data](#), most of our participants have dedicated their data to the [public domain \(CC0\)](#). Do not remove the public domain mark or provide misleading information about the copyright status.

## Data publication conditions

We care about data and we just want to make sure you do too. In order to [publish your data](#) through the Canadensys network, you should meet the following criteria:

- You are associated with a Canadian collection or organization.
- You are publishing a specimen or observation dataset, a taxonomic checklist, or metadata about one of these (i.e., one of the 3 types of datasets supported by the IPT).

# Module 1 – Making research data accessible

## Awareness issues – Rights holders

Who holds the rights to the data?

- *I can't share because it isn't my data*

**Analysis of Nutrient Levels in Canadian Coastal Waters  
and  
A Case Study on the Influence of Agricultural Activity on  
Nutrient Concentrations in Prince Edward Island**

Prepared by  
Michel Brylinsky  
Tanya Bryan  
and  
Scott Ryan

Acadia University  
Wolfville, Nova Scotia

for  
National Guidelines and Standards Office  
Existing Substances  
Environment Canada  
Gatineau, Quebec K1A 0H3

April 2006

Publication No. 83 of the Acadia Centre for Estuarine Research

### Need a FAQ section

Funding someone isn't the same as employing someone. If EC was just funding the work they would not be the copyright holder by default. They could have had an agreement to get copyright but it wouldn't be theirs without that agreement.

Similarly, if ACER is not employing those research associates, then ACER would not be the copyright holder by default.


If Acadia employs those researchers, and they have simply made the choice to affiliate with the centre who in turn received some funding to support the work ... then the researchers would be the copyright holder (unless they have signed a copyright transfer agreement with ACER or EC). It really depends on who employs the research associates.

# Module 1 – Making research data accessible

- What is a data policy?
- A document that clearly lists what you can do...
- Data is restricted and cannot be shared (species at risk, commercial/fishing). But the metadata describing the dataset can be shared
- Data is embargoed for defined period. But this ban can be lifted by the PI

# Module 1 – Making research data accessible

Ocean Tracking Network (OTN) data policy. PDF document accessible online

  
**OTN Metadata and Data Policy Highlights**

*OTN's data procedures and policies have, from the inception of the platform, been aligned with international standards and best practices and designed to meet the conditions of national and international funding agencies. In order to remain current, OTN must periodically review its data management, and update procedures and policies to meet evolving national and international standards and practices.*

*OTN's 2008 data policy was modeled on the international standards presented in OECD's 2006 Principles and Guidelines for Access to Research Data. Much has changed since 2008, especially with regards to the evolution of national and international standards for data sharing. In 2011 the Canadian Research Data Summit resulted in a unified position of Canadian funding agencies—including OTN's principal funding agencies CFI, NSERC, and SSHRC—to require that all data from publicly funded research be made openly available in a timely manner. Meeting this standard will be a condition for all grantees to maintain funding from these agencies. It is anticipated that this standard will be implemented at some point within the next year. However, it is also anticipated that there may be exceptions provided in some circumstances to the policy that permit investigators to restrict access to data for limited periods. Examples from the OTN context could include not reporting location information on an endangered species that network investigators are tracking to protect the animals from illegal harvesting, and protecting the thesis data for Highly Qualified Personnel who are in training.*

*Rapid growth in the OTN network and in the demand for OTN data services have greatly increased the volume of data collected. This requires the consideration of automation of data harvesting and the efficiency of automated data posting. Restricting access to data in these circumstances, as is presently done for limited periods for certain categories of OTN data, becomes problematic and expensive.*

*The assurance of Quality Assessment/Quality Control (QA/QC) for the global data feeding into the OTN requires the creation/operation of regional nodes. In order to provide guidance to the development of these regional data centers, OTN's data services, procedures, and policies must be kept current.*

*In the face of these changes, the OTN is now proposing a series of modifications to its current data policy. These are:*

1. Publicly funded (e.g. OTNGlobal or OTNCanada work supported by CFI, NSERC, SSHRC, and Dalhousie funding) deployment, tracking and data collaborators must accept the terms of the most current OTN data policy and provide in a timely fashion all required metadata and data forms to the OTN Data Centre (OTNDC). Collaborators who purchase instruments from VEMCO must also send both VEMCO and OTNDC a VEMCO equipment authorization form so that VEMCO may provide instrument, particularly tag specifications and metadata, directly to the OTNDC. If trackers are using another equipment or tag manufacturer, please contact the OTNDC.
2. OTNDC provides free and open access to all discovery, deployment, and release metadata, as well as detection data to scientists and members of the public who register to use it via the [OTN website](#). Exceptions to this rule are:
  - a. Tag IDs where OTNDC has received manufacturers' and trackers' metadata will be subjected to a two-year renewable embargo (two years after tag life expiration)
  - b. Where the OTN project coordinator has received proof of a scientific license for work on endangered species, release metadata and tag IDs will be subjected to a maximum ten-year renewable embargo\*  
\*Embargoed data can be obtained by contacting the principal investigator, or the embargo can be shortened at the request of the investigator
3. Metadata reports and tag detection histories (tracks and/or detections) will be sent via email to owners of identified tags whenever they are detected as part of routine processing by OTNDC. Real-time and or near real-time deployment operators (e.g. buoys, gliders) may submit detections to OTNDC and so trigger similar notification.

4. If a regional node (e.g. NEPacific, SAfrica = Windian + SEAtlantic, Australia = Indian + SWPacific, GULF) is required, then the node operator(s) must:
  - a. Provide their own data storage equipment and use a remotely managed copy of the Dalhousie-based software system (Red Hat Package Manager), or alternatively use a virtual system in the Dalhousie Computer Center
  - b. Identify the regional data administrator who provides local QA/QC and coordinates/collaborates with the OTNDC staff
  - c. Enable the timely extraction and rollout of data in the regional node by OTN Global initiated data processing procedures
5. By accessing or using OTN Data, all users agree to:
  - a. Give proper attribution to all data providers and to OTN using the preformed citations contained in OTN metadata reports and data records
  - b. Inform [OTN](#) of publications, data products (e.g. tables, graphs, maps, etc.), and commercial applications resulting from use of the OTN data
  - c. Acknowledge that neither OTN nor the provider is liable for inaccuracies in the data
  - d. Assume responsibility for investigating and understanding any limitations of the data
  - e. Report all problems with respect to data to [otnrc@dal.ca](mailto:otnrc@dal.ca)

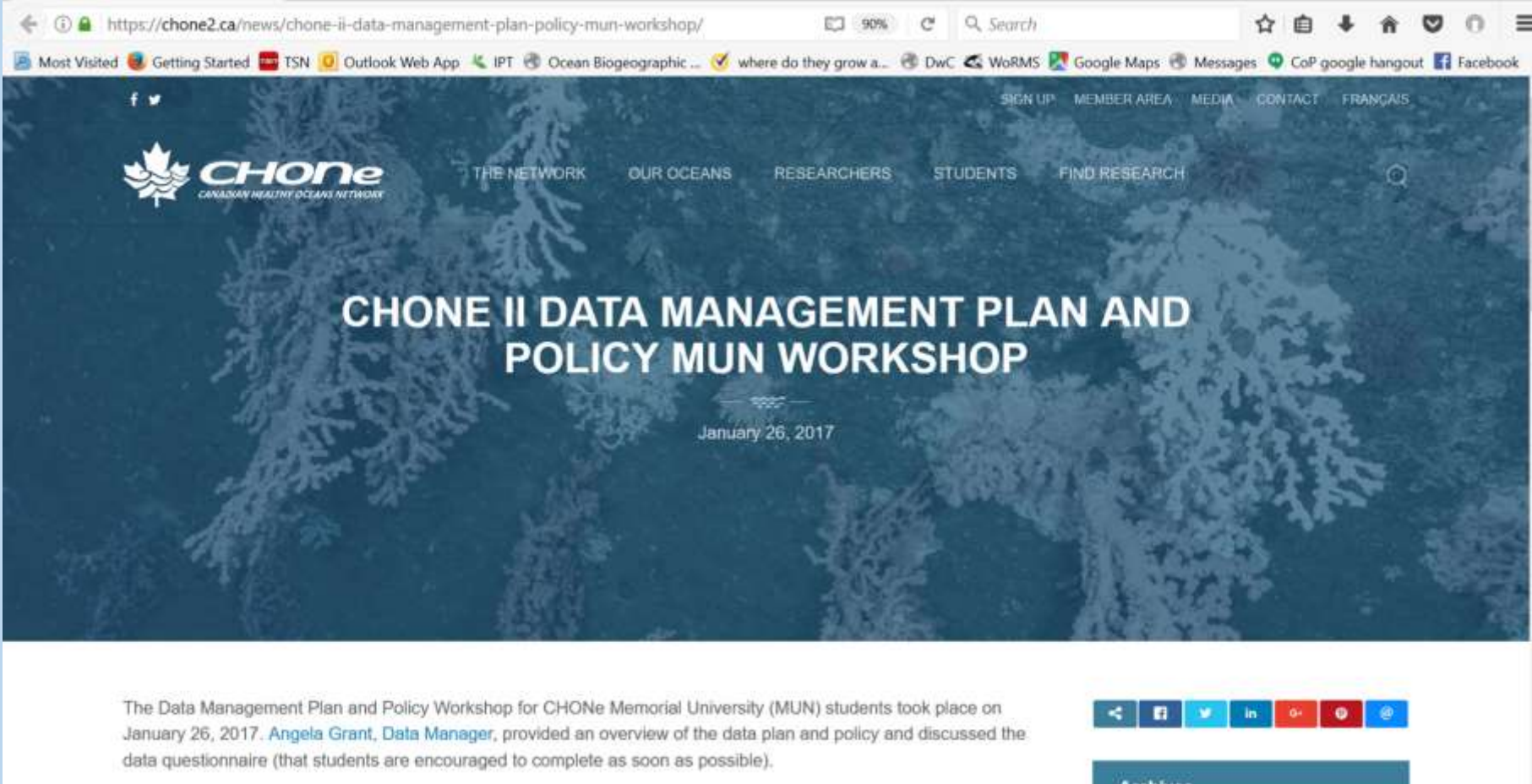
—  
Bob Branton

Director of Data Management, Ocean Tracking Network,  
Rm 7029, Biology Bldg, Life Science Ctr (LSC),  
Dalhousie University, 1355 Oxford Street,  
PO Box 15000, Halifax, NS, Canada B3H 4R2

Contact: work 902 494-7560; fax 902 494-1123; cell 902 225-3447; home 902 463-4657  
Email: [bob.branton@dal.ca](mailto:bob.branton@dal.ca)  
Website: <http://members.oceantrack.org>

# Module 1 – Making research data accessible

Canadian Healthy Oceans Network II (CHONe) data policy. (in process of being adopted)



The screenshot shows a web browser window with the URL <https://chone2.ca/news/chone-ii-data-management-plan-policy-mun-workshop/>. The browser's address bar shows a 90% zoom level and a search bar. The page header includes social media icons for Facebook and Twitter, and navigation links for SIGN UP, MEMBER AREA, MEDIA, CONTACT, and FRANÇAIS. The CHONe logo, featuring a stylized maple leaf and the text "CHONe CANADIAN HEALTHY OCEANS NETWORK", is prominently displayed. Below the logo, a navigation menu lists THE NETWORK, OUR OCEANS, RESEARCHERS, STUDENTS, and FIND RESEARCH. The main content area features a large, bold title: "CHONE II DATA MANAGEMENT PLAN AND POLICY MUN WORKSHOP", with the date "January 26, 2017" centered below it. At the bottom of the page, a paragraph of text reads: "The Data Management Plan and Policy Workshop for CHONe Memorial University (MUN) students took place on January 26, 2017. [Angela Grant](#), [Data Manager](#), provided an overview of the data plan and policy and discussed the data questionnaire (that students are encouraged to complete as soon as possible)." To the right of this text is a row of social media sharing icons for Facebook, Twitter, LinkedIn, and others.



# Module 1 – Making research data accessible

Do you know what your data policy is?

## COINAtlantic Data Accessibility Self-Assessment Tool

The COINAtlantic data accessibility self-assessment tool (CDAST) is for use in organizations to develop a benchmark to monitor progress in improving an organization's policies and procedures for providing effective and easy access to the data and information held by them.

For each principle one or more evaluation ladders have been developed with questions designed to assist the user in positioning their organization's policies and procedures on the ladder

<http://coinatlantic.ca/images/documents/COINAtlantic-Data-Accessibility-Benchmark-Data-Organizational-Self-Assessment-Tool.pdf>



ARTICLES

### The COINAtlantic Data Accessibility Self-Assessment Tool (CDAST)

The COINAtlantic data accessibility self-assessment tool (CDAST) is for use in organizations to develop a benchmark to monitor progress in improving an organization's policies and procedures for providing effective and easy access to the data and information held by them. Partners for this project included: Dr. Beirum MacDonald, School of Information Management, Dalhousie University for tool development, and Robert Branton, Emeritus, Ocean Tracking Network for tool testing and Google form development.

CDAST Presentation | Contact Form CDAST

Notes about the Tool

Data Accessibility Principles

Webinars

#### CDAST Components: Principals and Evaluation Ladders

The data accessibility principles used in the benchmark are an aggregation of four documents: the Organisation for Economic Co-Operation and Development's Principles and Guidelines for Access to Research Data from Public Funding<sup>1</sup>, the Government of Canada's Operating Principles for the Open Government site<sup>2</sup>, the United States Memorandum on Open Data Policy<sup>3</sup>, and the G8 Charter for Open Data<sup>4</sup>. For each principle one or more evaluation ladders have been developed with questions designed to assist the user in positioning their organization's policies and procedures on the ladder.

<sup>1</sup> OECD Principles and Guidelines for Access to Research Data from Public Funding <http://www.oecd.org/dataoecd/1/1/38500613.pdf>

<sup>2</sup> Operating Principles for the Open Government site <http://data.gc.ca/open-data/principles>

<sup>3</sup> Open Data Policy- Managing Information as an Asset <http://www.nf-research.gov/sites/default/files/ondp/memoranda/2013/11-13-13.pdf>

<sup>4</sup> G8 Open Data Charter and Technical Annex <http://www.gov.uk/government/uploads/attachmentes/2013/04/g8-open-data-charter-and-technical-annex>

# Module 1 – Making research data accessible

## Self-Assessment Tool: 11 Principles

- ▶ 1. Open Data by Default
- ▶ 2. Completeness
- ▶ 3. Primacy
- ▶ 4. Timeliness
- ▶ 5. Ease of Physical and Electronic Access
- ▶ 6. Non-discrimination
- ▶ 7. Licencing
- ▶ 8. Permanence
- ▶ 9. Usage Costs
- ▶ 10. Supporting Use
- ▶ 11. Evaluation

# Module 1 – Making research data accessible

## Self Assessment Tool Ladders

Choose your organization's position on the ladder:

### Principle 11: Evaluation

Periodic evaluation by user groups

5 - A regular program of evaluation is in place.

3 - Ad hoc evaluation is in place.

1 - Crisis-based evaluation is conducted.

0 - No evaluation.



7

# Module 1 – Making research data accessible

## How to Use the Tool

- ▶ Use repetitively over time to mark progress.
- ▶ The individual scores for each ladder are more important than the aggregate score: use the score to target areas to improve upon.
- ▶ Use separately to score multiple units within one organization.
- ▶ Have multiple persons from different perspectives (e.g. data manager, user, scientist) score within each unit.
- ▶ Weight scores based upon internal priorities.
- ▶ Some respondents may choose not to score some ladders (n/a).

8