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

Rhoda Alawiye¹ & Dani Kirsch²

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
  o Faculty, staff, postdoctoral fellows, graduate students, & undergraduates
• 9 publishers
  o Dryad most common (>80%)

PRELIMINARY RESULTS

Author & Dataset Representation across Colleges and Departments

Datasets with Multiple OSU Authors

Dataset Publication Trends over Time

OSU Co-Author Networks

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

METHODOLOGY

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

CONCLUSIONS

• DataCite metadata provide good starting point for assessing OSU research data sharing trends
  o More efficient than manual search
  o Identified high-activity researchers, departments, & colleges
• Technical limitations
  o Extracting information from JSON
  o Easily connecting datasets to publications that cite them
• Informational limitations
  o Data repository representation
  o 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.