Hands-on Activity 1: Accessing Data in the Literature

Associated DataONE Lecture: Lesson 1: *Why Data Management*

Objectives: Students recognize the value of accessibly archived data, by experiencing the challenges of accessing data from published papers.

Outcomes: (1) Students can explain why accessible data archiving is valuable. (2) Students can provide strategies for getting data from published papers, and anticipate challenges to accessing the data.

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

URLs: Any resource for searching scientific literature (e.g. *Web of Science*, *Google Scholar*, *JSTOR*, *BioOne*).

Additional Files Needed: None

Key Reading: Carly A Strasser and Stephanie E Hampton. 2012. The fractured lab notebook: undergraduates and ecological data management training in the United States. *Ecosphere* 3:art116. doi: 10.1890/ES12-00139.1

Notes and Instructions for Instructors:

An intended take-home lesson of this activity is that access to valuable original data can become difficult or impossible in a short period of time after a paper is published, but this loss of accessibility is avoidable. How easy it is to access original data depends on the field; some fields have developed a culture of data sharing and data accessibility, including genetics, climate studies, and geography. Others do not have this tradition. Because of these field-specific cultures, students' success at accessing data will depend on the topic and question they chose.

It may be worth reviewing with the students the different ways by which scientists access others' data: data tables or published data appendices within a paper, extracting (estimating) data from published graphs, online data archives or data streams (either restricted to journal subscribers or public), writing the author and requesting the data etc.

After students have completed the exercise (see *Student Instructions*, below), have students discuss the challenges that they faced in figuring out how to access data from the published literature that are relevant to their question, and ways the students came up with to deal with the challenges. This can be done as a 15 to 30 minute whole-class discussion or in small groups with a report-out. Things to note include whether accessibility to data varied depending on the question addressed, and whether accessibility depended on how long ago the paper was published. Perhaps culminate the discussion with questions about why data underlying

published papers are not always provided in an accessible form, and what are some solutions to make data accessible.

Optional extension: Rather than having students fill in the included worksheet (see below), provide an online form (e.g., Google form or spreadsheet) in which students can enter in their responses. Have them include additional information about the publications they found (e.g., ISI subject category, year of publication, number of citations). After all students have completed the assignment, have them use the data they've collected to explore and discuss possible relationships between these additional variables and the accessibility of the data associated with the publications. This extension will be especially interesting for courses or workshops with large numbers of students, or if the instructor maintains the data entered and reuses them over multiple offerings of the lesson.

Student Instructions:

This activity is intended to help you understand the challenges and opportunities for conducting research, using data that have already been collected by other researchers, usually for other purposes.

1. Pick a topic of interest to you (e.g. *invasive species*, *net primary productivity*) and come up with an interesting question that could potentially be answered with existing data (e.g. *How fast is invasive species x spreading? Does acid precipitation reduce net primary productivity?*).

2. Search the literature for **five** peer-reviewed research papers that use data that would be potentially helpful for answering your question.

3. For each paper, fill in the worksheet information. Note that if the proposed *Method of Data Access* is contact the author, make sure to provide the author's current contact information (which may have changed since the paper was published). If you conclude that the data are inaccessible, explain why (e.g. data not in the paper and authors have no current contact info).

4. Be prepared to discuss with classmates (a) difficulties in accessing data from the literature to address your question and (b) ways that data accessibility could be improved.

Worksheet for *Accessing Data in the Literature*

Topic Question:

PAPER 1

Citation:

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

Method of data access (*published in the paper, publically available online, contact author*)

PAPER 2

Citation:

Data type:

Method of data access:

PAPER 3

Citation:

Data type:

Method of data access:

PAPER 4

Citation:

Data type:

Method of data access:

PAPER 5

Citation:

Data type:

Method of data access: