



Data Collection and Use Policies For Plant Community and Trait Data

Plant Ecology Laboratory (PEL)
Department of Botany
University of Wisconsin-Madison
Vers. 1.4, 1 June, 2010



Data Collection and Use Policy

To meet evolving needs of researchers, we have developed this set of policies to guide additions to and use of the legacy ecological data we and our predecessors have collected. These reflect input from many contributors and users of the Plant Ecology Laboratory (PEL) data based in Birge Hall, Department of Botany, University of Wisconsin – Madison. It describes standards that we are seeking to establish, both for contributing new data to data sets, and for accessing and using existing and forthcoming data. We also address associated authorship and acknowledgement issues.

All scientists currently working with PEL data benefit from the fact John Curtis and his students carefully collected and archived their data. These data represent a treasure trove for ecologists seeking to understand community dynamics and the nature and drivers of ecological change. Our goal here is to extend and expand the utility of these data by establishing clear standards for acquiring new data and for using existing data while simultaneously protecting the intellectual rights of those engaged in collecting these data.

We first lay out general principles intended to guide the collection of data for PEL work, its incorporation into these data sets, and proper storage (**A. Principles for data collection, documentation, and storage**). We emphasize the importance of quality control and taxonomic accuracy in the data collected and standards to ensure that these data will be archived carefully to ensure that later researchers will have access to these legacy data. Succeeding sections draw on these principles to provide more specific guidelines for those seeking to either contribute data to the PEL archives (**B.**) or to use these data (**C.**). Finally, **section D.** outlines guidelines for collaborations, authorship, and acknowledgements.

Contacts:

Prof. Don Waller	608-263-2042	dmwaller@wisc.edu
Prof. Tom Givnish	608-262-5718	givnish@wisc.edu

A. Principles for data collection, documentation, and storage

We encourage collaborators working to collect data from the Curtis sites to share data with us and the wider scientific community.

1. **Quality.** Field-collected data must be carefully collected, vouchered, and checked to ensure high quality. All species should be accurately identified in the field with voucher specimens being regularly collected, confirmed, and deposited in the Wisconsin State Herbarium (WIS). These vouchers should include specific locale information (including Curtis site number) as well as a sample for possible genetic analysis (dried in a bag with silica gel).
2. **Description.** All data should be carefully organized for current and future access. Data files should have a clear, standard, and accessible format. The data themselves and their organization / format should be carefully described and fully documented. These meta-data should include a full description of the source and composition of the data as well as how the data files are organized and relate to each other.. For spreadsheets, this should include one or more additional pages with full and complete descriptions of each variable name and descriptions (or links to descriptions) of the methods used to collect the data (e.g., published papers and/or protocols).
3. **Archiving and redundancy.** Data should be duplicated and backed-up, ideally in both paper and electronic forms. We provide safely backed-up permanent storage for all our files and encourage partnering collaborators to do the same.
4. **Data corrections and updates.** All data files must have explicit version numbers and/or dates associated with them. If and when any collaborator finds an error in the data or has any other need to amend a data file, this should be documented in the meta-data and distributed to PEL

B. How to contribute data

We welcome contributions of quality data from Wisconsin plant communities to the PEL databases. In particular, all data must meet the criteria laid out in A. above. We suggest Excel and/or Access files as preferred formats. Please contact one of the researchers noted above and be sure to include full meta-data and associated references and reprints. We are also pleased to share pdf's of suitable papers via the PEL website: <http://www.botany.wisc.edu/PEL>

C. Guidelines for access and use of the PEL data

The collection, proofing, taxonomic synchronization, and organization of these data entailed considerable effort on the part of the primary researchers engaged in these projects. For these reasons, and to ensure careful and legitimate use of these data, we have developed the following policies regarding data access and staged release. These policies are designed to balance the needs of those who generate these data with the wider needs of the scientific community for access to those data. Thus, policies depend on just which data are sought and what use is to be made of them. At all times, we welcome inquiries regarding how these might be used and the potential for collaboration.

1. Curtis historical data on plant community composition

These data were collected and archived by the students and colleagues of John Curtis in the 1940s-1950s. Their original data files include raw data sheets, hand-drawn maps of many sites, and summary data sheets and cards, all in paper form. These data are stored in filing cabinets in a locked room in Birge Hall on the campus of the University of Wisconsin-Madison. Many of these data were also converted to digital format (Lotus 1-2-3 format) by Charles Umbanhower in the early 1990s. He generously makes these data freely available via his website:

<http://www.stolaf.edu/people/ceumb/PEL.html>

These digital files provide excellent tools for exploring the nature and potential utility of these data and for instruction (e.g., sample data sets). Users, however, should be aware that these files contain some omissions and errors stemming from the original recording and entering of the data (e.g., misspellings, unintelligible handwriting, errors in the location information, etc.). In addition, they have not been synchronized with modern taxonomy.

For these reasons, we recommend that colleagues who want to use these historical data for research contact us to make arrangements to access the original data sheets, e.g. by visiting Madison or requesting copies of particular files. We encourage the use of these paper data files but ask that they not leave Birge Hall. We are also pleased to share digital files of those historical data that have been re-coded and proofed. These data include the forested sites re-sampled by folks in the Waller lab and are available in Microsoft Excel or Access format. We hope to post these data publicly soon. In the meantime, please send an e-mail addressed to one of the contacts listed above.

2. 2000+ Resurvey data on community composition and environmental variables

Here, we lay out policies of staged access and release, reflecting the fact that these data are still being generated, processed into databases, and being analyzed for publication. Our general policies are that:

- a) Those that direct the collection and manipulation of a given data set earn a period of exclusive access to those data. The period of exclusive access should be long enough to allow reasonable time for data checking, data analyses, writing-up manuscripts, and

submitting the initial papers based on these data. However, the period of exclusive access does not extend indefinitely.

- b) As papers are published using a particular data set, those data will be deposited in a permanent data bank. All the features of the data (i.e., its metadata) are publicly available at this point, although the full data themselves may not be.
- c) A period of limited access to the full data may follow publication of the initial paper(s) that use a given data set if additional papers that use these data are planned and in process at that time. This period should allow reasonable time for these additional papers to be completed and submitted (typically <24 months from the initial paper's publication and the posting of the meta-data). A current list of working titles and authors for these papers is available via our website, and currently resides at:
http://docs.google.com/View?id=dg2xbhfc_74hjdk2tn8
- d) Following approved periods of limited access (under (c) above), the full data set and metadata will be made available, either via our own PEL web pages or via access to electronic or paper files by another route.

2a. Raw community resurvey data

These are the *raw* form of the contemporary re-sample data. Once that the lead researchers have publish their lead findings based on these data, they make these data available for further and wider use. In some cases, the researchers have a series of papers planned using the same data. Until these planned papers are published, they may prefer to continue to restrict access to these data to their prior and immediate collaborators for an interim period, not to exceed two years. After this window of time, the data will become publicly available. We still ask other users to inform the PEL of their use of these data, acknowledge the source(s) of the data plus the lead researchers involved, and post their papers, or at least a citation, to the list of PEL publications maintained on our website (<http://www.botany.wisc.edu/PEL/>).

These data are available for legitimate academic use. Those to anyone that submits a letter of request, along with this signed data user agreement, to the PI(s) associated with the selected data. Considerable time and knowledge usually goes into editing and synchronizing such raw field data before they are suitable for analyses. We therefore recommend that raw data only be used by persons with suitable experience and knowledge of the mid-Western flora.

2b. Synchronized community data

These are the versions of the data that we use in most of our publications. Most researchers making comparisons between the original PEL data and contemporary re-survey data spend months to years editing and taxonomically synchronizing these data sets to make them ready for fair comparisons and analyses. Taxonomic errors in identification were discovered for some species notoriously difficult to identify (e.g., *Viola* spp., *Carex* spp., *Dryopteris* spp., *Sanicula* spp., etc.). In addition, updated taxonomy may has resulted in new species being recognized now that were not recognized in the 40s-50s, or species recognized then may being lumped. To

address these issues, we made many decisions on whether or not to lump species, sometimes making different decision depending on the nature of the analyses involved (e.g., within or among sites). We document these decisions in the metadata associated with particular files and ask others to do the same. Any researchers with limited experience with the Midwestern flora and these plant communities those limited in the time they can spend making these decisions should seek out collaborators that have this detailed knowledge.

2c. Environmental and other data types

Researchers have collected a variety other data on both species and PEL sites. See the table below for who to contact regarding these data.

1. **Plant traits.** These data include 35 categorical and quantitative traits (species means) for the 235 most abundant forest understory species in Wisconsin.
2. **Landscape data.** Landscape variables for many of our forest sites have derived from detailed analyses of aerial photographs generally for a radius of 2 km from the center of the site. More limited data are available from early (1950 era) aerial photos. Coarser scale 'Wisland' data are also available for radii of 1, 2, and 5 km for many sites.
3. **Soil data.** Soil texture and nutrient content data are available for the majority of sites resampled in the 2000+ period. These may or may not correspond to soil conditions in the 1950s.
4. **Light data.** Light availability (generally % open sky, as estimated from hemispherical photos) is available for many of the resurveyed southern and northern upland sites.

D. Guidelines regarding collaborations, authorship, and acknowledgement

We encourage collaborations and wide use of these datasets. These data are available for all educational purposes and for researchers seeking to test hypotheses that will advance our understanding of these ecological systems. Anyone interested in for-profit use of these data should discuss details of these uses with one of the contacts listed above.

As noted in C.1., the original / historical **Curtis plant community data** are all fully available for use by any researcher or educator. Such users should acknowledge the original researchers who collected and analyzed those data (Table 1).

Table 1. Researchers involved in the initial plant community surveys with J.T. Curtis.

Community	Researchers		
Northern Upland Forests	F.W. Stearns	R.T. Brown	M.L. Gilbert
	W.E. Randall	O. Anderson	
Northern Lowland Forests	H.A. Goder	J. J. Jones	E.M. Christensen
	R.T. Ward		
Southern Upland Forests	R.P. McIntosh	P.W. Whitford	G. Cottam
	M.L. Partch	M. L. Gilbert	W.E. Randall
Southern Lowland Forests	G.H. Ware	R.A. Dietz	
Prairies	O. Anderson	M.L. Partch	P. Green
	L. D. Bard	B.G. Wagner	H.C. Greene
Oak Savannas	J. R. Bray	G. Cottam	A.B. Stout
Pine Barrens	R.T. Brown		
Boreal Forest	P.F. Maycock		

The data generated by later researchers also involve several different plant communities as well as a variety of other kinds of data related to local site and surrounding landscape conditions (Tables 2 and 3). The lead individuals involved in designing these studies and in collecting, checking, and synchronizing these data with the historical surveys are listed. The Tables also list timelines for fully releasing these data for open public access.

Table 2. Contacts for use of community resample data and scheduled release year for plant data.

Community	Research contacts		Release date
Northern Upland Forests	Dave Rogers	Don Waller	2012
Northern Lowland Forests	Matt Bushman	Emmet Judziewicz	2012
Southern Upland Forests	Dave Rogers	Don Waller	2013
Southern Lowland Forests	Sarah Johnson	Don Waller	2014
Dry Prairies	Sarah Kraszewski	Don Waller	2010
Mesic Prairies	Mark Leach	Tom Givnish	TBD
Oak Savannas	Mark Leach	Tom Givnish	TBD

Cedar Glades	Jason Mills	Tim Allen	TBD	
Apostle Islands	Sarah Johnson	Erika Mudrak	Don Waller	2012
Beech Forests	Dave Rogers	Don Waller	2012	

Table 3. Contacts for use of site and species data, and scheduled release years for those data.

Data Type	Research Contacts			Release date
Trait data – Species means	Sarah Johnson	Kathryn Amatangelo	Don Waller	2012
Trait data – Individuals	Sarah Johnson	Kathryn Amatangelo	Don Waller	2013
Soil data – Upland forests	Dave Rogers	Don Waller		2011
Soil data – Lowland forests	Sarah Johnson	Don Waller		2014
Landscape data – Wiscland	Dave Rogers	Voker Radeloff		2011
Landscape data – Hand digitized	Sarah Johnson	Kathryn Amatangelo	Don Waller	2014
Light data	Dave Rogers	Don Waller		2012

For those wishing access to these data in advance of these release dates and/or opportunities for collaboration, please contact the researchers involved to discuss your particular interests and goals. We welcome collaborations and can provide expert guidance regarding the power, utility, and potential limitations of these data. To avoid unwanted overlap or duplications of effort, we also post a list of planned forthcoming papers that are making use of these data at:

http://docs.google.com/View?id=dg2xbhfc_74hzdk2tn8

We ask our colleagues and collaborators to be open and direct about their plans and expectations, e.g., by establishing clear *a priori* agreements for each authors' role in a paper. As a courtesy and to assure good communication, we ask those using these data to join us in posing their titles and authors at this site. Those publishing work based on these data should also acknowledge the appropriate individuals listed in Table 1. Finally, we ask collaborators using these data not to share these data with other collaborators without first requesting permission from the appropriate researchers.

Authorship decisions clearly depend on context. We generally adhere to guidelines followed by many peer-reviewed journals, e.g., everyone who is listed as an author should have made a substantial, direct, intellectual contribution to the work. For example they should have contributed to the conception, design, analysis and/or interpretation of data. Honorary or guest authorship should be avoided.

Acknowledgement Policies

When researchers involved in collecting and developing these data are not themselves authors, it is understood that they and their data sets should be properly acknowledged, e.g.,

Creator(s)
 Title of dataset,
 Year of data publication,
 Dataset identifier (e.g., URL)

Historical Curtis Community Data

The PIs associated with data collection from the 1940s-50s should be acknowledged.

Contemporary community and site data

The PIs associated with collecting and processing the contemporary data should be acknowledged. We ask that you list the granting agency and funding number:

Community	Agency	Grant
Northern Upland Forests	NSF	DEB-9974041
	NSF	DEB-0236333
Southern Upland Forests	NSF	DEB-0236333
	USDA	NRI # 2003-02472
Southern Lowland Forests	USDA	NRI #2008-35320-18680
Prairies	NSF	DEB-9107379
Oak Savannas	NSF	DEB-9107379
	NSF	DEB-9509550
Plant traits	NSF	DEB-0717315
Digitized landscape data	USDA	NRI #2008-35320-18680

Special mention might also be given to the public and private landowners that provided permission for data to be collected on their property.

**Plant Ecology Laboratory (PEL)
Department of Botany
University of Wisconsin-Madison**



Statement of data request

Name:

Affiliation:

Email:

Phone:

Data type(s) requested

- 1.
- 2.
- 3.
- 4.

Intended use and proposed publication topics/papers:

List of any secondary authors, contact information, and their role in the research

- 1.
- 2.
- 3.
- 4.
- 5.

Signature of person requesting data: _____

Date: _____

Signature of PI accepting request for data: _____

Date: _____