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

**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 lack time or resources (e.g., you may be close to retirement or the original data owner may be deceased or absent).

**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) Is data entry required?    ☐ No    ☐ Yes

    If yes, how much data entry is required?

```
┌─────────────────────────────────────────────────────┐
│                                                     │
│                                                     │
│                                                     │
└─────────────────────────────────────────────────────┘
```

## Section C: The intern

(10)    These data skills are required for the internship (check all that apply):

☐     Programming (list languages: _____)

☐    Geospatial (e.g. ArcGIS, QGIS, geospatial R packages, etc.)

☐    Relational database (e.g. MySQL, etc.)

☐     Statistical

(11)   This conceptual background is required/desirable for the internship (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) What constraints are there, if any, related to the physical location of the internship?

　　　　　☐　　　The intern will be based at our organization, located:

　　　　　_____

　　　　　☐　　　This internship can be done entirely virtually

　　　　　☐　　Other (describe):

_____

(13)   The intern will need fluency in (check all that apply):

☐ English:

|  | Read | Write | Oral |
|---|---|---|---|
| Required |  |  |  |
| Preferred |  |  |  |

☐ French:

|  | Read | Write | Oral |
|---|---|---|---|
| Required |  |  |  |
| Preferred |  |  |  |

(14)   Please list any other requirements of the intern that have not been covered.

|  |
|---|
|  |

**Section D: The internship process**

(15) Who at your organization will be primarily responsible for mentoring the intern? This person should take the lead in supplying the data and answering questions about the data and data collection, and meeting with the intern on a weekly basis.

_____

(16) If the intern will be based at your organization, and will need to travel to do so, is your organization able to contribute to any travel and accommodation expenses? The Living Data Project has only limited funds for these costs.

_____

(17) In addition to putting data management and archival skills in practice, what other potential benefits will the intern gain from participating in this internship. We are particularly interested in our interns experiencing non-academic workplaces and the potential to be involved in data analysis, publications and public outreach.

**Section E: Outcomes**

(18) List the desired outcomes (products) of the internship. These will form part of the formal agreement with the intern, so please indicate if the outcome is required or optional if time permits. Examples could include: creating a data management plan for future data archiving, determining which variables should be archived, combining multiple data sources, performing quality control on the data, creating metadata for the project, 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.

(19) 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 (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 (Scholars Portal) Dataverse. Federal government agencies are typically mandated to ensure that archived data in a repository is findable via the Open Government Portal.

(20)   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).