

University of New Mexico

UNM Digital Repository

Long Term Ecological Research Network

Museums and Research Centers

3-2-2013

LTER Network Data Availability Correspondence

Emery Boose

Follow this and additional works at: https://digitalrepository.unm.edu/lter_reports



Part of the [Other Ecology and Evolutionary Biology Commons](#)

Recommended Citation

Boose, Emery. "LTER Network Data Availability Correspondence." (2013).
https://digitalrepository.unm.edu/lter_reports/163

This Handout is brought to you for free and open access by the Museums and Research Centers at UNM Digital Repository. It has been accepted for inclusion in Long Term Ecological Research Network by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Dear Folks:

In addition to the answers to Emery and Scott's questions (attached), I put forth below some ideas for your consideration.

I have spent some time with the PASTA development staff over the last two days, and I believe that we have come up with an approach to registering for data download that addresses all of the core issues. These core issues, as we see them, include the two that Gus mentioned:

1. Maximizing data access
2. Identifying and tracing data use

as well as three more that we consider equally important:

3. Providing flexibility for sites to determine accessibility to their own data
4. Providing an option for users to choose whether they wish to receive updates on data they have downloaded.
5. Defining and encouraging ethical data use.

We suggest the following elements are necessary to address all of these core issues.

Provide clear instructions for data users on the Data Portal page. These instructions will explain our expectations of data users, define ethical behavior, encourage examination of metadata and the LTER Data Policy, and offer the option for users to register voluntarily. We believe that this information is best presented before searching for data. In particular, we do not wish to force the user to read the data policy before each data download. That would certainly constitute cruel and unusual punishment. Please see the attached document (Portal / Access / Registration Text) for an idea of what such text might look like.

Provide a mechanism to collect information for identification and tracing of data users. Registration when using the LTER Data Portal provides the means to collect information to identify data users and trace data use. This can be accomplished through a registration screen that includes name, institution, e-mail address, and general categories of use (see Portal / Access / Registration Text). This procedure does not provide specific information on data use (e.g., "I am writing a paper for Science and need a test data set") for each data set downloaded, but does allow the data provider to request such information directly if they wish to keep closer tabs on data use.

Allow sites to determine which data sets require registration. Sites can restrict access to data sets in PASTA through settings in the associated metadata, which provides flexibility for sites to behave differently. For example, some sites may require registration for all data sets, some for none, or there may be a mix. This allows maximum control to sites, and does not

force sites to conform to a majority decision. In practice, this would mean that an unregistered user would only have access to those data sets in PASTA that were completely unrestricted. Voluntary registration at the Data Portal would provide access to all data sets except those completely restricted by the sites, such as endangered species data. If a user did not register voluntarily, he would be unable to obtain those data sets marked for identification and tracing by sites until he did register. The attached flowchart (Tracking Data Downloads) shows how this would work.

I should point out that mandatory registration for any data sets, if adopted, may restrict access to these data sets by external applications using the PASTA API. Resolving this issue is complicated and would require further discussion.

I hope this provides enough flexibility to satisfy all.

Regards,

Bob

On 3/2/2013 5:06 PM, Boose, Emery wrote:

Hi Mark, James & Duane,

The issue of tracking data downloads from PASTA was discussed again at the EB meeting this week. Several technical questions came up that perhaps you could shed some light on:

1. What information (if any) does PASTA currently record when a data package is downloaded (e.g. date, time, file name, user IP number)?
2. If the user were presented with a form asking for information (e.g. name, email address, institution) and/or one or more checkboxes (e.g. acceptance of LTER data policy) before data could be downloaded, would this prevent web crawlers from gathering the information required for LTER data packages to appear in search engines such as Google?
3. If the answer to #2 is yes, are there technical workarounds? Would these be feasible for PASTA?
4. Or, given the anticipated size and number of data packages in PASTA, would it be better to exclude web crawlers regardless of whether user information is tracked?

Thanks for any information you can provide. Please feel free to reframe the discussion if I haven't asked the right questions.

Best regards,
Emery

