MAKING FIELD STATION DATA AVAILABLE TO THE WORLD

ESF
State University of New York
College of Environmental Science and Forestry

EDI

Introduction

The Environmental Data Initiative (EDI) supported six Summer 2018 (June 4 – August 13, 2018) Fellowships hosted and mentored by Organization of Biological Field Stations (OBFS) in its ecological data management training program. The fellows received training in data management and gained hands on experience through participation in data preparation, archiving and publishing with scientist.

Project Summary

Long time series of data are essential to document ecological change, explore scientific questions and recognize unusual events. Many field stations have collected data for decades but have limited resources to manage and archive data; other facilities have on-site archival but lack the ability to make the information publicly available. The SUNY ESF Adirondack Ecological Center (AEC) has over 85 years of wildlife, plant and climate records in digital and raw format and was awarded an NSF-funded Environmental Data Initiative (EDI) Fellow in 2018. In this summer EDI fellowship, we implemented practical recommendations for archiving and publishing data to reduce errors and increase efficiency. The fellowship objectives were to:

- Transform raw data into technically correct and consistent data
- Create concise, compact and detailed metadata
- Keep data and metadata separate
- Generate formatted output

The output consists of three parts: a data set in a plain text file format with Data Object Identifier (DOI), EML (Ecological Metadata Language) metadata, and graphs for analysis and visualization. Using best practices for organizing, archiving, and publishing long-term data can dramatically increase available information to science and improve the efficiency of data management without a loss of data quality.

Adirondack Ecological Center (AEC) at Huntington Wildlife Forest (HWF)

Adirondack Park is six million acres, covers one-fifth of New York State and equal in size to Vermont. It is the largest National Parks in the US; larger than Yellowstone, Yosemite, Grand Canyon and Olympic National Parks combined. It is established in 1892 to protect the forest and headwaters of five major watersheds.

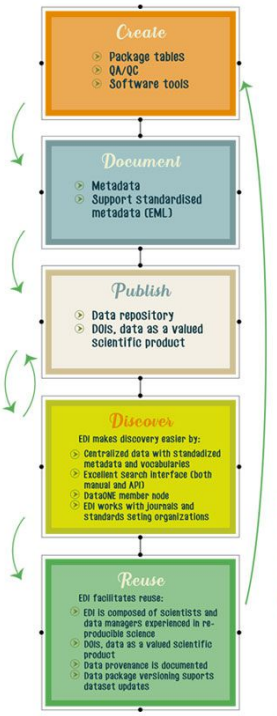
Adirondack Park contains millions of acres of largely undeveloped land and waters along with over 100 towns and villages. Research sites on the field station include 350-year-old unmanaged sugar maple, American beech, yellow birch and many conifers, compared to experimentally managed forest stands ranging in harvest age, land use history and species composition.

Huntington Wildlife Forest (HWF) is a 6,000 ha field station in the center of the Adirondack Mountains.

There are complex relationships between animals, their habitat and climate, predators and prey, pollinators, seed dispersers and plants. Long-term records are important to characterizing and assessing population trends and changes to the ecosystem from management, fragmentation and other impacts. Scientists began collecting data at Huntington Wildlife Forest experimental station in 1932.

Methodology

The Environmental Data Initiative (EDI) is an NSF-funded project to curate and archive data from field-based biological research projects. EDI is using R Languange and EML library. The methodology explain in this chart:



Result

- > 50,000 records / datasets already uploaded to the EDI portal
- 7 lakes were sampled on lake ice date dataset which are Arbutus, Catlin, Deer, Military, Rich, Wolf Lake and Lodo Pond
- 6 small mammal sites were sampled which are AD, EF, HA, MS, RM, SB
- 6 bird sites were sampled which are AD, HA, MS, NA, NAW, SB
- 156 plant species and 86 bird species
- 6 dataset already uploaded in production and staging server, here are the link:

Production :

1. Grouse (1984 - 2017) : https://portal.edirepository.org/nis/mapbrowse?scope=edi&identifier=220&revision=1
2. Small Mammal (1981 - 1987) : https://portal.edirepository.org/nis/mapbrowse?scope=edi&identifier=221&revision=1
3. Small Mammal (1981 - 1996) : https://portal.edirepository.org/nis/mapbrowse?scope=edi&identifier=221&revision=2
4. Lake (1874 - 2016) : https://portal.edirepository.org/nis/mapbrowse?scope=edi&identifier=231&revision=1
5. Bird (1952 - 1964) (1983 - 2008) : https://portal.edirepository.org/nis/mapbrowse?scope=edi&identifier=232&revision=2

Staging :

6. Phenology (1929 - 2016) : https://portal-s.edirepository.org/nis/mapbrowse?scope=edi&identifier=221&revision=5

Sharing data. AEC plans to compare the data collection methodology and spatio-temporal coverage of AEC datasets with unrelated but similar ecological datasets (e.g., NEON, LTER) to determine the degree to which data synthesis for large-scale analysis is feasible. AEC maintains 85 years of long-term data on forest ecology, wildlife population change, phenology, biogeochemistry, limnology, and more. Yet the AEC and its home institution has limited capacity to manage and make public long-term datasets. The EDI Fellow greatly increased AEC's ability to get well-documented "dark data" online and available to the scientific community, students and public.

Rationale

Adirondack Ecological Center

Involvement

Providing housing and travel arrangements; scientific mentoring to provide context for how the data are collected and used, including some field visits to see sites and projects first-hand. Required planning ahead to identify and prioritize datasets.

MENTOR

Benefit

The fellowship permitted AEC to expand access to several long-term datasets which were never publicly available (other than to researchers who contacted our staff). AEC research staff also partnered with the EDI fellow who shared tips on using the R environment to manage data and the EDI metadata creation process, with the objective of increasing onsite staff capacity to manage data. Mentoring the fellow enabled the advisor to determine how best practices in data publishing could be achieved and will enable future data management planning. It was a pleasure to work with a fellow who possessed high-caliber informatics expertise and a strong work ethic.

FELLOW

Benefit

I feel immensely grateful to be part of the EDI Summer Fellowship 2018. This program has given me the unique opportunity to transition from more computer-based study in R Language to the real ecological data and its issue in field stations. I would not have been able to finish the internship without the Ecological Metadata Language (EML) training. The training and the internship not only has sharpened my skill in data management but also enhanced knowledge and awareness in ecology and conservation. Finally, the opportunity to regularly interact with such a diverse group of fellows and scientists has been invaluable to my personal and professional development.

Authors: Prita Novita, prita_an@my.uri.edu
Supervisor: Stacy McNulty, Smcnulty@esf.edu
Affiliation:
1. Adirondack Ecological Center, the State University of New York College of Environmental Science and Forestry, Newcomb, New York
2. Environmental Data Initiative (EDI)

Acknowledgments