Data Management Plan

I. Policies for Data Sharing and Public Access

Our data sharing and public access policy is consistent with Creative Commons Attribution 4.0 International (CC-BY 4.0), a public license that makes the data freely available (i.e. use without permission) provided the source of the data are attributed properly, and that the user does not claim an endorsement by the Smithsonian Institution for their specific application of the data (unless stated otherwise). The policy was adopted because the data are collected by both US Federal and non-Federal employees, making the application of Federal copyright law to the data unclear. CC-BY 4.0 is used by Federal and State agencies, and non-profit organizations (https://creativecommons.org/licenses/by/4.0/).

II. Data Collected, Formats and Standards

Raw data are archived after being subjected to QA/QC processes. **Derived data** are calculated from raw data, which we archive along with the code (SAS or R) used to execute the calculations. Derived data are reduced by replicated unit (plot or chamber) and reported as single values. We do not archive treatment-level summaries of the data as they are easily calculated by the user. The data are stored in Excel-format files, each of which also contain a **Metadata** worksheet listing variable names, variable units, missing data codes, and other information that places the data into context. The Excel-format is used only for easily associating the data with the metadata; the files do not contain formulas, formatting, or other proprietary features and are easily exported in a standard comma delimited (csv) format.

We adopted the Harvard Forest metadata standards (http://harvardforest.fas.harvard.edu/data-archive/policies-guidelines). Raw data collected on paper are scanned and archived on a Smithsonian Institution network (see below) in Archive PDF (pdf/a) format. These scanned documents are not posted for public download but are available from any of the LTREB PIs upon request. The data are archived and publically available two years (or less) after the samples are collected. This period allows 12 months for analysis and QA/QC, and 12 months for publication of novel results. All files are posted on a Smithsonian-maintained site dedicated to the project (http://serc.si.edu/GCREW/data), with a sub-page dedicated to the data from each experiment (Table 1). Archived variables are listed in Table 2.

Table 1. Websites where data from the three long-term experiments are freely available to the public.

Experiment	**URL**
eCO_2 x Community	https://serc.si.edu/gcrew/CO2data
eCO_2 x N	https://serc.si.edu/gcrew/nitrogendata
eCO_2 x N x *Phragmites*	https://serc.si.edu/gcrew/phragmitesdata

III. Data Storage, Preservation, and Dissemination

Standard procedures for processing samples, entering raw data, performing QA/QC analyses, reducing data, and tracking corrections are documented in files associated with the respective datasets on the Global Change Research Wetland website (Table 1). All files posted on the website are assigned a name that ends in the posting date (dd-mm-yyyy). Once posted, the original raw and derived data files are locked by the managing PI as read-only. Any subsequent changes to the file are made by the managing PI and logged into a "Change Log" worksheet found in the same Excel file that contains the data and the metadata. Subsequent changes trigger a new version of the dataset identified by the new date. Retired versions of the file are not archived on the public website but are available upon request. All data files are saved to a Smithsonian Institution network drive; we do not depend on local drives other than for temporary transfer from sensors, devices or instrument memory to the network. Changes to files stored on the network drive are backed up by the Institution every evening, and a complete backup occurs every weekend. Backup tapes are stored offsite and the backup data are kept for one year.

Primary responsibility for managing the data rests with Dr. Patrick Megonigal, a PI on the LTREB grant and a full-time employee of the Smithsonian Institution since 2000. Responsibility for the data will be passed to PI Genevieve Noyce, an early career federally funded Senior Scientist, continuing

our history of deliberate succession planning that began when Dr. Bert Drake, who began the longest of the datasets in 1987, retired and data curation responsibilities passed to Dr. J. Patrick Megonigal.

IV. Planned Upgrades to Data Storage, Preservation, and Dissemination

Our standard procedures for data curation as described in section III have been successful as shown by the large number of synthetic studies that discovered and used our data (see Project Description). However, we will implement new procedures that leverage the workflows and processes developed for the Coastal Carbon Research Coordination Network (https://serc.si.edu/coastalcarbon), the focal activity of which is the Coastal Carbon Atlas. The Atlas is a stable repository of >16,000 curated coastal marsh, mangrove and seagrass soil profiles worldwide that follows FAIR (i.e. Findable, Accessible, Interoperable, Reusable) standards. Starting in 2025, we will migrate LTREB data from our project website to a stable FigShare site repository that assigns digital object identifiers (https://smithsonian.figshare.com/). The move will significantly improve the permanence and the discoverability of the data by committing the Smithsonian Institution to care for LTREB data in perpetuity.

Table 2. Dataset status. Data that are collected, archived and made public from three long-term experiments that began between 1986 and 2010, and are supported exclusively by LTREB funding. Level-0 indicates that the data are organized into spreadsheets and subjected to initial quality control (QC), but not to a higher level of QC required for public posting. Data that are publicly posted are considered Level-1 data. QC for Level-0 data is limited to programmed checks for unrealistic values while Level-1 data is subjected to visual and statistical analyses before releasing it publicly.

Dataset	Experiment	Level-0	Level-1	Notes
Net Ecosystem Exchange	Native *e*CO_2	1986-2008	1986-2008	Dataset was formally sunset. This dataset ended before the LTREB grant began.
Peak season shoot biomass	Native *e*CO_2 Native *e*CO_2xN Phrag *e*CO_2xN	1986-2024 2005-2024 2011-2024	1986-2024 2005-2024 2011-2023	Biomass data from the *Phragmites* experiment has been collected and processed. The delay in public posting is deciding on a proper data format for this relatively complex dataset.
Root production	Native *e*CO_2 Native *e*CO_2xN Phrag *e*CO_2xN	1986-2024 2006-2024 2011-2024	1986-2024 2006-2024 2011-2024	These data are straightforward to assess for QC and to format. All data are posted publicly.
Tissue C & N content	Native *e*CO_2 Native *e*CO_2xN Phrag *e*CO_2xN	1986-2023 2006-2023 2011-2023	1986-2019 2006-2019 2011-2019	The methods and instruments for analyzing C, and N have varied over time. We are conducting an in-depth assessment of these data to account for any methods-induced errors.
Tissue $\delta^{13}C$ & $\delta^{15}N$	Native *e*CO_2 Native *e*CO_2xN Phrag *e*CO_2xN	2005-2023 2006-2023 2011-2023	2005-2018 2006-2018 2011-2018	
Porewater solutes	Native *e*CO_2 Native *e*CO_2xN Phrag *e*CO_2xN	2002-2024 2006-2024 2011-2024	2002-2024 2006-2024 2011-2024	The methods and instruments for analyzing porewater chemistry have varied over time. We completed an in-depth assessment of these data for QC and posted them publicly.
Soil Elevation	Native *e*CO_2xN Phrag *e*CO_2xN	2006-2024 2011-2024	2006-2019 2011-2019	are collaborating with the USGS and FWS to adopt community-wide standards for QC and analysis before posting these data.
Temperature, PAR, rainfall	Site Wide	1986-2024	1986-2019	These data are extracted from a nearby weather station operated by the National Weather Service.