Notes toward a data plan: Lakeshore Nature Preserve

Description of data

General Ecology 460 students have been sent into the Lakeshore Nature Preserve for some years, once per academic-year semester, to do flora and fauna counts and other observations. The resulting data (from rather fewer years) have been collected in spreadsheets, one per semester. Other data, such as soil and water analyses, related to the Nature Preserve have been collected as part of the Biocore Curriculum honors program.

Context for curating data

These data are fertile ground for time-series investigations into the impact of campus land use or climate change on the Lakeshore Nature Preserve. They could also serve as real-world experience in data-management best practices for students, given that requirements are rising for research-data management and sharing.

Nature and amount of data

The data currently reside with the course professors. File formats vary, though most data are in the form of Excel spreadsheets. Some data from early surveys exist only on paper. The General Ecology 460 digital datasets are well-organized, and comprehensive methodology documentation exists in the form of instruction sheets and maps for students. The Biocore datasets require further organization and description; some materials (such as soil-analysis reports) have lost their research context entirely.

Minimal storage space is required to keep these datasets.

Recommendations

The Biocore datasets are not viable candidates for preservation and dissemination until they are organized by research project and described. This could be a useful class project; certainly current and future classes should be encouraged to document their research-project methods and results with a view toward compiling the kind of time-series that exists for General Ecology 460.

Excel spreadsheets are an acceptable but not ideal file format for preservation. For best preservability, formulas should be replaced by values, and the

spreadsheets saved into an open, documented format such as Open Document Format. Best of all would be a text-based format such as comma-separated values; since the spreadsheets are in multiple pages, however, this would require some thought given to data modeling and column descriptions.

Perhaps the best way to reformat the data for easiest access and best preservation, as well as to rekey the data trapped on paper, would be to ask General Ecology 460 students to answer a research question based on the historical datasets. To keep the workload manageable for a class, it may make sense to have classes sample at multiple-year intervals (e.g. current year, 3 years back, 6 years back...). Over time, the entirety of the analog data would be rekeyed, and the entirety of the digital data reformatted into more easily computable forms.

The datasets should be added to a digital repository for dissemination and long-term curation; relying on individuals creates a serious risk of accidental data loss. MINDS@UW (http://minds.wisconsin.edu/) is an acceptable candidate, though relevant disciplinary repositories may exist. If an article is published based on these datasets, Dryad (http://datadryad.org/) will accept them. Analog materials may be accepted into the University Archives.