|  |  |
| --- | --- |
| **Data Ethics: Using Data Responsibly**  |  |

## *"WIRED data visualization illustrations" by CLEVER ° FRANKE is licensed under CC BY-NC 4.0*

## Objectives

Upon completion of this module, each student should be able to:

* Identify legal and ethical uses of a dataset in a repository's Terms of Service (TOS) and in a dataset's Creative Commons licensing
* Explain how metadata is utilized in ethical data use
* Compose a data citation using a digital object identifier (DOI)

**Introduction**

Data and datasets, as with any information, have both legal and ethical restrictions on their use. Though you are always responsible for ethical behavior in research and academic work, you may find that the data or the repository which publishes the data will provide guidelines for how the data may be ethically used.

These guidelines may be found as terms of service for the data repository, and in the license assigned to a dataset. A common licensing platform for ecological data is Creative Commons licensing. It always best to look for this information before you begin work with an existing dataset.

There are also ethical considerations with creating the data. Data creators must protect sensitive information, not only personal information such as people's names, addresses, or social security numbers, but also protected biodiversity data. A researcher of endangered species may collect population or nesting location data but may broaden those locations when the data are shared so that the species are kept safe. Data creators explain any methods they have taken to protect the sensitive data in metadata, or data about data. This metadata may be found within the data repository with the dataset and may also be downloadable as a READ ME file along with the data files.

A best practice for ethical use of a dataset is to provide a data citation when using that data in your own work. You can find formats for writing data citations in style guides (such as APA or MLA style), or suggested citations can be found in data repositories when downloading the dataset.

**Activity**

* Go to Dryad: <http://www.datadryad.org/>
* Go to the very bottom of the page to Dryad's Terms of Service, 3.2 Content Distribution

Explain who is responsible for the use of the datasets that can be viewed or downloaded. Is it Dryad? The data creators? Those who downloaded the data?

* Search Dryad for "Rudbeckia hirta" and choose a resulting dataset

Note the citation, including the DOI.

* Open the Data Publication (PDF) associated with the dataset

What is the license given for this dataset? Briefly describe what the license allows.

* Download the datafiles and unzip them

Are there READ ME files, or any other files associated with the data? What do they contribute to the data files?

**References:**

Braak, K. (2013). *Publishing sensitive data: Training course on data cleaning and data publishing, Nairobi, February 2013* [PowerPoint slides]. Retrieved January 16, 2020, from <https://slideplayer.com/slide/4471792/>

Chapman, A. D., & Grafton, O. (2008). *Guide to best practices for generalising sensitive species occurrence data, version 1.0*. Copenhagen, Netherlands: Global Biodiversity Information Facility (GBIF). ISBN:8792020062. Retrieved January 16, 2020, from <https://www.gbif.org/document/80512/guide-to-best-practices-for-generalising-sensitive-species-occurrence-data>

Creative Commons. (n.d.). *CC licenses and examples*. Retrieved January 17, 2020, from <https://creativecommons.org/share-your-work/licensing-examples/>

DataONE. (2012). *DataONE education module: Data citation*. Retrieved January 16, 2020, from <https://www.dataone.org/sites/all/documents/education-modules/pptx/L08_DataCitation.pptx>

DataONE. (2012). *DataONE education module: Legal and policy issues*. Retrieved January 16, 2020, from <https://www.dataone.org/sites/all/documents/education-modules/pptx/L10_Policies.pptx>

Lamar Soutter Library, University of Massachusetts Medical School. (n.d.). *New England Collaborative Data Management Curriculum: Module 5: Legal and ethical considerations for research data*. Retrieved January 16, 2020, from <https://library.umassmed.edu/resources/necdmc/modules>

Martone, M. (Ed.), and Data Citation Synthesis Group. (2014). *Joint declaration of data citation principles*. San Diego, CA: FORCE11. Retrieved January 17, 2020, from <https://www.force11.org/datacitationprinciples>

Olesen, S. (n.d.). *Ethics of data publication: Same same or different?* [PowerPoint slides]. Retrieved January 16, 2020, from <https://publicationethics.org/files/COPE%20presentation_S%20Olesen_for%20distribution%20%281%29.pdf>

**Assessment Pages**

**Summative Assessment**

Use the activity answers as a summative assessment.

Alternatively, have students write short reflection papers or a 3-2-1 paper:

* Three things the students learned from the lesson
* Two things the students would like to learn more about
* One question the students still have about the material

**Instructor Support Materials (as separate documents):**

1. Answer keys for formative and summative assessments
2. Slide set
3. Slide recording:
<https://chipcast.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=df2bc9ee-5586-4805-95ed-ab6001116ad1>