Hands-on Activity 2: Data Sharing

Associated DataONE Lecture: Lesson 2: *Data Sharing*

Objectives: Students recognize the value of shared data, and develop some skills at searching for data, by searching data repositories for datasets related to their own research question.

Outcomes: (1) Students can search multiple data repositories for data relevant to a question of interest. (2) Students can explain how each of their selected datasets should be cited.

Time Needed: One hour out of class, 15 – 30 minutes in class discussion.

URLs: DataBib http://databib.org/index.php; ONE*Mercury* https://www.dataone.org/find-data , Dryad http://datadryad.org/ ; Knowledge Network for Biodiversity (KNB) https://knb.ecoinformatics.org/index.jsp

Additional Files Needed: None

Key Reading: Stephanie E Hampton, Carly A Strasser, Joshua J Tewksbury, Wendy K Gram, Amber E Budden, Archer L Batcheller, Clifford S Duke, and John H Porter 2013. Big data and the future of ecology. *Frontiers in Ecology and the Environment* 11: 156–162. http://dx.doi.org/10.1890/120103

Notes and Instructions for Instructors:

This exercise will provide students the opportunity to explore the diversity of data available in data repositories. They may also discover that with some data archives it is easier to find data than others, and that some datasets can be freely downloaded, while others require contacting the archive or the data author to obtain the data.   Students will also become familiar with how to cite a dataset.

In class, students can discuss what kinds of data they were looking for and how successful they were in finding it.   They can describe the archives they tried, and what the archives contained (to give other class members an idea of the range of data available).

Optional extension: The students could be asked to review the metadata for their selected datasets to determine whether there is additional information they would want to have before using the data. This additional activity is most appropriate when students have already learned about metadata; if so the students can also discuss metadata completeness.

Student Instructions:

Define a research question that you would like to be able to answer using existing data that may exist in a data repository. You may use the same question that you investigated for Exercise 1. Your goal is to identify three datasets from at least two repositories that you could use to address your research question.

Review the list of available data repositories at DataBib; there is a wide range of repositories listed here, covering topics from genomics to environmental and social sciences. Alternatively, visit major data clearinghouses and repositories such as ONE*Mercury*, Dryad, or KNB. Choose at least three data sources to explore in more detail.   As you search for data for your research question, consider these questions:

1) What options do each repository offer for data searching?   Can you search by title, keywords, or what else?   What makes one repository easier to find data in than another?

2) Once you have located a dataset of interest, examine the metadata to determine whether it is openly accessible.   If you cannot directly download the data, how would you obtain the dataset?

3) Do the metadata provide sufficient description of the data for the research project you proposed? For example, do you understand how the data were collected, the units of measurement for all of the variables, how missing values are represented, or other questions that come to mind?

4) For each of your datasets, describe how you would cite it in a paper.   The metadata often contains this information, or it may be provided on the repository web site.

5) Document the three data sets you found in the following worksheet (see below).

DataONE

Worksheet for *Data Sharing*

Topic Question:

DATA SOURCE 1

Dataset citation:

Data description (*what the dataset consists of*)**:**

Data source repository:

Method used to locate data set (*e.g. search method and keywords used*)**:**

How to access the data (*publically available, register and request, contact author etc.*):

DATA SOURCE 2

Dataset citation:

Data description:

Data source repository:

Method used to locate data set:

How to access the data:

DATA SOURCE 3

Dataset citation:

Data description:

Data source repository:

Method used to locate data set:

How to access the data:

DataONE