# Policy on Data Sharing and Access for Collaborations

This project relies on several major types of data: those that belong to individual colleagues and are not public domain, those that are publicly accessible but have sharing rules specified by individuals, and those with comparatively free open access (e.g. NASA).  As part of our funding agreement with XXXX, this group has agreed to comply with the XXXX Data Policy (url), but that still provides a fair amount of discretion to the group.  With this document, we describe our policies about how the data will be used, and how they will be shared. This is a living document – e.g. technologies and perspectives will change as we move forward - we will fully document any policy changes and we will post all updates to the data policy to this document in our Google drive folder (url).

First, we clarify that we all agree to the following principles that guide our decisions:

1. We respect the ethical rights and confidentiality restrictions that are associated with all data that are made available to the group

2. Open data sharing is fundamental to effective scientific synthesis

3. Reproducibility of our findings requires that all data and code associated with a finding are made openly and publicly available unless it conflicts with the rights in (1)

In the remainder of this document, we outline our policy for specific situations.

## Individual data not in the public domain

Some data being shared are generated through existing collaborations that are covered by data policies within existing Memos of Understanding (MOU's), and MOU policies take precedence.

The individual data sets that are not in the public domain or already openly licensed (e.g., via CC0) belong to the individuals contributing them; use of these data requires explicit permission (on a project by project basis) from them, and may involve an invitation for co-authorship (see co-authorship policy document to help determine whether co-authorship will be warranted in the end). Before requesting permission for use, the project collaborators will discuss the request with the full group to ensure that we are not duplicating each other's efforts or creating complications for each other.

While these data sets will be available on our restricted access collaborative site – the following rules apply:

- These data will never be shared outside of the group without explicit permission from the colleagues who own them;

- Individuals within the project proposing to use the data should discuss the proposed project with the team before embarking on analyses – please consider that this is a big group in which several people may already be undertaking similar analyses;

- We will enthusiastically support efforts to make these data public, *at any time* that the owners choose to make them public.

## Individual data with public access

We agree to abide by the attribution requests outlined for any publicly accessible data resources. We will maintain a shared document that outlines the attribution requirements for each data set included in our analyses, and note in detail how the requirements have been met (e.g. who contacted the owner, when, for which sub-project and specific uses, and details of the response including acknowledgement that derived data will be made publicly available). Those data sets that are more fully open-access will be included in this list with details on the attribution wording suggested by the contributor.

Modifications of documents in:
Cheruvelil, K. S., Soranno, P. A., Weathers, K. C., Hanson, P. C., Goring, S. J., Filstrup, C. T., & Read, E. K. (2014). Creating and maintaining high-performing collaborative research teams: The importance of diversity and interpersonal skills. *Frontiers in Ecology and the Environment*, *12*(1), 31-38. DOI: 10.1890/130001

**Derived data resulting from this project**

Our goal in data management is to be an exemplar for sound data practices, not only to facilitate our own collaboration but also to advance the field. Making data public and linked across repositories is relatively new territory for [field], and we need the enthusiasm of the full team to achieve these goals.

*Making Data Public* – Derived data are the data used for our analyses that could not be back-transformed to arrive at the original data (e.g. summer mean temperatures from multiple locations in one of our models). The team is committed to making our derived data public, upon publication if not before. Therefore there is a premium on documenting and organizing these data in standardized formats from the very beginning. Transformations, interpolations, etc, need to be documented in detail in order for these derived data to be useful within our project and for others.

*Management, Documenting and Archiving* – Different fields vary in their advancement of data archiving and sharing practices. Therefore we need to recognize that the fields represented by our team now may have different data practices and there is not a 'one size fits all' approach to data management. Our goal is to meet the gold standards of data management and data sharing across the fields represented in our project, while maximizing interoperability and cross-disciplinary discovery of the different data types. Thus, broadly speaking, we are pushing genetic data down a different pathway than non-genetic ecological data, and will do our utmost to provide links (e.g. URLs) that associate genetic data with non-genetic data across repositories.

1) Experimental or other non-genetic ecological data will be documented in Ecological Metadata Language (e.g., using the desktop application Morpho), and will be archived in DataONE (via the Arctic Data Center). Relevant links to data in other repositories will be established in these records.

2) Phylogenies resulting from our work with be housed in TeeeBASE. Relevant links to data in other repositories will be established in these records.

*Pre-publication data sharing within the team* – Active non-genetic data sheets and their associated metadata will be housed on the pass-worded collaborative site. Genetic data should not be housed on the collaborative site (those data sets are too big!) – URLs for these data can be grouped together in appropriate locations, e.g. within tasks.

We encourage team members to look at these working data sheets, and to consider ways in which they could collaborate, following these rules:

1) This should be obvious, but… Don't upload any changes to someone else's data files – this would be very difficult to do by accident on our collaborative website – if you want to make a comment on a data file or suggest an analysis or transformation, just keep it separate from the working data file;

2) If you would like to collaborate on an analysis, contact the investigators before you get too far into thinking about it (see Collaborator policy);

3) Team members and collaborators will not be allowed to share or make available in any form (other than through publication of research results) any pre-publication data unless discussed with the team.

*Public data sharing and attribution* – When data are publicly accessible, the metadata for each data set will include explicit statements about our expectations for use, e.g. attribution and notification. These expectations may be decided on a project by project basis, or this document later may be modified to create a blanket statement. The big repositories like the Arctic Data Center provide a mechanism for traditional citation formats, such that data providers can receive credit for data use. Additionally, it's likely that data

providers will be contacted with invitations for collaboration. However, the data will be in the public domain and there may be situations in which our expectations are not met. A study of data sharing in the LTER system found that very few conflicts arise over the use of published data, and those that do almost entirely occur between established collaborators (who presumably have not carefully followed our guidelines here!), so we should set our expectations reasonably but not anticipate problems.

Modifications of documents in:
Cheruvelil, K. S., Soranno, P. A., Weathers, K. C., Hanson, P. C., Goring, S. J., Filstrup, C. T., & Read, E. K. (2014). Creating and maintaining high-performing collaborative research teams: The importance of diversity and interpersonal skills. *Frontiers in Ecology and the Environment*, *12*(1), 31-38. DOI: 10.1890/130001