UC Office of the President
CDL Staff Publications

Title
Ecological data in the Information Age

Permalink
https://escholarship.org/uc/item/7f41z0h2

Journal
Frontiers in Ecology and the Environment, 10(2)

Authors
Hampton, Stephanie
Strasser, Carly
Tewksbury, Joshua

Publication Date
2012-03-01

Copyright Information

Ecological data in the Information Age

Most of us can close our eyes and imagine a future in which ecological data are at our fingertips, as easy to access online as current weather conditions, satellite images of our field sites, or cute videos of kittens. In this future we can easily find information about the temporal trends of aphid outbreaks around the world or the spatial occurrences of Atlantic cod, to use in research or in the classroom. We can imagine services that operate "behind the scenes" to intelligently aggregate data and allow for tailored datasets and analytics, such as averaging soil respiration rates within a certain elevational band, or estimating the age structure of local deer populations.

Although we can easily imagine these developments, and perhaps happily anticipate them, very few of us act as though we are the generation of scientists that will make them happen.

We are not documenting, sharing, or archiving our data in ways to ensure that this future is attainable. Surveys demonstrate that we record our data in highly idiosyncratic ways, with much of the information about the data collection (metadata) remaining as oral history within a research group. After the data have been used for publication, their electronic datasheets steadily atrophy in hardware and software that become obsolete, and associated metadata are often forgotten. After careful planning and collection by experts, valuable ecological datasets are gathering dust in small piles all over the world.

Why are we so reluctant to preserve our data for the long term in public archives? The reasons given for not doing so are diverse, and many authors have preceded us in analyzing the anachronistic lack of data sharing in ecology. In short, ecologists see few near-term rewards and many costs associated with making data public. Making matters worse, very few ecologists have a clear idea of how they would go about sharing data even if they wanted to. Thus, while ecologists are as active as any other group in embracing the Information Age – from smartphones for field research to distributed software development for statistics – paradoxically we are still keeping our hard-won ecological data hoarded in idiosyncratic lockboxes.

If this world of readily accessible ecological data is coming "someday", how do we get there from here? How can we become the generation of ecologists that creates the bridge to that future of ecological data access?

We propose a first step: treat your ecological data as a real product of research, not just a precursor to a set of publications. Many of our datasets are irreplaceable, documenting organisms, patterns, and ecosystems that are rapidly changing. With this in mind, manage your data with the conviction that they will have a lifetime far beyond your own. Document your data for someone – a stranger – who will discover the information decades from now.

As a simple starting point, read Borer *et al.* (2009; *Bull Ecol Soc Am* **90**: 205–14) – an introductory paper with sound and practical suggestions – with your research group or journal club. To see the level of detail you might need to consider, examine sources like *Ecological Archives*, the Ecological Society of America's journal geared toward publishing self-describing datasets. Start creating machine-readable metadata early in your project – even before data collection begins – through free software like Morpho (http://knb.ecoinformatics.org/morpho portal.jsp). This step facilitates sharing your data and metadata with colleagues now, and uploading to public repositories when you are ready. Repositories such as *Ecological Archives* connect your dataset to many others in the Knowledge Network for Biocomplexity (http://knb.ecoinformatics.org), and opportunities for connecting datasets will grow with improvements in cyberinfrastructure – such as the US National Science Foundation's (NSF's) DataNet project known as DataONE (www.dataone.org), which will confederate environmental data from many repositories.

The NSF's recent requirement that data management plans be included in all proposals is a logical step toward data publication, and many believe that it may provide a tipping point for broad adoption of data publication policies in the US. In our opinion, widespread sharing of ecological data will be a welcome change.

If ecologists are to have societal relevance – alerting managers to unusually rapid range expansions in nonnative species, illuminating long-term oscillations in fish production, or determining the historical distribution of a plant with newly discovered medicinal properties – we need to get serious about bringing our discipline into the Information Age now, by documenting and sharing data. What may be perceived as a waste of time to this generation of ecologists will not be viewed that way by future researchers, who will be prepared to engage in more transparent and open science.

Stephanie E Hampton
NCEAS,
University of California,
Santa Barbara, CA

Joshua J Tewksbury
University of
Washington,
Seattle, WA

Carly A Strasser
California Digital
Library,
University of California,
Oakland, CA