Integrating Ecological Data:
Notes from the Grasslands ANPP Data Integration Project

Judith B. Cushing[1], Nicole E. Kaplan[2], Christine Laney[3], Juli Mallett[1],
Ken Ramsey[4], Kristin Vanderbilt[5], Lee Zeman[1]
Jincheng Gao[6], Judith Kruger[7], Carri LeRoy[1], Daniel Milchunas[8], Esteban Muldavin[5]

[1]The Evergreen State College, Olympia, WA 98505, [2]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO 80523, [3] Jornada Experimental Range, New Mexico State University, Las Cruces, New Mexico 88003, [4]Department of Biology New Mexico State University, Las Cruces, New Mexico 88003, [5]Department of Biology, University of New Mexico, Albuquerque, New Mexico, 87131, [6]Department of Biology, Kansas State University, Manhattan, KS 66506, [7] South African National Parks, Scientific Services, Skukuza, South Africa, 1350, [8]Department of Forest, Rangeland, & Watershed Stewardship, Colorado State University, Fort Collins, CO 80523

ABSTRACT

Trends in annual aboveground net primary productivity (ANPP) at regional and global scales are an important component of the structure and function of ecosystems across spatial and temporal gradients in a changing world. Ecologists are interested in conducting cross-site or large-scale integration and analysis of annual ANPP values, but are often hindered by the lack of standard methodologies for data collection, data management practices and detailed metadata documentation across sites. The Grasslands ANPP Data Integration (GDI) project has brought together experts in ecology, information management, and computer science to address the challenges of integrating ANPP data. Together, we have created a centralized database of annual ANPP data and metadata from five national and international Long Term Ecological Research (LTER) grassland sites. The database contains ANPP data at a level of granularity appropriate to each site, but standardizes vegetation species codes and sampling location metadata to facilitate cross-site comparison. This approach is important to local ecologists and information managers as no data are lost, and data can still be aggregated to the proper level of granularity for statistically valid cross-site analysis. The GDI database facilitates transformation, integration, and exploration of site-specific ANPP data, and preliminary cross-site statistical analyses and synthetic research. The GDI team has created processes and tools that will enable future warehousing of ANPP data by streamlining data insertion, update, integration, and standard metadata documentation and species information. This paper presents a description of the GDI data model, data transformation and integration techniques, and quality assurance standards. Lessons learned that might be applicable to other ecological and scientific data integration are also included.

Keywords

Ecological informatics, database integration, ecological synthesis, ANPP, biotic data semantics

1. Introduction

The GDI (Grasslands ANPP Data Integration) project is a joint effort among ecologists interested in annual aboveground net primary production (ANPP), Long Term Ecological Research (LTER) Information Managers, and computer scientists interested in data integration and semantics. ANPP datasets represent core areas of research in many programs, including the Long Term Ecological Research Network,[1] Oak Ridge National Laboratory,[2] and the Global Terrestrial Observing System,[3] and are an important measurement for assessing ecosystem structure and function, biodiversity and ecosystem services, including carbon sequestration (Parton et al. 1995).

The influence of changing climate on ANPP is a question of great interest to ecologists. Knapp and Smith (2001) assessed the temporal dynamics of ANPP across eleven LTER sites in the United States, and suggested that grassland ANPP

---

[1] http://www.lternet.edu/coreareas/coreintro.html

[2] http://daac.ornl.gov/NPP/ html_docs/npp_stat.html

[3] http://www.fao.org/ gtos/NPP.html

will be very responsive to future climatic changes. Their analyses were conducted with total annual ANPP values from each site. More refined analyses of ANPP values might also be of interest; for example, ANPP broken down by plant species and life forms would help assess community and population responses to variability in precipitation and help predict how grasslands might respond to global change phenomena. Synthesizing long-term datasets of species or life form level ANPP data from different regions and ecosystems is a critical first step towards conducting this research. However, ecologists and information managers across different LTER sites have in the past experienced challenges to integrating ANPP data from multiple sources.

Given the importance of ANPP to the ecological community, computer scientists, information managers, and ecologists both within and peripherally connected to the LTER and ILTER networks initiated a project to integrate ANPP data from several sites so that synthesis research across sites could be more easily conducted. Project objectives were to make the integration process more efficient, enable cross-site analysis, conserve fine levels of data granularity, and eventually accommodate ANPP data from sites outside the grassland biome, as well as other grassland sites. A reliable and useful integrated data product requires documenting the data as they are loaded, determining a statistically valid level of comparison, and transforming the data into a standardized format. A long term sustainable data warehouse, however, is not feasible without semi-automated tools for data insertion, integration, documentation, and validation. It is also inadvisable to take on such a project without advice from the ecologists familiar with the data to be incorporated. Our collaboration among ecologists (responsible for experimental design and data analysis), information managers (accountable for data access), and computer scientists (responsible for producing technical solutions) was thus important from the onset of the project.

Here, we present preliminary products and results of the GDI project. We discuss the design, development, and implementation of a centralized database and the tools created to support the integration and importation of these disparate datasets. We also share the lessons that we learned in the integration process, which may be applicable to other ecological and scientific data integration efforts.

2. The Grasslands Data Integration (GDI) Project

Our initial goal was to combine ANPP data from three LTER sites (Jornada Basin [JRN]; Shortgrass Steppe [SGS] and Sevilleta [SEV]), and to include two other grassland sites if possible (Konza Prairie [KNZ] and Kruger National Park ILTER in South Africa [KRU]). The data from each site were collected under different experimental methods under different climate and vegetative conditions, were described using different semantics for experimental units, species names, and ecosystem types, and were made available to the project in incompatible syntactic formats. The large number of records from each of the sites led us to explore issues of identifying and fixing data quality problems, and highlighted the need for new tools that would enable both data producers and consumers to explore the data sets in a multi-site database. Exploring data in a multi-site database exposed data quality problems at the site level and raised questions about changes in data collection or reporting over the life of a data set, which we think could lead to improvements in the quality of the data warehoused in the GDI database. In the following sections, we describe activities that took place during the project, the solutions developed to address each problem encountered, and methods used to validate the process and tools that we developed.

2.1 Data Collection and Information Management at the Site Level

Methods for ANPP field data sampling are designed independently by ecologists at each site, and typically change over time. Data are collected by field technicians, sometimes processed or aggregated by the responsible investigators, and then placed into a database local to an LTER site and validated by an information manager. Where observational data are coded by species, a table of species codes and information about the species they represent is also maintained and bundled with the ANPP data.

Previous cross-site integration work by ecologists (e.g., Knapp and Smith, 2001), involved manually acquiring data from each site and combining these data into a new, single-purpose and static database. Even today, most current ANPP data are kept in a local site database and are available upon request in a file format commonly used for exchange between database packages, most often comma-separated-value (CSV) format. The schema of these tables vary greatly among sites and preclude a simple path to automated integration. A lack of semantic metadata regarding the experimental design, how sample replicates should be grouped into statistically-relevant experimental units, and details about how best to aggregate ANPP values can obfuscate the comparison of seemingly-equivalent data between sites. Even species information, which is almost universally gathered, is difficult to integrate because of the use of site-specific codes. There are few processes available for managing species tables, but the accuracy of this information is critical for data analysis.

2.2 Cross-Site Data Integration Issues

In integrating data from multiple sites, we faced various challenges, such as differences in data granularity (whether data were collected by species or growth habit) and differences in site-specific experimental design. In some cases, as at the

SGS and KNZ sites, ANPP is measured directly by harvesting total standing crop biomass (Milchunas et al. 1994). At other sites, e.g., SEV, JRN, and Kruger, ANPP is estimated based on species-specific regression relationships between biomass and plant volume or coverage (Muldavin et al. 2008, Huenneke et al. 2002). In still other cases, ANPP is estimated from remotely sensed images and the use of indexes (Paruelo et al. 1997). These different methodologies for collecting ANPP data are conducted at various spatial scales (e.g., one-quarter square meter vs. hectare), and at different temporal (e.g., seasonal or annual) and biological (e.g., species or life form) resolutions. Such differences are common among measures of biotic data.

Each of the sites participating in this project bases ANPP on field measurements, but each has a different number of experimental units at which they collect data. A site might have many plots and each plot many sub-plots; sub-plots might be even further subdivided. We called the lowest level of sub-plots where data were collected the "experimental unit". For the integrated database, however, the responsible ecologists emphasized that one should not analyze values at this level but instead aggregate site data at the experimental unit level to a unit appropriate for ecological analysis. We called this level the "sampling unit", and for our data validation and preliminary analyses averaged ANPP across comparable experimental units, and reported ANPP at various "locations" (sampling units) across sites. Table 1 shows how sampling units and experimental units varied for the sites we worked with. Subsites in the database are distinct because they are considered to have natural differences in ecological characteristics. Some sites perform experiments and certain plots with the same ecological characteristics might be differentiated as a control plot or treatment plot (e.g., burns, livestock grazing).

Most data within the GDI database have been collected and integrated at the species level, but sites typically use different codes to record species level data. In these cases, observational data are coded by site-specific species codes, and a table or list of those codes and information about the species they represent (the species table) is available. The USDA PLANTS database is used as the cross-site species table, and we built a general-purpose tool, *Specifik*, to map each site's species codes to the USDA PLANTS codes. Of course, the PLANTS species codes are applicable only to species typically present in the U.S., so adding an international site requires updating the codes database to cover species or plant forms not present in the U.S.

3. Methods and Techniques

While constructing a database of one site's ANPP data is relatively straightforward, merging many sites' data at the observation and species levels, and properly combining experimental units into sampling units, is dauntingly difficult and time consuming. As a result, few ecologists to date have analyzed ANPP across sites and those who have typically have limited the granularity of their analysis, leaving many potentially important variables among sites unexplored. For example, the EcoTrends database aggregates ANPP data into total annual values per site, vegetation type or treatment and does not

Site	Sampling Method	Times Measured per year	Years of Data	Number of Vegetation Types or other Relevant Treatments	Number of Sub-Sites	Number of Sampling Units (replicates)	Experimental Units* in each Sampling Unit (plots per rep)	Total Number of Experimental Units (plots)
Kruger National Park (Kruger)	Regression relationships	1	17	35	35	35	9-41	315-1435
Konza Prairie (KNZ)	Biomass harvest	2	5	1	1	2	40	80
Jornada Basin (JRN)	Regression relationships	3	17	5	15	15	49	735
Sevilleta Wildlife Refuge (SEV)	Regression relationships	3	8	3	3	15	16	720
Shortgrass Steppe (SGS)	Biomass harvest	1	23	1	6	3	5	90

Table 1. Site-Specific methods and number of years of data. The vegetation types sampled and sampling units at each site were determined by site ecologists; these determine replicates for statistical analysis. The number of experimental units is the number of plots or quads within each sampling unit or replicate.

* The GDI database, as shown in Figure 1, refers to Experimental Unit as "location".

differentiate by species.

We identified three key areas for the GDI project: schema design, a robust process for integrating data, and species integration. The first step was to develop a data model that represents the complexity of ANPP data across sites and that is simple to explain and use; the second to write scripts to process each site's dataset as this processing had to be repeated several times – to correct data errors in the original datasets and to add additional years of data as they became available. These scripts will be used in the future as data for additional years and additional sites become available. The third step was to develop a program to map site-specific species codes to PLANTS codes.

3.1 The GDI Data Model

Because of data quality issues (primarily data type and referential integrity errors) identified during the first data loads and integration, we determined that the best design would be a single, centralized database to be updated periodically, rather than a virtual database or index from each LTER site's online ANPP database.

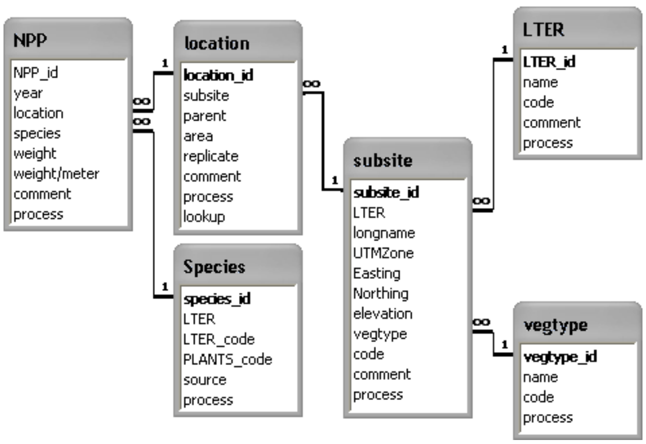

Figure 1. GDI Schema Design

We chose to design the ANPP integrated database so that the NPP table is its primary focus. Each row of that table contains information on how much NPP was collected per year, for what species (or growth habit), and where and when it was collected. Figure 1 shows this schema, as generated by MS Access. Each table name appears in grey highlight above a list of attributes (rows) for the table; the cardinality of all relationships between tables is many (∞) to 1, as shown by solid lines connecting tables. All tables contain a unique identifier (primary key), which is first in the list of attributes and is the name of the table followed by "_id". Many tables also include a comment field, which at this point is free text, and a process field, which is a set of numbers corresponding to processes listed in a metadata document, so that it is possible to recreate exactly the queries and scripts that affected each row to process the data.

A record in the **NPP table** contains the following attributes and relationships to other tables: a unique identifier NPP_id, Year when NPP was measured, location of the measurement(s) as a foreign key relating to the location table, code for species or plant type as a foreign key relating to the Species table, weight of NPP as measured (units differ per site), and weight of NPP adjusted to grams per meter squared. Thus, from this table, a researcher can determine observed ANPP (in a common unit, grams of biomass per meter squared) within a given location (an area defined on a per-site basis, referred to as the sampling unit) over one calendar year. While growing season is used by many sites to report NPP, we did not report data at this level of time granularity since we could not compare growing season across sites as they differ.

The **Location table** contains information about the experimental unit to which the measured NPP value is associated. In particular: subsite relates to an area of ecological interest for which biome and climate data are available in the Subsite table, area is the size of the (plot) location in meters squared, replicate denotes membership in one of the sampling units within the subsite; each sampling unit contains 5-50 plot-locations. lookup is used for validation during the integration process; it is not meaningful to the ecologist. parent is not currently used; before clumping many experimental units into one subsite, location was organized into a hierarchy of plots, subplots, subsubplots, etc.

The **Subsite Table** describes the subsite to which (plot) locations belong. It identifies the [LTER] research site at which the measurements were taken, and is typically identified at the research site by a name and code. The Subsite table also contains geographic coordinates and information (UTMZone, Easting, Northing, and elevation) as well as vegtype (vegetation type or treatment of interest) for comparative analysis within the larger LTER research site. vegtype (vgetation type) is a coded value described in the Vegtype table, which contains a full name and code for the vegetation type. Sampling units (replicates) and experimental units (location) are further explained in Section 3.2, where we articulate the data transformation process from site-specific data to the integrated database.

The **Species table** relates the standardized USDA PLANTS species code in the NPP table to a site-specific species code. Each PLANTS species code could correspond to multiple LTER site-specific codes at different sites, and (over time) any one LTER code could relate to more than one PLANTS code.

The major challenge we identified in loading data in the GDI schema was cleaning up referential integrity for species codes and deciding which USDA PLANTS code to associate with each site-specific code. That conversion process prompted us to create a tool which may be useful for broader work with species-coded data in the botanical domain. This tool is explained in the Section 3.3.

3.2 Data Transformation and Integration

Prior to adding each site's data to the integrated database, we carried out two steps: *transformation* to the NPP observation table format and *integration* with data from other sites. The separation of these steps allowed easy re-integration of data when changes occured in the experimental design, as location, sub-site and species code tables were loaded only once. Each LTER provides biomass data, either calculated or directly measured. If it is not in a yearly format, seasonal data is combined into years. These transformations are handled with ad-hoc scripts specific to each site's data. Future data submissions will be required in the "observation" format and validated and processed with a common script.

Once each dataset is formatted as a series of NPP observations, it must be integrated into the central database with enough contextual data to allow for meaningful statistical comparison across research sites by *location* or by *species*. Individual plot areas range from a quarter of a meter squared to two hundred meter line transects, so plots are not directly comparable. Each individual plot (location) is assigned to a statistical sampling unit, which is an aggregation of co-located plots with similar soil and vegetation types to allow statistically meaningful analysis. Individual sampling units are designated by the ecologists as containing enough data to be statistically meaningful and contain between five and fifty plots. In the GDI database, we call the sampling unit the replicate. In addition, each plot (location) is assigned to a subsite within the LTER – an ecologically meaningful geographic designation for which biome, geographic location, and climate data are available. This contextual data allows aggregation of data (beyond species or plant code) for analysis at three additional levels: research site (LTER), subsite, or sampling unit. Mechanical parsing of most species information proved successful, as species information is generally provided in CSV format, though occasional human intervention was required where a site's species table was not syntactically self-consistent.

3.3 Species Code Conversion

Site-specific species tables are typically not congruent with USDA PLANTS codes since most sites use site-specific codes. The USDA PLANTS codes, on the other hand, use rules defined by the International Code of Botanical Nomenclature for assigning botanical names. Although some taxonomists have defined new standards for botanical naming in line with information science best-practices[4], binomial names remain the dominant standard and are the only system of plant taxonomy that is generally-accepted. Complicating integration, even binomial names used correctly today may become inaccurate over time. Recent years have seen trends towards plant reclassification based on DNA evidence, making some names obsolete. If it is discovered that a species has been incorrectly assigned to a genus then the genus given as part of its binomial name would lead to inaccurate analysis between genera.

The USDA PLANTS database contains placeholder markers for those obsolete names, and encodes information about which currently-accepted name is synonymous. We decided that the best practice would be to use the USDA PLANTS species code in our database, and create a table that provides the correspondence of that code to the name/code used by the LTER site, and thus to make use of the USDA PLANTS synonym infrastructure. The USDA PLANTS codes have been established as the standard for identification of plants species for the four U.S. LTER sites[5] in the GDI database, and the correct USDA PLANTS code must be determined to avoid falsifying species coded data. Some information of interest for many ecologists, such as carbon pathway, seems to be absent from USDA PLANTS, and must be maintained externally. This presents problems as we found no single authoritative source for such information. However, some information not kept by many sites, but useful in analysis and reporting such as threatened, endangered and invasive status and common name, is available from the USDA PLANTS database.

Just as data errors are common, spelling mistakes are frequent enough in binomial names for species at LTER sites to have required a manual process at many points in the conversion. A manual (or at least interactive) process is likewise highly desirable as it brings species errors to the attention of the information manager and the ecologist. The responsibility to make a determination of correct equivalent USDA PLANTS code in such situations is not a technical decision, and will rest with the contributing site. To facilitate this process we have developed a web-based application, *Specifik*, described in Section 4.2.

---

[4] See Phyocode (http://www.ohiou.edu/phylocode/index.html and http://www.ohiou.edu/phylocode/PhyloCode4b.pdf) and Biocode http://www.bgbm.org/iapt/biocode/.

[5] The non-U.S. iLTER site in our database, Kruger, did not record ANPP by species, so the fact that the USDA PLANTS codes do not cover South Africa was not an issue. If we include other ANPP datasets for non-U.S. sites that provide ANPP by species, we will need a species table similar to PLANTS for those sites.

4. Results and Discussion

The GDI database, as of March 20, 2008, contained 113,500 distinct NPP observations from five sites, and was 73 MB in size. Because the database creation process brought data errors to light, the finished database contained fewer errors than the source data. Normalization highlighted errors of absence: data missing from certain plots or certain years and blank species or mass data. Integration highlighted errors of context: species with entries in the data but not the species tables, data from mislabeled or nonexistent plots, or plots with bad coding information. In addition, some basic validation checks removed observations with negative or zero weight, and observations for years outside the known span of the experiments. Questionable data were resolved by conversations with data providers. As we go to press for this paper, the GDI database is not yet released for general download, but it can be requested from the Sevilleta LTER Information Manager.

4.1. Preliminary Validation

The database allows comparison of ANPP between LTER sites and vegetation types, and we found that the discipline of data integration and preliminary scientific analysis can lead to improving data quality for better subsequent analysis. For example, an early statistical comparison of three LTER sites, JRN, SGS and SEV erroneously suggested that Jornada was significantly more productive than similar grassland sites despite the fact that it is the warmest and driest. This led the Jornada site to update its regressions and helped to emphasize the fact that the prevalence of a single species (*Yucca elata*) at one site can influence cross site analysis.

4.2 *Specifik*

Specifik is a web-application that we wrote to assist a user in adding USDA PLANTS species codes to a species table. Given a CSV-formatted table of species information, it asks the user a few simple questions to determine the dataset's taxonomic ontology. It then asks users to select a USDA PLANTS code for each species from a list of likely matches. If a user is unsure which alternative is correct, the tool allows the user to defer the assignment, and to provide a manually processed code at a later time. Once codes have been selected for every species, users are given a copy of their species code table with USDA PLANTS codes added. We hope that by contributing an easy-to-use tool to facilitate this process, more sites will maintain USDA PLANTS codes in their own databases, facilitating future work by us and others who hope to integrate data and analyze species-coded data. *Specifik* defers to the USDA PLANTS database on issues of taxonomy, as there are taxonomists and biologists who ensure that the species information therein is current and correct. The USDA PLANTS database also contains most species metadata documented by each LTER site, such as genus, author, family and form, but that information may be incomplete for many species. *Specifik* is open source and freely available for download from the internet.[6]

5. Lessons Learned, Conclusions, and Future Work

Perhaps the most critical lesson learned from this project is the fact that ANPP is but one of a class of ecological data dubbed *response variables*, e.g., measurements of primary productivity (NPP, biomass, cover, etc.) or diversity (species richness, species diversity, community dynamics, etc.). While useful in and of themselves, ecological response variables are significantly more useful if environmental drivers, or context variables, are available in such a way that correlations can be drawn between environmental drivers and responses. To model change in ANPP over time, contextual data is required. Climate is a key driver of year-to-year changes in production, so models of change in ANPP over time would include climate data such as Palmer Drought Severity Index (PDSI), precipitation (e.g., growing season vs. non-growing season, totals, extreme events), temperature data (e.g., growing season vs. non-growing season, minimums, maximums). Other important contextual data might include landform for each location (e.g., slope, aspect, soils, and elevation), and land history (e.g., grazing and fire). Soil moisture, and atmospheric and soil chemistry could also impact ANPP. Links between a GDI database website and data stores such as EcoTrends,[7] ClimDB/HydroDB,[8] and the National Atmospheric Deposition Program[9] could facilitate analyses of queried NPP data with contextual data.

Response variables (aka biotic data) are significantly more complex than environmental drivers such as precipitation or temperature which have a history of data recording and reporting, but both are needed for answering important ecological questions (Peters et al, 2008). Secondly, we discovered that the process of putting complex data into a database can go a long way to improving data quality; analysis of cross-site data integration products provides additional quality control. Of course,

---

[6] http://alala.evergreen.edu/~mallettj/specifik/

[7] http://www.ecotrends.info

[8] http://www.fsl.orst.edu/climdb

[9] http://nadp.sws.uiuc.edu/

the additional effort required to standardize codes (e.g., for species) and semantics (e.g., sampling units or plot granularity, vegetation type) is significant.

Thirdly, interdisciplinary collaboration and teamwork during the data integration design process are key to success. Without any one of our stable three-prong foundation (ecologists, information managers, and computer scientists), this project could not have succeeded. The challenges to make the GDI successful and useful required a team approach with communication among ecologists, information managers and computer scientists. Fourthly, we assert that once a general schema has been developed and well tested, and easy to use tools and processes established, the burden of uploading data to a data repository, such as the GDI, should lay with the site rather than with a central curator. While a curator must take ultimate responsibility for including a certain site's data in a repository, only the local information manager has adequate in-depth understanding of the data to perform the required data transformations.

In conclusion, we suggest that the GDI project has created a sustainable, streamlined system for transforming, integrating, validating and analyzing ANPP data. It also required time and attention to identification of data quality issues during data transformation and integration, and a fairly deep understanding of how the data would eventually be analyzed. This collaboration also resulted in a species code standardization tool that can be used for other synthetic research projects.The GDI project has helped information managers become more aware of LTER synthesis projects, such as EcoTrends, as ANPP represents a core area of research for the LTER Network and a long-term ANPP dataset would serve as a foundation for cross-site and synthetic research future scientific endeavors (Baker, BioScience). Standardized, structured, centralized and accessible repositories of such data facilitate the work of both information managers and ecologists.

6. Acknowledgements

NSF Canopy Database Project (NSF Grants: DBI-0417311, DBI-0319309), JRN-LTER (NSF Grant: DEB-0080412)**,** KNZ-LTER (NSF Grant: DEB-0218210), SEV-LTER (NSF Grant: DEB-0080529), and SGS-LTER (NSF Grant: DEB-0217631). We also thank Alan Knapp, Debra Peters, Mark Servilla, Susan Stafford, Bob Waide, and the LTER and Kruger National Park field crew members and technicians, for helpful suggestions or work on this project.

7. References

Baker, K.S., B.J. Benson, D.L. Henshaw, D. Blodgett, J.H. Porter, and S.G. Stafford. 2000. Evolution of a multisite network information system: the LTER information management paradigm. BioScience, 50:11, pp 963-978.

Huenneke, L.F., J.P. Anderson, M. Remmenga & W.H. Schlesinger. 2002. Desertification alters patterns of aboveground net primary production in Chihuahuan ecosystems. *Global Change Biology* 8:247-264.

Knapp, A.K. and M.D. Smith. 2001. Variation among biomes in temporal dynamics of aboveground primary production. *Science* 291:481-484.

Knapp, A.K., M.D. Smith, S.L. Collins, N. Zambatis, M. Peel, S. Emery, J. Wojdak, M.C. Horner-Devine; H. Biggs, J. Kruger, S.J. Andelman. Generality in Ecology: Testing North American Grassland Rules in South African Savannas. *Frontiers in Ecology and the Environment* 2(9):483-491.

Milchunas, D.G., J.R. Forwood & W.K. Lauenroth. 1994. Productivity of long-term grazing treatments in response to seasonal precipitation. *J. Range Manage*. 47:133-139.

Muldavin, E.H., · D.I. Moore, S.L. Collins, K.R. Wetherill & D.C. Lightfoot. 2008. Aboveground net primary production dynamics in a northern Chihuahuan Desert ecosystem. *Oecologia* (155):123-132

Parton, W.J., J.M.O. Scurlock, D.S. Ojima, D.S. Schimel, D.O. Hall, M.B. Coughenour, E. Garcia Moya, T.G. Gilmanov, A. Kamnalrut, J.I. Kinyamario, T. Kirchner, S.P. Long, H-C. Menaut, O.E. Sala, R.J. Scholes, and J.A. van Veen . 1995 . Impact of climate change on grassland production and soil carbon worldwide. *Global Change Biology*. 1: 13-22

Peters, D.P.C., Guest Editorial, Frontiers in Ecology and the Environment, Special Issue on Connectivity and Continental Scale Research, June 2008.

Paruelo, J.M., Epstein, H.E., Lauenroth, W.K., & I. C. Burke. 1997. ANPP Estimates from NDVI for the Central Grassland Region of the United States. *Ecology* 78(3): 953-958.

USDA, NRCS. 2008. The PLANTS Database (http://plants.usda.gov, 29 April 2008). National Plant Data Center, Baton Rouge, LA 70874-4490 USA.