Managing Scientific Research Data:
Data Packaging and Organizing Materials for Curation

Nicole E. Kaplan[1], Daniel C. Draper[2] , Karen S. Baker[3] , Shea Swauger[2], John C. Moore[1], and Dawn Bastian Paschal[2]

[1]Natural Resource Ecology Lab, Colorado State University, [2] CSU Libraries, and [3] University of Illinois, Urbana-Champaign

LTER

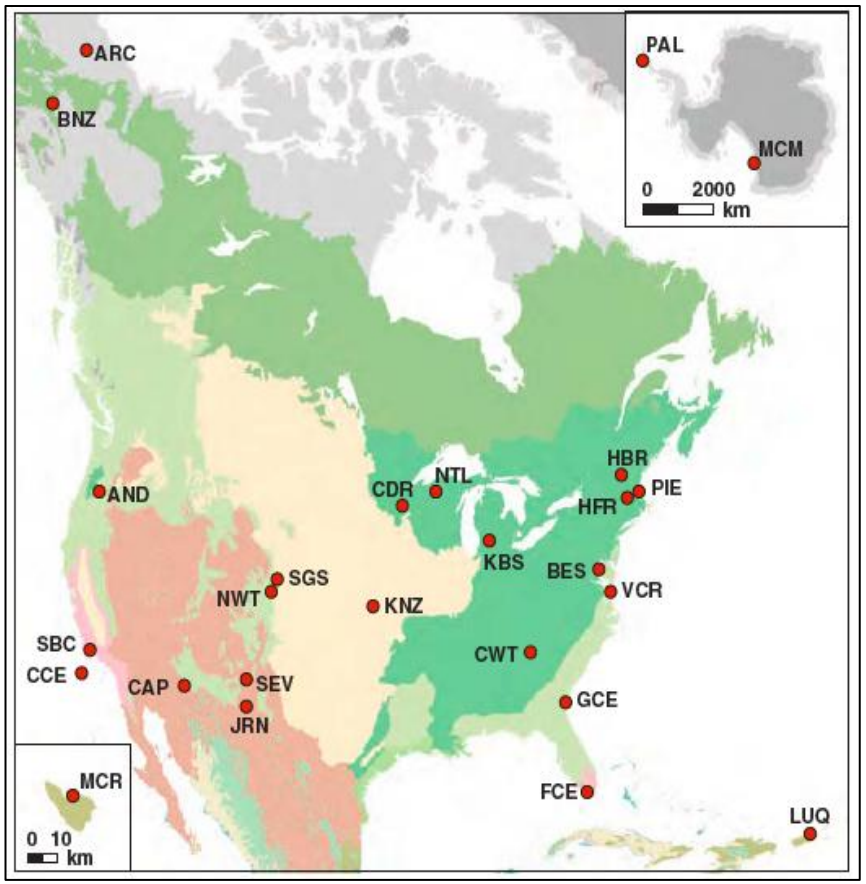
Colorado State University

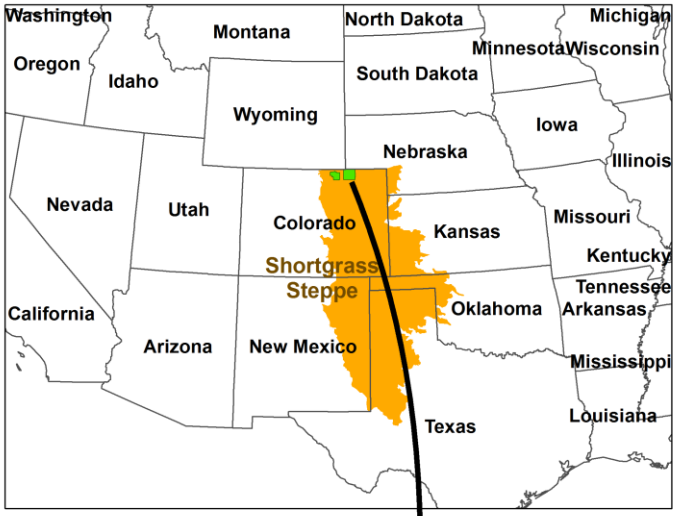
NSF

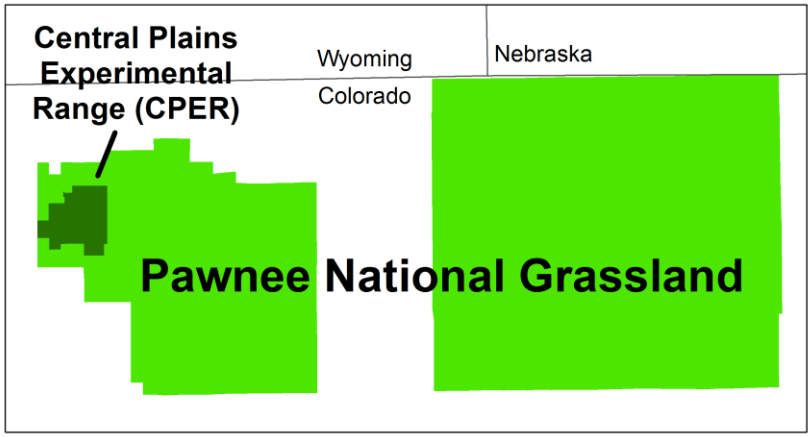
UIUC

International Biological Program (1960-70s)

The Shortgrass Steppe LTER Research Site







Sunset for SGS LTER = A New Horizon

The Participants

- **SGS-LTER**
  - local information management
- **CSU IR**
  - expanding collections to include data
- **LTER Network Information System (NIS)**
  - a secondary domain repository
- **University of Illinois at Urbana-Champaign**
  - data curation education at research centers

Broader Goals

- Develop ***an information infrastructure*** that supports new approaches, tools, and services for collections of scientific data and related artifacts.
- Provide digital ***access to all project materials*** as well as the data packages.
- Use a ***collaborative team approach*** that includes information professionals and scientific researchers.
- Contribute to a ***web of repositories.*** (Baker & Yarmey 2009.)

The Focus

- **Needs Requirements**
- **CSU as the Source**
- **Pilot Study**

Needs Requirement

- Open Access for LTER
  - LTER Data Access Policy
- Stay Involved
  - Access to Data Curators
  - Ability to Participate in Curation
  - Ability to Append, Revise or Version
- Keep it Local
  - CSU as Publisher
  - Control Embargos and Attribution
  - Establish CSU as Source Repository

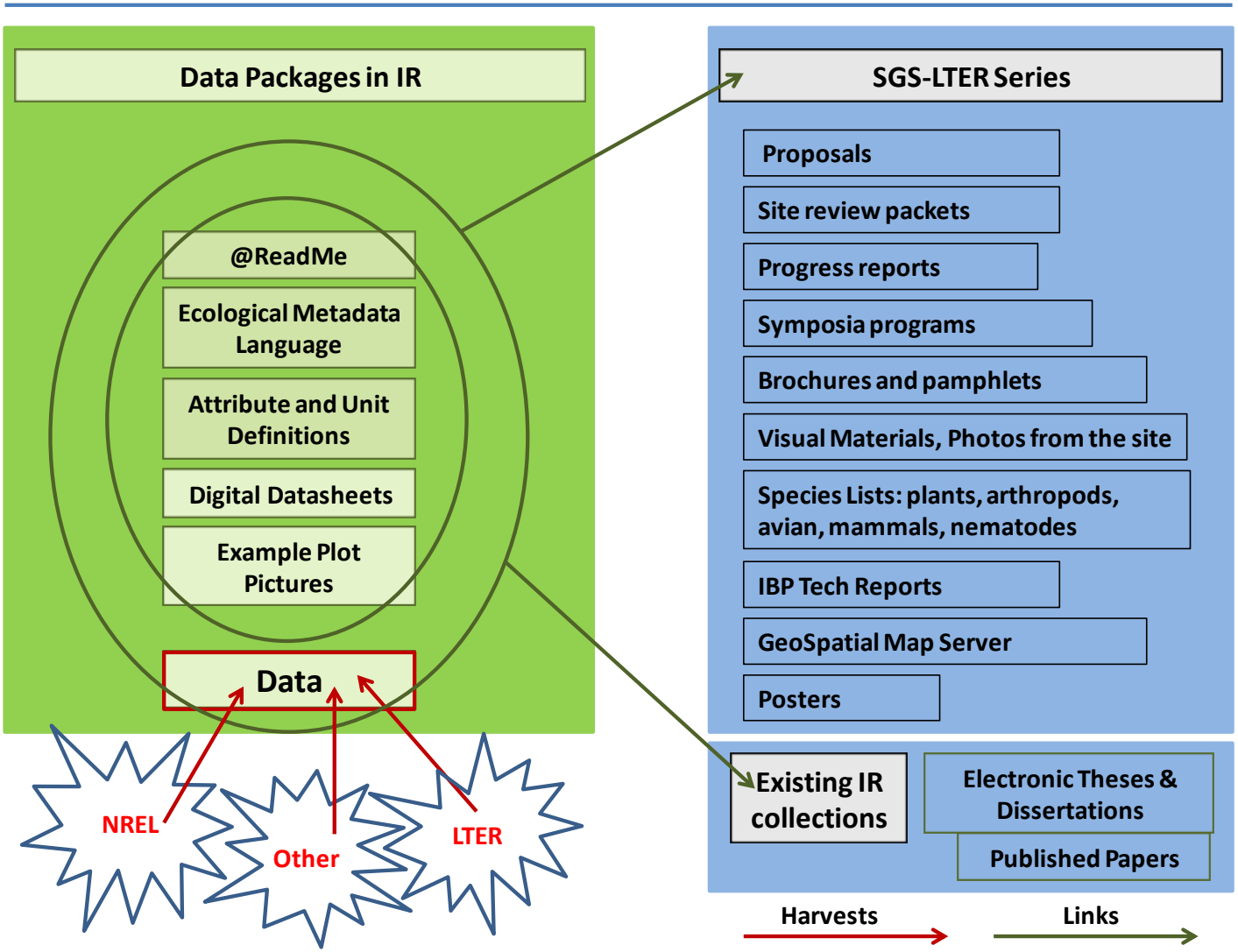
Needs Requirement cont.

- Use Rich Metadata
  - Map Ecological Metadata Language to Qualified DC
- Satisfy Interoperability
  - NIS built on PASTA, Data Packages replicated within DataONE, accessible by local IBIS

CSU IR as Source Repository

- DigiTool
  - workflow tailored to the project
- Data Package
  - Descriptive information
  - Only useful with data table
  - Data table on the outside (i.e. interoperability)
- Links to other series or established collections (ETDs, IBPs)
- Links to published research (papers) that supported by data
- Persistent identifiers provided by Handle.net

Research Data and Related Artifacts



We're Live!

- Hop to IR!

A Pilot Study

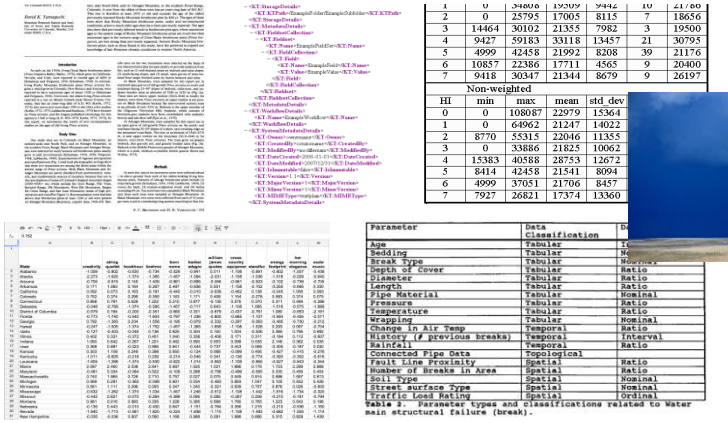
- Lessons Learned
- Report and poster produced
- Inter-collection linking
- Interoperability between repositories via programmatic harvest
- Find and work with local data managers

Interoperability, Data Discovery and Citation

Inter-repository data sharing in three steps:

CSU IR

SOURCE REPOSITORY



1	0	54808	19509	9442	10	21780
2	0	25795	17005	8115	7	18656
3	14464	30102	21355	7982	3	19500
4	9427	59183	33118	13457	21	30795
5	4999	42458	21992	8208	39	21176
6	10857	22386	17711	4565	9	20400
7	9418	30347	21344	8679	9	26197

Non-weighted

HI	min	max	mean	std_dev
0	0	108087	22979	15364
1	0	64962	21247	14022
2	8770	55315	22046	11355
3	0	53886	21825	10062
4	15383	40588	28753	12672
5	8414	42458	21541	8094
6	4999	37051	21706	8457
7	7927	26821	17374	13360

Parameter	Data Classification
Age	Tabular
Bedding	Tabular
Break Type	Tabular
Depth of Cover	Tabular
Diameter	Tabular
Length	Tabular
Pipe Material	Tabular
Pressure	Tabular
Temperature	Tabular
Wrapping	Tabular
Change in Air Temp	Temporal
History (# previous breaks)	Temporal
Rainfall	Temporal
Connected Pipe Data	Topological
Fault Line Proximity	Spatial
Number of Breaks in Area	Spatial
Soil Type	Spatial
Street surface Type	Spatial
Traffic Load Rating	Spatial

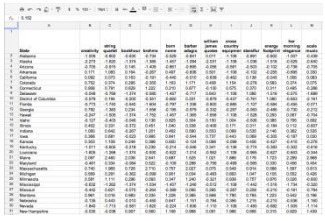
Table 2. Parameter types and classifications related to Water main structural failure (break).

1. **Data Package is registered with a handle @ CSU IR**
2. **Contents of an LTER data package is transferred to LTER**
3. **The data subset is registered with a new DOI (DataCite) in LTER and TR**

Data users find data in source repository (CSU) or secondary repository (LTER)

LTER

SECONDARY REPOSITORY



EML 2.1.0

Next Steps

DM Tasks:

- ingest more data packages
- continue to share our story

Local Data Management System:

- no landing page for the collect
  - create web page at the lab to serve as portal with descriptive information about the project and its history

IR Policies:

- link to GeoSpatial Centroid
- articulate issues that arise in interoperability between source and secondary repository and duplication or replication of data packages (i.e. provenance of data)

Thanks!

- NSF Grant DEB-1027319
- CSU Libraries
- UIUC Data Curation Education at Research Centers (DCERC IMLS Award #RE-02-10-0004-10).