

# Data Curation Issues in Transitioning a Field Science Collection of Long-Term Research Data and Artefacts from a Local Repository to an Institutional Repository

Nicole E. Kaplan<sup>1</sup>, Daniel C. Draper<sup>2</sup>, Dawn Bastian Paschal<sup>2</sup>  
John C. Moore<sup>1</sup>, Karen S. Baker<sup>3</sup>, and Shea Swauger<sup>2</sup>

<sup>1</sup>Natural Resource Ecology Lab,  
Colorado State University

<sup>2</sup>Colorado State University Libraries

<sup>3</sup>University of Illinois Urbana-Champaign

Nicole.Kaplan@colostate.edu



## The Challenge

Transition a local 32 year project, the Shortgrass Steppe Long-Term Ecological Research (SGS-LTER), with over 100 data packages and related digital artefacts, to an Institutional Repository (IR) at Colorado State University (CSU) Libraries to ensure persistent, reliable, and interoperable access to our collection of scientific data.

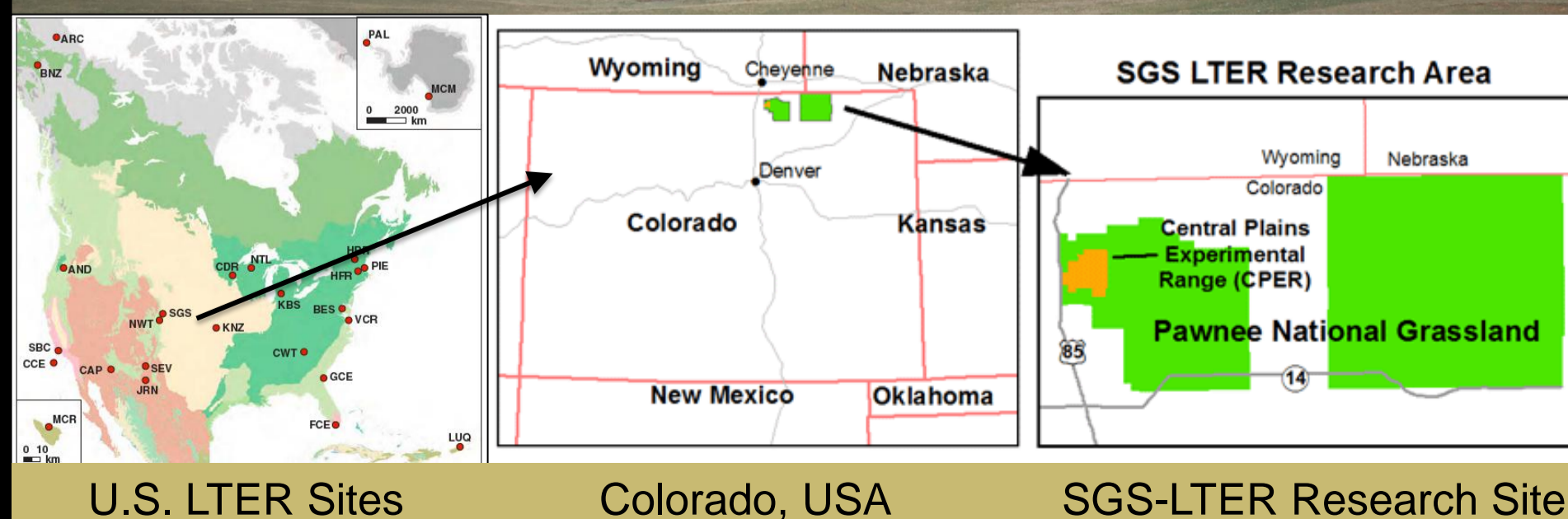
## The Participants

- **SGS-LTER** represents a unique scientific project with a local information management component, which produces organized & well-described data with Ecological Metadata Language (EML).
- **CSU IR** is expanding support for preservation and access to scholarly output, including scientific research data.
- **LTER Network Information System (NIS)**, a related secondary domain repository, harvests, ingests, and disseminates data from all 26 LTER sites in the network.

## Broader Goals

- Develop *an information infrastructure* that supports new approaches, tools, and services for collections of scientific data and related artefacts.
- 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*.

## Sites Maps



## The Paradigm

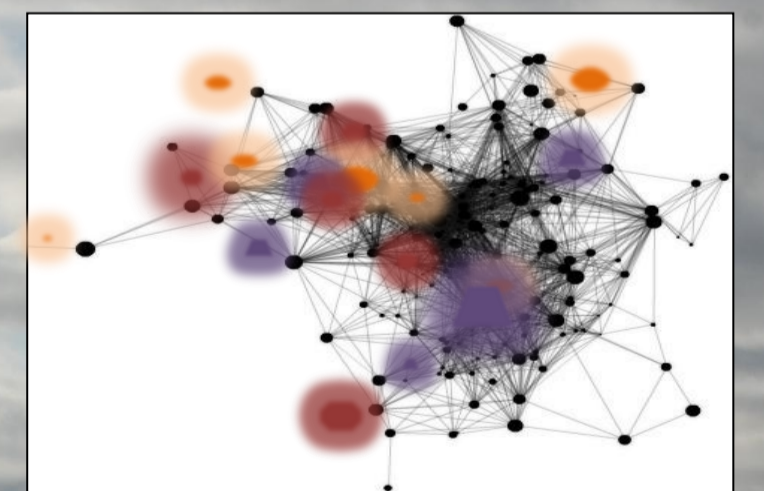
Data and artefacts are organized in series that contribute to a SGS-LTER project collection within the CSU IR. Data tables are packaged with metadata and are machine-accessible by the LTER NIS and beyond.

## Issues Encountered **Significant**

1. Developing the capacity to incorporate project collections of scientific data and related artefacts
2. Developing and making visible working standards (EML metadata, Qualified Dublin Core, best practices for ecological research projects, controlled vocabularies)
3. Organizing, describing, and relating project artefacts
4. Linking objects in other collections in the IR (theses, articles) within SGS-LTER data packages
5. Ensuring discoverability and interoperability for searches and harvests (e.g. Google, LTER)
6. Establishing relationships among repositories
7. Working with local research groups to ensure analysis and visualization systems have programmatic access to data (e.g. NREL)
8. Planning for versioning and/or appending ongoing scientific dataset series

## The Big Picture

Our collaborative team envisions being part of a larger information environment, which enables sharing of knowledge and data – a web of repositories.



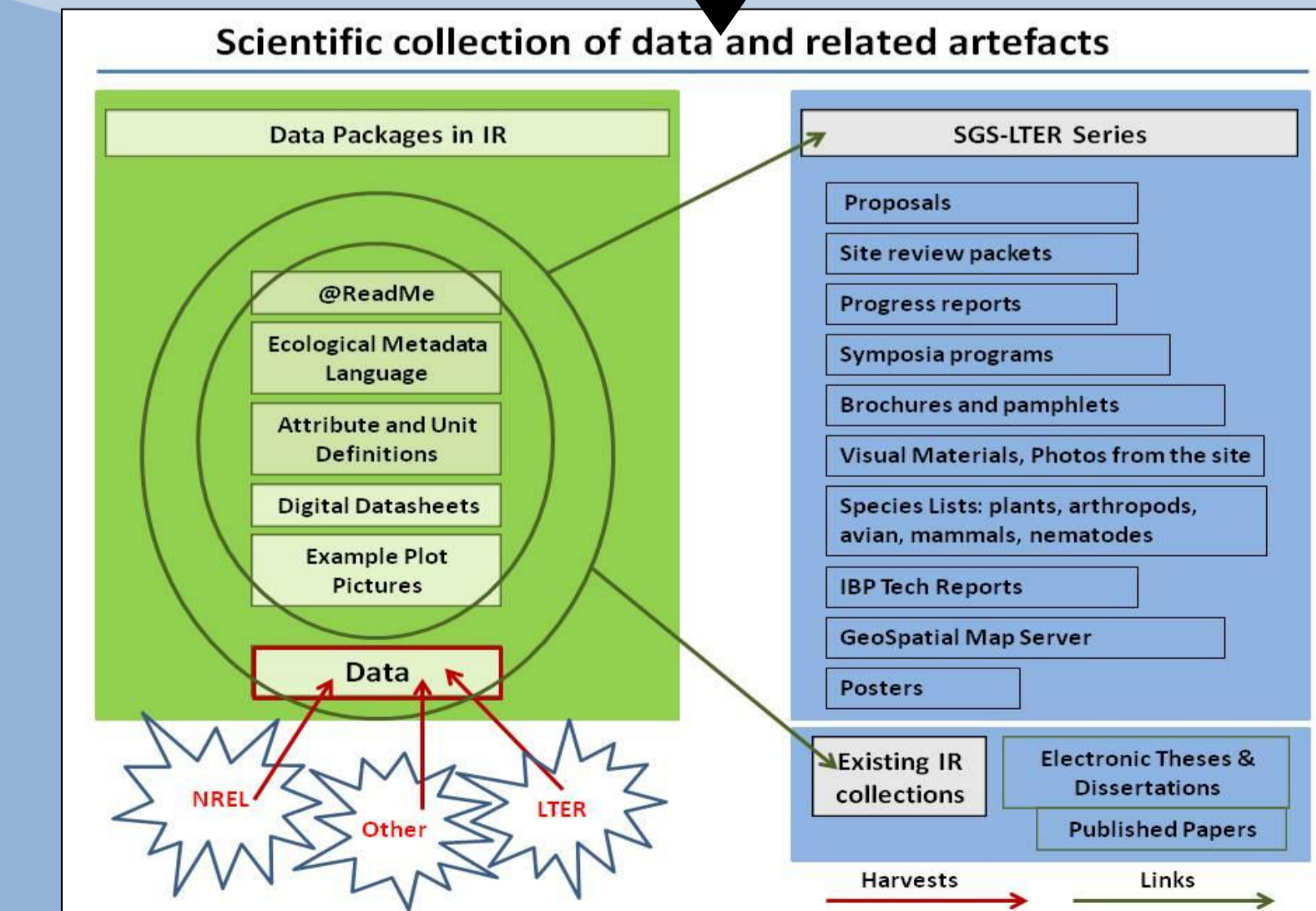
Baker, K. S., & Yarmey, L. (2009). Data stewardship: Environmental data curation and a web-of-repositories. *International Journal of Digital Curation*, 4(2), 12-27.

## Acknowledgements

This work is supported by NSF Grant Number DEB-0823405, Colorado State University, and the UIUC Data Curation Education at Research Centers (DCERC IMLS Award #RE-02-10-0004-10).



Photos contributed by N. Kaplan, M. Lindquist, P. Stapp, and M. Johnson.



A schematic showing a SGS-LTER data package bundled with information necessary to use the data (left). Links to project artefacts organized in series and to other collections in the CSU IR (right). Data tables are available for harvest by LTER, local data systems (e.g. NREL), or other secondary repositories (bottom left). The scientific collection in the IR is documented with Qualified Dublin Core.