Appendix A: Metadata and data for the communities included in the analysis.

CLASS I. DATA SET DESCRIPTORS

A. **Dataset identity:** Species composition and abundance of manipulated animal communities compiled from published literature.

B. **Dataset identification code:**
   Suggested Data Set Identify Code: Manipulated Animal Community Database (MACD)

C. **Dataset description:**
   6,698 records indicated the presence and abundance of animal species, including representatives across trophic groups and size classes documented at 254 sites throughout the world, encompassing a variety of habitats. We accessed peer-reviewed articles, government publications, and theses that were freely available with the Utah State University library subscription and were published in English. We extracted data from articles that reported species-level abundance for a control community and at least one manipulated community. Published data were often summed or averaged over replicates, rather than reporting abundance separately for each replicate. We originally recorded data from 562 terrestrial animal communities representing 91 references. For this dataset, we eliminated communities where > 10% of individuals were not identified to the species level or where the total area sampled for paired control – experiment communities was unequal. We only used data where raw abundance was reported as a summed total for each species or where mean abundance was reported across the replicates, excluding percent cover, biomass, and presence-only data. When mean abundance was reported using "<" symbols (*e.g.* < 0.01) we assumed the value was at the top of that bin. SADs are difficult to characterize when the number of species or total abundance is very low (ref), so we included only communities with $S \geq 5$ and $N \geq 30$. We compare pairs of communities from sites that were sampled at the same spatial scales and at similar temporal scales (*e.g.*, we did not compare data from different seasons or across a time-series) to avoid complications due to differences in sampling intensity or timing.
   The data here represent a single data point each for the control treatment and the manipulated treatment(s) in each study. Data came from a wide variety of sites including artificial experiments (i.e., caged exclosures, habitat modules, nutrient addition) and human-mediated "natural" experiments (*e.g.,* wildfire or controlled burn, logging, grazed plots, pollution). Sites represent all continents except Antarctica, and widely varying terrestrial animal groups (arachnid, insect, herpetofauna [reptiles and amphibians], mammal, and bird).

D. **Keywords:** abundance, community, community structure, composition, experiment, manipulation, environmental change

CLASS II. RESEARCH ORIGIN DESCRIPTORS

A. **Overall project description**

Identity: Species composition and abundance of manipulated animal communities compiled from published sources.
Originators: Sarah R. Supp and S. K. Morgan Ernest
Period of Study: Data compiled in this project were published from 1982 – 2010. The authors intend to continue adding to the database, and encourage others to contribute their data and/or recommendations for additional sources to this effort.
Objectives: To provide data for macroecological analyses of experimental/manipulated community assembly and structure
Taxonomy: Taxonomy follows that reported in each published paper.
Sources of Funding: Utah State Ecology Center Fellowship and Utah State School of Graduate Studies Fellowship

CLASS III. DATA SET STATUS AND ACCESSBILITY

A. Status

Latest update: 29 January 2013 for the final format of all files
Latest Archive date: January 2013
Metadata status: The metadata are complete and up to date.
Data verification: Data quality has been carefully checked as described in class V, section B, below.

B. Accessibility

Storage location and medium: Copies of the latest version of the data file are being stored on the principal investigator's personal computer, Dropbox, and on GitHub (https://github.com/weecology/experimental-rads.git) in Microsoft Excel, Text, and CSV formats
Contact person: Sarah R. Supp, Department of Biology and the Ecology Center, Utah State University, Logan, UT, 84322-5305; sarah@weecology.org
Copyright restrictions: None.
Proprietary restrictions: None.
Costs: None.

CLASS IV. DATA STRUCTURAL DESCRIPTORS

REFERENCE DATA

A. Data Set File

Identity: ref_data_analysis.csv
Size: 47 records, not including header row, 12,774 bytes.
Format and storage mode: ASCII text, comma delimited
Header information: The first row of the file contains the variable names. See section B below for detailed descriptions of the column contents
Alphanumeric attributes: Mixed.
Special characters/fields: If no information is available for a given record, this is indicated by NULL.

Authentication procedures:

A. Variable information

Variable name	Variable definition	Storage type	Missing value codes
referenceID	Unique identifier for each reference; links to sites and references tables	Character	N/A
reference_source	Search engine and keywords used to locate the reference	Character	NULL
authors	Names of authors, in the order given	Character	N/A
year	Year of publication, if published, otherwise null	Integer	NULL
title	Title of publication, if published, otherwise "unpublished"	Character	N/A
source	Journal, government agency, book, or university that published the reference, if published	Character	NULL
vol	Volume number of source, if applicable	Integer	NULL
first_page	First page number	Integer	NULL
last_page	Last page number	Integer	NULL
num_sites	Total number of sites in reference for which data are included	Integer	N/A
num_manips	Total number of unique experiments/manipulations in the reference for which data are included	Integer	N/A

SITES DATA

B. Data Set File

Identity: sites_data_analysis.csv
Size: 253 records, not including header row, 42,011 bytes.
Format and storage mode: ASCII text, comma delimited
Header information: The first row of the file contains the variable names. See section B below for detailed descriptions of the column contents
Alphanumeric attributes: Mixed.
Special characters/fields: If no information is available for a given record, this is indicated by NULL.
Authentication procedures:

C. Variable information

Variable name	Variable definition	Units	Storage type	Variable codes and definitions	Missing value codes
referenceID	Unique identifier for each reference; links to sites and references tables	N/A	Character	N/A	N/A

siteID	Unique numeric code for each site; links to sites and experiments tables	N/A	Integer	N/A	N/A
country	Country in which site is located	N/A	Character	Full names used except for the United States (USA)	NULL
state	State/province where the site is located	N/A	Character	Full names used except for the United States (USA), for which standard state abbreviations are used	NULL
location_details	Details of site location and names given in the original paper	N/A	Character	N/A	N/A
latitude	Latitude where the site is located	Decimal degrees, WGS 84	Fixed point	N/A	NULL
longitude	Longitude where the site is located	Decimal degrees, WGS 84	Fixed point	N/A	NULL
uncertainty_radius	The approximate radius of uncertainty for the given coordinates	kilometers	Integer	N/A	NULL
elevation_min	The minimum, or estimated minimum, elevation of the study area	meters	Integer	N/A	NULL
elevation_max	The maximum, or estimated maximum, elevation of the study area	meters	Integer	N/A	NULL
spatial_extent	Approximate spatial extent of the trapping grids/transects/webs	square meters	Integer	N/A	NULL
first_year	The first year of data collection	N/A	Integer	N/A	NULL
end_year	The final year of data collection	N/A	Integer	N/A	NULL

notes	Miscellaneous notes about the site data	N/A	Character	N/A	NULL

EXPERIMENTS DATA

A. Data Set File

Identity: experiments_data_analysis.csv
Size: 253 records, not including header row, 30,057 bytes.
Format and storage mode: ASCII text, comma delimited
Header information: The first row of the file contains the variable names. See section B below for detailed descriptions of the column contents
Alphanumeric attributes: Mixed.
Special characters/fields: If no information is available for a given record, this is indicated by NULL.
Authentication procedures:

B. Variable information

Variable name	Variable definition	Units	Storage type	Variable codes and definitions	Missing value codes
referenceID	Unique identifier for each reference; links to sites and references tables	N/A	Character	N/A	N/A
siteID	Unique numeric code for each site; links to sites and experiments tables	N/A	Integer	N/A	N/A
experiment	Indicates the type of treatment the data represents	N/A	Integer	0 = control/unmanipulated; 1 = artificial manipulation; 2 = 'natural' or human-mediated manipulation	N/A
experiment_type	Indicates the category of manipulation	N/A	Character	N/A	N/A
experiment_description	Brief description of the experiment or manipulation on the community	N/A	Character	N/A	N/A
replicates	Indicates the number of replicates aggregated or	N/A	Integer	N/A	NULL

	averaged over in the abundance data for the community table				
taxa	Indicates the taxonomic group represented by the data	N/A	Character	N/A	N/A
biome	Indicates if the study is terrestrial or aquatic	N/A	Integer	0 = aquatic; 1 = terrestrial	N/A
habitat	Brief description of general habitat category	N/A	Character	N/A	N/A
raw_abundance	Indicates data type	N/A	Integer	0 = not raw abundance, 1 = raw abundance given	N/A
mean_abundance	Indicates data type	N/A	Integer	0 = not mean abundance, 1 = meaned abundance given	N/A

COMMUNITY DATA

A. Data Set File

Identity: community_data_analysis.csv
Size: 5,990 records, not including header row, 325,430 bytes.
Format and storage mode: ASCII text, comma delimited
Header information: The first row of the file contains the variable names. See section B below for detailed descriptions of the column contents
Alphanumeric attributes: Mixed.
Special characters/fields: If no information is available for a given record, this is indicated by NULL.
Authentication procedures:

B. Variable information

Variable name	Variable definition	Units	Storage type	Variable codes and definitions	Missing value codes
referenceID	Unique identifier for each reference; links to sites and references tables	N/A	Character	N/A	N/A
siteID	Unique numeric code for each site; links to sites and experiments tables	N/A	Integer	N/A	N/A
initial_year	Initial year of sampling for	N/A	Integer	N/A	NULL

	the associated data; some sites may have several years of data presented in aggregate				
family	Family to which the species listed belongs	N/A	Character	N/A	NULL
genus	Genus to which the species listed belongs	N/A	Character	N/A	N/A
species	Specific epithet	N/A	Character	N/A	N/A
id2species	Indicates whether the taxon has been identified to the species level	N/A	Integer	0 = not identified to species; 1 = identified to species; 2 = identified to species pair or morphospecies (i.e., either of two possible species, or c.f.)	N/A
abundance	Abundance data	N/A	Fixed point	N/A	NULL

COMPARISON DATA

A. Data Set File

Identity: comparison_analysis_data.csv
Size: 155 records, not including header row, 3,143 bytes.
Format and storage mode: ASCII text, comma delimited
Header information: The first row of the file contains the variable names. See section B below for detailed descriptions of the column contents
Alphanumeric attributes: Mixed.
Special characters/fields: If no information is available for a given record, this is indicated by NULL.
Authentication procedures:

B. Variable information

Variable name	Variable definition	Storage type	Missing value codes
reference	Unique identifier for each reference; links to sites and references tables	Character	N/A
control_site	Unique numeric code for site; represents the control or unmanipulated community; links to community table	Integer	N/A

comparison_site	Unique numeric code for site; represents an experimental or manipulated site to be compared to the control site; links to community table	Integer	N/A

CLASS V. SUPPLEMENTAL DESCRIPTORS

A. Data Acquisition
The published literature was searched using the databases and keywords listed in the following table:

Database	Keywords	Access
Google Scholar	spider, community, experiment	Online; accessed 12/2011 – 3/2012
Google Scholar	butterfly, community data, experiment	Online; accessed 12/2011 – 3/2012
Google Scholar	butterfly, burn*, community	Online; accessed 12/2011 – 3/2012
Google Scholar	carabid*, community, experiment	Online; accessed 12/2011 – 3/2012
Google Scholar	grasshopper, community, experiment	Online; accessed 12/2011 – 3/2012
Google Scholar	lizard, experiment, community	Online; accessed 12/2011 – 3/2012
Google Scholar	lizard, burn, community	Online; accessed 12/2011 – 3/2012
Google Scholar	bird, experiment, communit*	Online; accessed 12/2011 – 3/2012
Google Scholar	mammal, experiment*, communit*	Online; accessed 12/2011 – 3/2012

B. Quality assurance/quality control procedures:
Each record was entered by the author, and then carefully double-checked against the original reference at a later date.

C. Related material: N/A

D. Computer programs and data processing algorithms: N/A

E. Archiving: Data files and metadata have been archived and are under version control on GitHub (https://github.com/weecology/experimental-rads.git)

F. Literature Cited: Contained in the references table

G. History of data set usage:
Data set update history: N/A
Review history: N/A
Questions and comments from secondary users: N/A