

The MOR2 Database: Building integrated datasets for social-ecological analysis across cultures and disciplines

**Melinda J. Laituri^{1,2,3,4}, Sophia Linn^{2,5}, Steven R. Fassnacht^{1,2,3,6,7},
Niah Venable^{8,9}, Khishigbayar Jamiyansharav^{10,11}, Tungalag
Ulambayar^{10,12}, Arren Mendezona Allegretti^{14,15}, Robin Reid^{1,3,16}, Maria
Fernandez-Gimenez^{10,13}**

¹Ecosystem Science and Sustainability, Colorado State University (CSU), Fort Collins, CO, USA, 80523-1476

²Geospatial Centroid at CSU, Fort Collins, CO, USA, 80523-1019

³Natural Resources Ecology Laboratory, CSU, Fort Collins, CO, USA 80523-1499

⁴<Melinda.Laituri@colostate.edu>

⁵<Sophia.Linn@colostate.edu>

⁶Cooperative Institute for Research in the Atmosphere, Fort Collins, CO, USA 80523-1375

⁷<Steven.Fassnacht@colostate.edu>

⁸EASC-Watershed Science, Colorado State University, Fort Collins, Colorado USA 80523-1482 ⁹<niah.venable@gmail.com>

¹⁰ Forest and Rangeland Stewardship, CSU, Fort Collins, CO, USA, 80523-1472

¹¹<jkhishig@gmail.com>

¹²<tungaa.sg@gmail.com>

¹³<Maria.Fernandez-Gimenez@colostate.edu>

¹⁴Graduate Degree Program in Ecology, CSU, Fort Collin, CO, USA, 80523,

¹⁵<amendezona@gmail.com>

¹⁶<Robin.Reid@colostate.edu>

ABSTRACT

This paper describes the construction of a complex database for social-ecological analysis in Mongolia. As a National Science Foundation (NSF)-funded Dynamics of Coupled Natural and Human (CNH) Systems, the Mongolian Rangelands and Resilience (MOR2) project focused on the vulnerability of Mongolian pastoral systems to climate change and adaptive capacity. To study this phenomenon, our team is made up of a group of hydrologists, social scientists, geographers, and ecologists collecting data across the Mongolian landscape over three years. This dataset is unique in that it captures multiple types of field data: ecological, hydrological and social science surveys; remotely-sensed data, participatory mapping, local documents, and scholarly literature. We describe the content, structure, and organization of the database and explain the development of data protocols and issues related to access and sharing. Descriptions of data analysis are included to demonstrate the utility of the database as well as its limitations. We conclude with a description of the challenges in creating a cross-cultural and multi-disciplinary database and lessons learned.

Keywords: database, interdisciplinary, Mongolia, social-ecological analysis

INTRODUCTION

This paper describes the database created for the National Science Foundation (NSF)-funded Mongolian Rangelands and Resilience (MOR2) research project. Comprehensive, complex databases are needed to address complex problems. The MOR2 database is designed to provide researchers with data for assessing socio-ecological aspects of climate change and herder adaptations in Mongolia. Data were collected to examine different management activities and outcomes on Mongolian rangelands in places that adopt community-based rangeland management (CBRM) and those adhering to traditional herding practices (non-CBRM). The MOR2 database is comprised of multiple types of data, organized into different thematic datasets and databases, which have been gathered by an interdisciplinary, multi-cultural research team (e.g., hydrologists, ecologists, geographers, and social scientists; US and Mongolian scientists) using different data collection methods, scales and units of analysis, and analytical techniques.

The database has evolved organically over the course of the five-year project. Database discussion and development has been ongoing, adaptive, and reactive to data collection activities. A central component of the database is the multiple types of field and secondary data collected by sub-teams (Table 1). Note that the research teams are fluid, with some members participating in multiple aspects of the project, strengthening the interdisciplinarity of the research.

This paper provides an overview of the MOR2 database and discusses some key challenges and lessons learned. Data content and collection are described for the study sites across Mongolia. For thematic parts of the database, data collection, access and sharing protocols have been created. Examples of preliminary analyses and challenges are presented. The paper concludes by highlighting the creation of a comprehensive database built into a simple storage structure of multiple thematic folders. Central to the database design is a common set of geographic study sites, which created the foundation for developing coded linkages between datasets. This database was constructed prior to many of the technical advances made in web-based data sharing and cloud technology for data repositories. These new options will be considered for archiving and long-term data maintenance.

STUDY SITES, DATA CONTENT AND COLLECTION

The MOR2 study sites span four ecological zones across the Mongolian landscape: mountain and forest steppe, steppe, Eastern steppe, and desert steppe. In each zone, paired *soums* (counties) were selected with and without formal CRBM groups for a total of 36 *soums* (18 with CBRM and 18 non-CBRM). Within each *soum*, 3-9 herder groups (CBRM) or traditional neighborhoods (non-CBRM) were selected for study. Geospatial data were used to visualize sample study locations using a Geographic Information System (ESRI, ArcGIS Version, 10.2) (Figure 1). Extensive fieldwork took place over four field seasons where ecological, social, physical, and boundary data were collected. Each research team collected data using a mixture of tools including global positioning systems (GPS); photographs; digitally recorded and paper survey questionnaires, focus groups and interviews for social and herder observation data; and ecological and hydrological sampling protocols. Teams collected existing *soum*- and group-level data and documents, government livestock and population statistics. A nested social database includes three social tiers: 1) household questionnaires, 2) organizational profiles of CBRM and non-CBRM groups that synthesize qualitative data collected through focus group discussions and leader interviews, and 3) *soum*-level survey from focus groups, and demographic and socio-economic data. Social data were collected from 142 pastoral groups and 706 member households. Ecological data include vegetation cover, biomass, plot level soil pit descriptions, soil surface and environmental characteristics collected at three different grazing distances from 143 winter camps in four different ecological zones.

Physical data include meteorological and stream flow measurements collected in selected study sites to examine sub-watershed stream dynamics. Precipitation, temperature and hydrologic data collected from the Global and Mongolian governmental databases were used to derive national level, point and interpolated climate datasets. Tree cores were collected from two sites for comparison and analysis with existing datasets stored in the International Tree-Ring Data Bank. Remotely-sensed satellite data spans Mongolia and includes data with multiple resolutions (AVHRR at 1km and MODIS at 250m). Participatory mapping resulted in hard copy maps, digital maps, global positioning system (GPS) points, and interview data used to analyze social, ecological, physical and administrative boundaries. In addition to field-collected data and existing statistics and data sources, the database includes additional GIS databases for the country, government and NGO reports, and scholarly literature.

In the field, data were collected on datasheets for both ecological and social information and entered into digital databases after the conclusion of fieldwork each year. Two Microsoft Access databases, with input forms (for the household questionnaire and organizational profiles) were developed and data from hard copy forms were entered into these databases. Quantitative data from household questionnaires and organizational profiles are stored in two MS Access databases and exported to SPSS for statistical analysis. Ecological data were entered into the Database for Inventory, Monitoring & Analysis (DIMA) developed by the Agricultural Research Service (ARS), at the Jornada Experimental Range, New Mexico (<http://jornada.nmsu.edu/monit-assess/dima>). DIMA provides automated analysis routines for vegetation, biomass, and soils indicators. A separate DIMA dataset was developed for each ecological field season (2011, 2012, 2013) to facilitate consistency of data entry and quality assurance. The ecological data are stored in both DIMA and in exported Excel files and imported into SAS or SPSS for analysis. Quality assurance and control procedures (QA/QC) were established for transferring data from hard copy forms to DIMA. All data were documented with metadata.

STRUCTURE, ORGANIZATION, TYPES, AND MANAGEMENT

Data discussions were ongoing throughout the project creating a flexible and adaptive structure in which to manage and store the data. These data are housed at Colorado State University on a shared drive within the Warner College of Natural Resources. Data are protected and backed up on external hard drives and on the CSU-WCNR computer servers. Data are organized thematically within a series of folders. Versioning of datasets follow a naming convention based on dates (DD/MM/YYYY) and new folders created by field season for the ecological data. The social data are organized by data type: household questionnaire, organization profile, and *soum* data. Spatial data obtained from other sources were assessed and organized thematically for reference use and spatial analysis.

To create linkages and relationships between the distinct ecological, social and physical databases, a numbered coding system was established to create an unambiguous label for each administrative unit (*soum* and *aimag* or province), organization, household, winter camp and ecological plot (Table 2). The coding system enables cross-referencing and merging of data across the various databases, using a spatially explicit hierarchy, to facilitate integrated analysis of social, ecological and physical data. The coding system is used in each of the project's separate databases. The code is the "key" that links the spatial data, the ecological data, and the social data (both the household and organizational database).

DATA PROTOCOLS AND ACCESS

Formal written protocols were created for all data collection activities to ensure consistency. Protocols for the database were also created, including the development of internal metadata standards. The metadata for the database has been created through a series of dynamic README files that are located within each folder. Team members can access the different data in the folders, but any changes or updates are recorded in the README file within that folder. This essentially creates a distributed, living data dictionary. However, it is incumbent upon team members to maintain these README files. Additionally, as new data have been added and transferred to the latest version of the database, an overall spreadsheet is maintained to track these data changes. All thematic databases have a data steward to oversee these processes.

A Data Ownership and Use Protocol provides guidance for database use and encourages young researchers to develop their research skills. Each researcher must apply to use the data and explain their research questions, methods, and analysis. The protocol provides for and ensures oversight of analytical approaches, scientific peer review, and appropriate referencing of credit for data development. The process is meant to be flexible and to further the adoption of sound scientific approaches in using a unique database.

DATA ANALYSIS

The research teams are conducting numerous types of analyses using these compiled MOR2 datasets. GIS layers and physical landscape data provide the basis for analysis to create derived data products. For example, flow analysis, stream networks, and digital elevation models were created for Khangai Mountain region river basins using the suite of ArcGIS Spatial Analyst tools (ESRI, ArcGIS 10.2). Household-level data are aggregated and combined with organizational profile data to analyze social outcomes of community-based management. Ecological field data are analyzed to assess effects of grazing gradients and community-based management, and combined with remote sensing to compare patterns in ground and remotely-sensed data. Participatory maps are coded using visual grounded theory to identify a typology of boundary types associated with herder management practices. Next steps include merging full ecological and social databases to examine reciprocal relationships between social and ecological functions and how these differ between CBRM and non-CBRM communities.

CHALLENGES

The MOR2 database is diverse and complex, which has inherent challenges. MOR2 illuminates cross-cultural challenges that include language barriers, differences in scientific approaches between the US and Mongolian researchers, communication barriers due to different cultural norms and differing political and social contexts. These issues are beyond the scope of this paper but are embedded in the praxis of trans-disciplinary and multicultural research. The MOR2 project reveals the complexity of transdisciplinary, trans-cultural research and database development.

An examination of interdisciplinary communication by a MOR2 graduate student through transcribing and coding meeting minutes and interviews reveals new data types, data collection methods and underlying team dynamics. The word map depicts themes on data sharing, communication, access, and its challenges (Figure 2). Common themes involve concerns about data sharing among the MOR2 team, differing scientific standards, analyses, authorship, and publications. Sharing and accessing data, particularly for our Mongolia colleagues remains challenging as the data are stored and processed in a shared drive housed within CSU. While more in-depth narratives underlie

these themes, our challenges reflect our efforts to collaborate and communicate among MOR2 team members with differing norms, epistemologies, and ways of approaching science.

Language barriers provide a unique set of disciplinary and cultural challenges. Disciplinary vocabulary demands definitions to address both communication and conceptual misunderstanding. The different languages and alphabets of Mongolian and English, influences data collection, development, sharing, and access. However, the MOR2 project supports a number of Mongolian graduate students and researchers who have been invaluable in bridging this cultural divide through providing translation, oversight, and participation in data entry.

The long term status of database for maintenance, access, and research need resources and support. Recent advances in web-based data sharing provide multiple venues to explore for long-term data storage and access. The development of derived data for long-term management is another concern. For example, proprietary data from Mongolian institutions were provided for hydro-climatic analyses. A key question is the format for derived data that maintains the integrity of the original data sources but protects the confidentiality of the original data.

LESSONS LEARNED

The lessons learned from three general categories: 1) data collection; 2) database development; and 3) data use. Training and education on data collection is essential to high quality and consistent socio-ecological data. In a multi-lingual environment, rigorous translation and back-translation of surveys, and field testing and revision before field use are needed to reduce ambiguity. A simple database design organized around research themes of the project provides a comprehensive, organizational framework. The coding scheme creates a spatially-explicit hierarchy to link the different thematic datasets, allowing for the creation of customized linked databases to address specific research questions.

Methods for database versioning are needed to ensure the best data is captured and organized for use and analysis. Naming conventions with dates should be adopted early in the data development process to maintain metadata, data provenance, and reduce redundancy.

Access and use of the data must come with disclaimers. Users must be aware of the limitations of the database, what is included and what is not included and guidelines on how the database can be use. The involvement of many researchers means that data are used for multiple purposes where future use by project partners may require a level of oversight and management not yet identified.

The MOR2 database is not explicitly an integrated database, but integration is achieved through analytical approaches that link the social, ecological, and physical databases. Integration is also achieved through the team-based approach to the research. The team dynamic has strong and long-term collaborative research relationships built on trust and respect. These are fundamental to the long-term use of the MOR2 database as access, curation, maintenance, and potential further development will need to be undertaken. This paper focuses on a description of the database and only alludes to the complex, underlying issues associated with building a transdisciplinary database in a cross-cultural environment. An examination of issues associated with large research teams, the challenges of international research, and designing approaches to complex problems are all avenues for future research.

ACKNOWLEDGEMENTS

This project would not have been possible without the support from the herder groups in Mongolia that provided their insights, time and hospitality. The authors express their

thanks for all the work done in the field to collect the data that comprises the MOR2 database – colleagues and students in Mongolia and the United States. We are also grateful for the support from those that provided support in developing the structure and framework of data collection tools and database framework. This project is supported by funds from the NSF Dynamics of Coupled and Human Systems (CNH) Program award BCS-1011801.

Table 1. MOR2 sub-teams and field data collected

Sub-teams	Field Data
Social	Focus groups, interviews, and surveys at <i>soum</i> , herder group and household levels
Ecological	Rangeland soil, plant community composition, diversity and production
Physical	Hydroclimatic surveys/assessments and tree-ring analyses
Herder observation	Surveys and interviews of herder observations of hydroclimatic and rangeland soil and vegetation changes
Boundary	Participatory mapping of herders' socio-ecological boundaries
Interdisciplinary team observations	Interviews with US and Mongolian research team members, Participant observation, transcripts of research team meetings

Table 2. Coding system for survey data

AIMAG	SOOM	ORGANIZATION	HOUSEHOLD
Given a TWO DIGIT code, based on alphabetical order. <i>Aimag</i> codes are assigned beginning with the number 10 through 32. ¹	Given a FOUR DIGIT code. The first two digits = <i>aimag</i> code. The second two digits are based on alphabetical order of the <i>soum</i> names. ²	Given a SIX DIGIT number. The first two digits = <i>aimag</i> ; the second two digits = <i>soum</i> ; the third two digits are assigned randomly. Each year new organizations are added. ³	Given an EIGHT-DIGIT code. The first two digits = <i>aimag</i> ; second two digits = <i>soum</i> ; third two digits = organization. The final two digits will be associated with household winter shelter number (the 1-st digit) and plot numbers (the last digit) ³

¹Ulaanbaatar is included in this list, even though it is not officially an *aimag*.

²The districts of Ulaanbaatar are assigned two-digit codes as if they are *soums*.

³Alphabetical order is difficult to maintain for organizations and households.

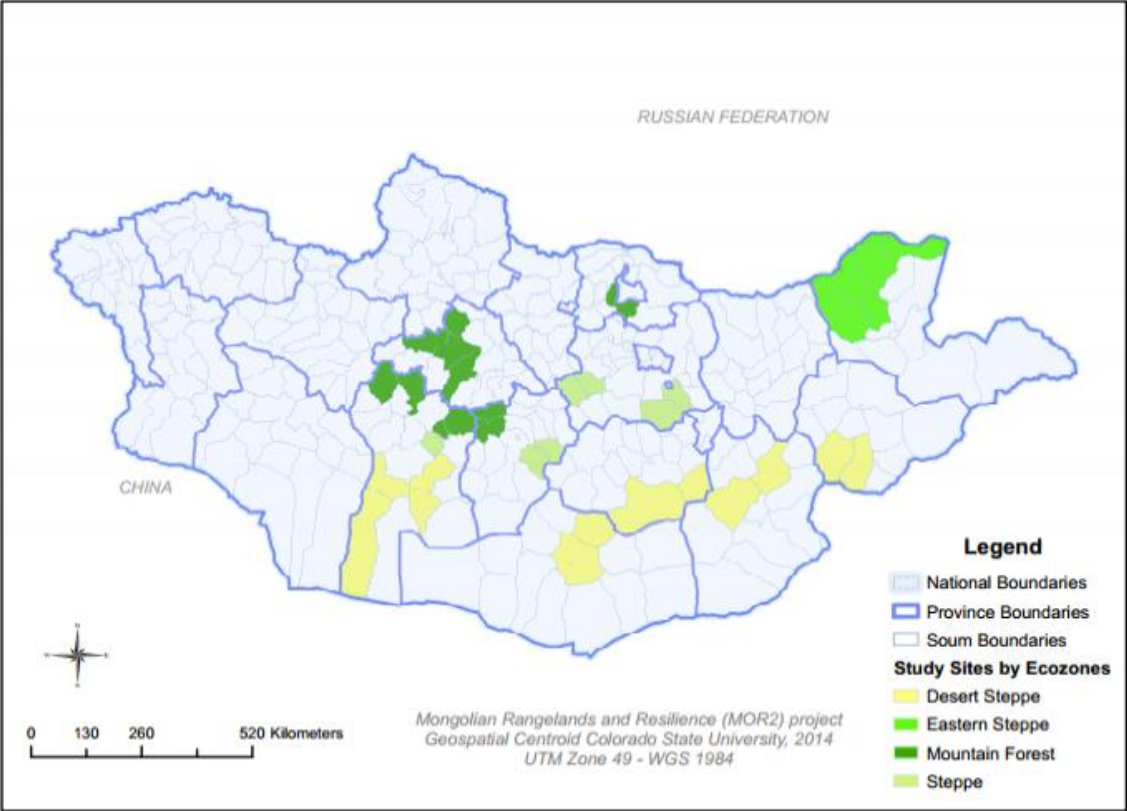


Figure 1. MOR2 paired *soum* study sites.