Using Publicly Available Metadata to Analyze Data Sharing Practices at Oklahoma State University

Rhoda Alawiye[1] & Dani Kirsch[2]

[1] MS in Business Analytics & Data Science, OSU; [2] Research Data Specialist, OSU Library

INTRODUCTION

- **Sharing and publication of research data** increasingly required by funding agencies and academic publishers
- Recent growth in the offering of **research data services**, often at academic libraries, to support this data sharing
- However, **reporting publication of datasets not as common** as for articles
- OSU is a very active R1 institution, but we have **limited insights** on research data sharing among OSU researchers
- **DataCite** provides digital object identifiers (DOIs) to many types of scholarly works, **including datasets**, and **maintains metadata** on these works

OBJECTIVES & OVERVIEW

Primary goals of this project:

1. Improve our understanding of research data sharing at OSU by using publicly available metadata from DataCite
2. Collect DOIs for published datasets and associate them with faculty profiles in the OSU Experts Directory

Data Overview:

- **176** published datasets from 2011-2023
- **150** unique OSU authors
  - Faculty, staff, postdoctoral fellows, graduate students, & undergraduates
- **9** publishers
  - Dryad most common (>80%)

METHODOLOGY

Datasets derived using DataCite API → Data cleaning & preprocessing → OSU author profile retrieval from various sources → Exploratory data analysis and network analysis

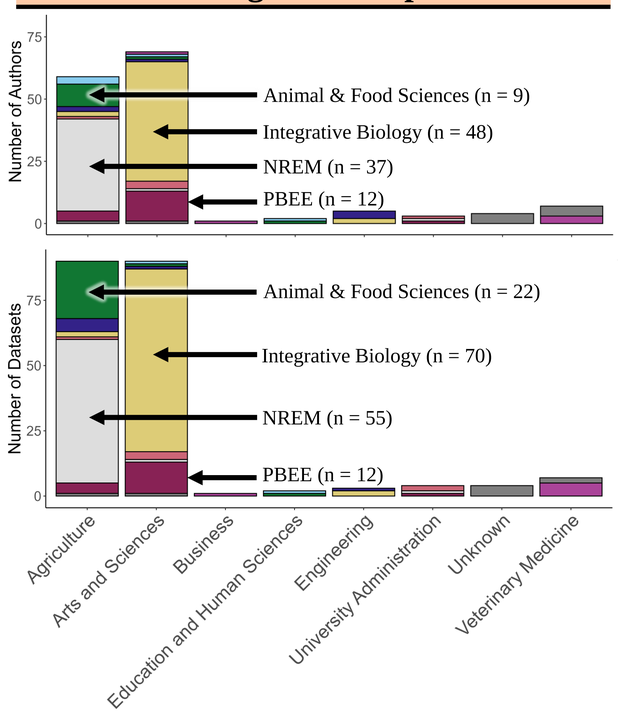
- Created **custom query** for DataCite API
  - Restricted results to datasets with at least one author affiliated with OSU
  - Used Research Organization Registry (ROR) ID to identify OSU affiliations
- **Extracted relevant data** from resulting JSON file using RStudio
- Cleaned dataset in Excel
  - Standardized author names, removed non-OSU authors, removed duplicate entries
- Matched authors to **Experts Directory** to assign college & departmental affiliations

Software Utilized

- Tableau used to visualize datasets with multiple OSU co-authors
- VOSviewer used to create & visualize co-author networks

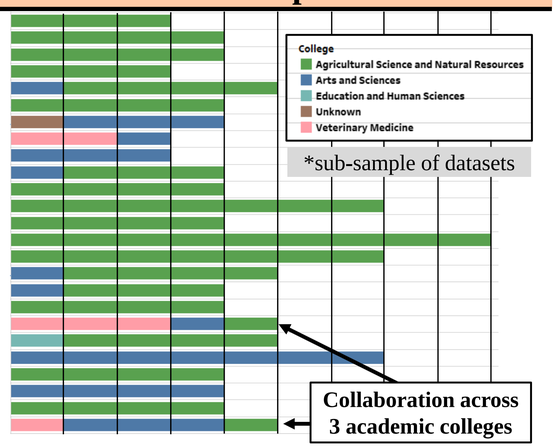
PRELIMINARY RESULTS

Author & Dataset Representation across Colleges and Departments



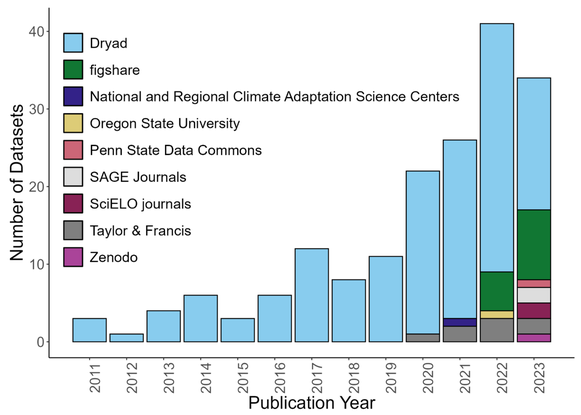
- Colors represent different departments in each college
- Co-authored datasets may be counted multiple times
- NREM = Natural Resource Ecology & Management
- PBEE = Plant Biology, Ecology, & Evolution

Datasets* with Multiple OSU Authors



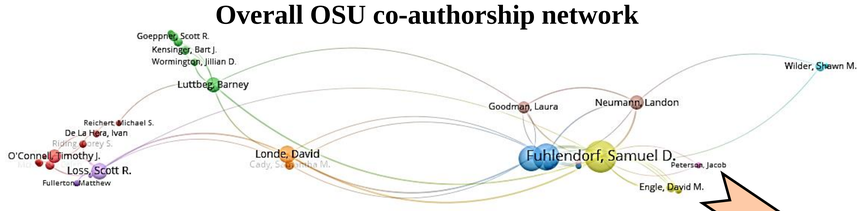
- Node size indicates relative number of co-authored datasets
- Top 3 recurring authors affiliated with Natural Resource Ecology & Management (NREM) department in College of Agriculture

Dataset Publication Trends over Time



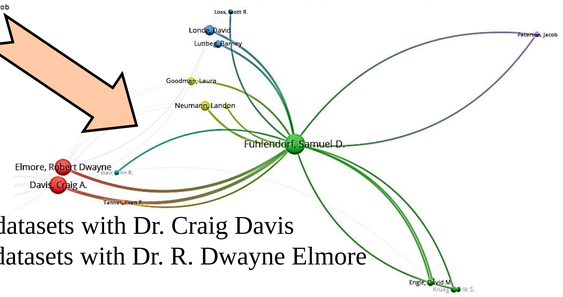
OSU Co-Author Networks

Overall OSU co-authorship network



Dr. Samuel Fuhlendorf, Regents Professor

- 19 datasets published with 16 other OSU authors



- 15 datasets with Dr. Craig Davis
- 11 datasets with Dr. R. Dwayne Elmore

CONCLUSIONS

- **DataCite metadata provide good starting point for assessing OSU research data sharing trends**
  - More efficient than manual search
  - Identified high-activity researchers, departments, & colleges
- **Technical limitations**
  - Extracting information from JSON
  - Easily connecting datasets to publications that cite them
- **Informational limitations**
  - Data repository representation
  - Inconsistency of metadata
    - Variable reporting standards
    - Quality of dataset usage metrics
- **Next phase**: uploading dataset metadata to Experts Directory

ACKNOWLEDGEMENTS

We thank **Clarke Iakovakis** (Scholarly Services Librarian at Oklahoma State University), **Kelly Stathis** (DataCite Representative) and **Isaac Wink** (Research Data Librarian at University of Kentucky), for their contributions and guidance during various phases of this project.