

**Living Data Project: Application to host a data rescue project**

**Section A: About you**

(1) Your name: \_\_\_\_\_

Your organization: \_\_\_\_\_

(2) Who is the original data owner and what is their organization? If not you, please explain why the original data owner is not applying (e.g. deceased) and your interest and involvement in the data.

(3) Explain why you or your organization is not able to archive this data without involvement from the Living Data Project. For example, you may lack specialized training in data management and reproducible research and/or, you may be close to retirement or the original data owner may be deceased.

**Section B: The data**

(4) Years over which the data was collected: from \_\_\_\_\_ to \_\_\_\_\_

(5) Spatial coverage of the dataset (if just one location, list; otherwise please explain area over which data was collected)

(6) In broad terms, what types of variables were collected in this dataset (e.g., geolocation, temperature, abundance, diversity, growth, etc)?

(7) What was the primary purpose of collecting this dataset? Please indicate in your answer if this is survey or experimental data.

(8) Why do you think this data is important to preserve for future generations of scientists? Some reasons might include: historical data suitable for establishing baselines for current/future change; time series suitable for understanding dynamic processes in ecology; classic experiment that had profound impacts on theory; includes species or ecosystem(s) of high importance for conservation, cultural or commercial reasons.

(9) The data to be rescued is currently in this format (check all that apply):

- Excel spreadsheets
- Other format spreadsheets (list: \_\_\_\_\_)
- MS Access
- SQL database
- Paper datasheets
- Within published reports
- Other: (list \_\_\_\_\_)

(9a) I verify that all data to be rescued is in digital form and can be transmitted online to the Living Data Project. If this data is on defunct digital media (e.g. magnetic tapes, floppy disks) I will take responsibility for extracting.

- Yes    No

### Section C: Required skills

(10) All graduate students will be trained in R-based tools for data validation and compilation which we believe will be adequate for most data rescue projects. If you believe that additional computational skills are needed for your dataset, please check any that apply.

- Programming (list languages: \_\_\_\_\_)
- Geographic Information System (e.g. ArcGIS, QGIS, geospatial R packages)
- Structured Query Language (SQL, used in relational databases)

Other: \_\_\_\_\_

(11) This conceptual background is required/desirable for the project (check all that apply):

- Entomology / insect ecology
- Ornithology/ avian ecology
- Mammalogy/ wildlife ecology
- Forestry / forest ecology
- Fisheries / Marine ecology
- Agriculture / agroecology
- Botany / Plant ecology
- Limnology / Aquatic ecology
- Population ecology
- Community ecology
- Landscape ecology
- Climate change biology
- Evolutionary biology
- Phylogenetics
- Population genetics
- Genomics
- Other: \_\_\_\_\_

(12) The language used in the dataset is (check all that apply):

- English:
- French:

(13) Please list any other skills or requirements needed to rescue your dataset that have not been covered.

**Section D: The project process**

(14) Who at your organization will be primarily responsible for mentoring the project? This person should take the lead in supplying the data and answering questions about the data and data collection, and meeting with the LDP instructors and graduate student rescue team. We anticipate five to ten meetings over the fall.

---

(15) In addition to putting data management and archival skills in practice, what other potential benefits will the graduate students gain from participating in this project? We are particularly interested in our students learning about non-academic careers and the potential to be involved in data analysis, publications and public outreach in the future.

**Section E: Outcomes**

(16) List the desired outcomes (products) of the project. Indicate if the outcome is required or optional if time permits. Examples could include: creating a data management plan for future data archiving, combining multiple data sources, performing quality control on the data, writing a data paper as part of the archiving process. Note: All projects have

mandatory outcomes of depositing the data in a public repository and writing a short public description of the researcher and dataset for the CIEE website.

(17) Do you have a preferred (or employer mandated) data repository for the data to be archived in? If so, please list here (otherwise leave blank, and we can decide together):

---

We accept any data repository that employs best practices in the permanent storage of and open access to data. A global registry of such repositories can be found at [r3data.org](http://r3data.org) (please contact us if your preferred repository is not on this list). Popular data repositories in ecology, evolution and environmental science include: KNB, GBIF, Environmental Data initiative, Genbank, Pangaea, Sequence Read Archive. Popular data repositories for scientific or academic data in general include: Zenodo, Dryad, Figshare, and Borealis. Federal government agencies are typically mandated to ensure that archived data in a repository is findable via the Open Government Portal.

(18) Do you have any concerns about this data becoming publicly accessible? Please let us know, so that we can work with you on possible solutions. For example, concerns about the potential misuse of data can sometimes be dealt with by coding into the data any information that should be conveyed to future users (e.g. changes in collection method or data reliability). Concerns about the locations of endangered or commercially valuable species can be addressed by decreasing the spatial resolution. Concerns about allowing time to complete a publication before data is made open could be addressed by the embargo facility of Pangaea's repository, or the no-questions 1 year embargo policy of Dryad and Figshare. Concerns about commercial use of data can be addressed by the correct [Creative Commons license](#) (CC-NC prevents commercial use).

(19). I acknowledge that it is my responsibility to determine the variables to be archived, and the exact datasets to be archived, and that further data cannot be added to the project once it has been started.

Yes  No