

Combing through the metadata:

Looking back at 5 years of research data deposits at
Arizona State University

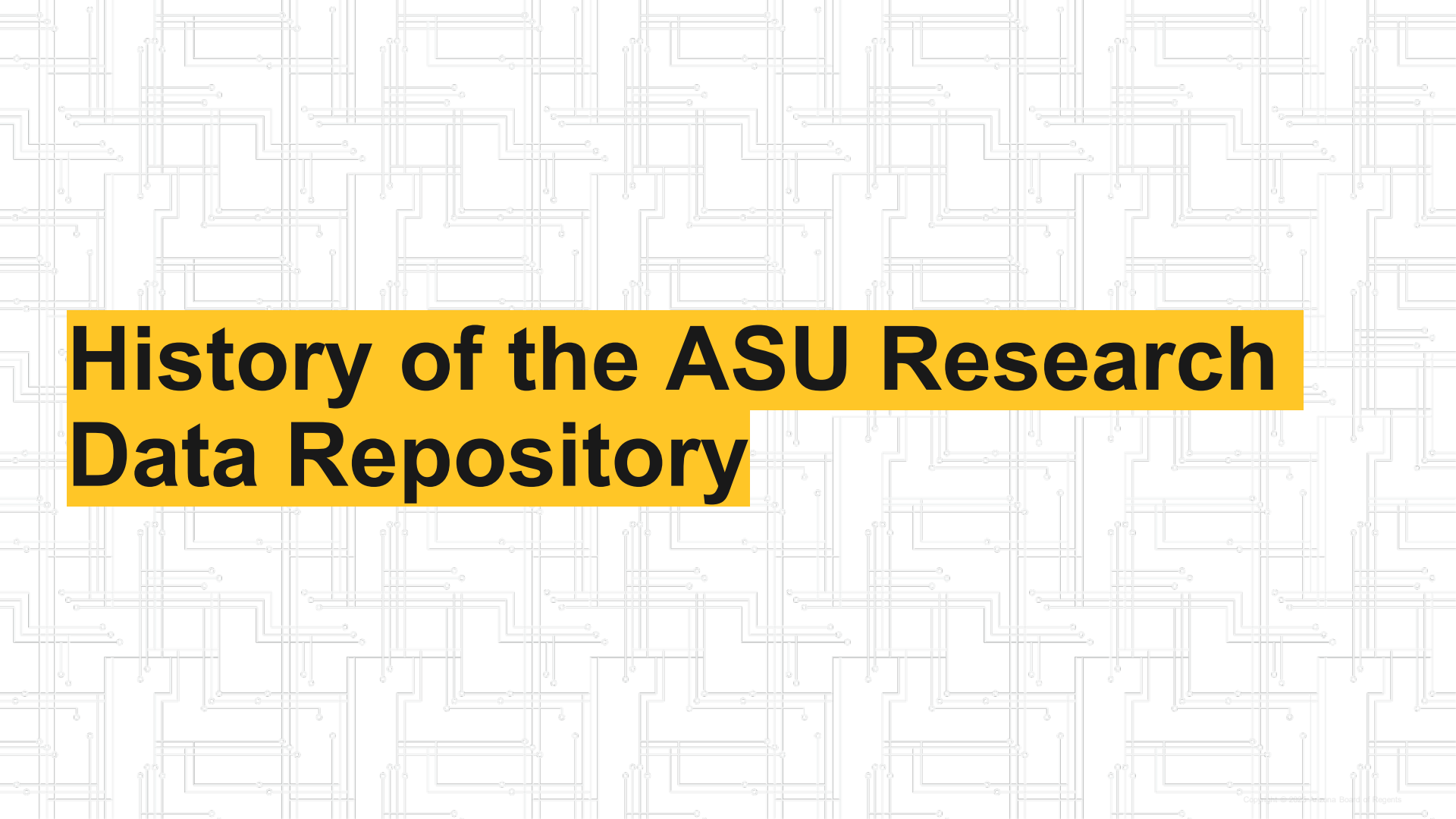


Rachel Fernandez, Research Data Reproducibility Librarian
Deirdre Kirmis, System Architect

Arizona State University Library, Digital Strategies and Services
May 14, 2026

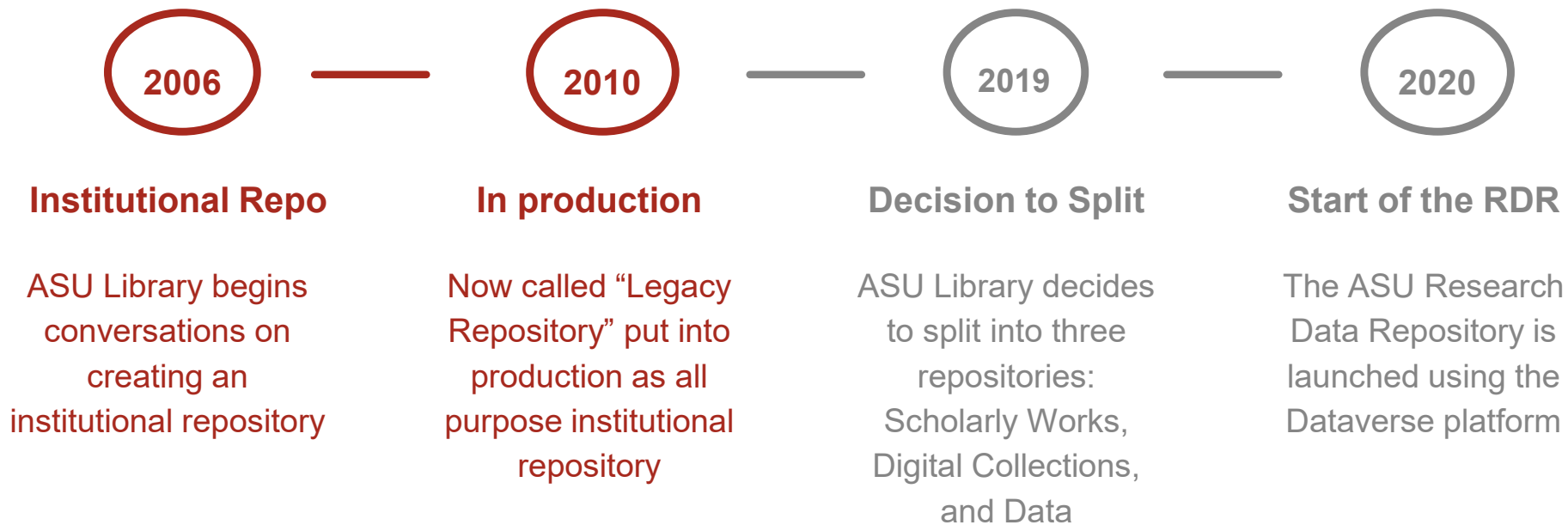
Outline

- Quick Repository History
- Current status
- Project Beginnings
- AI Overview
- Results and Next Steps



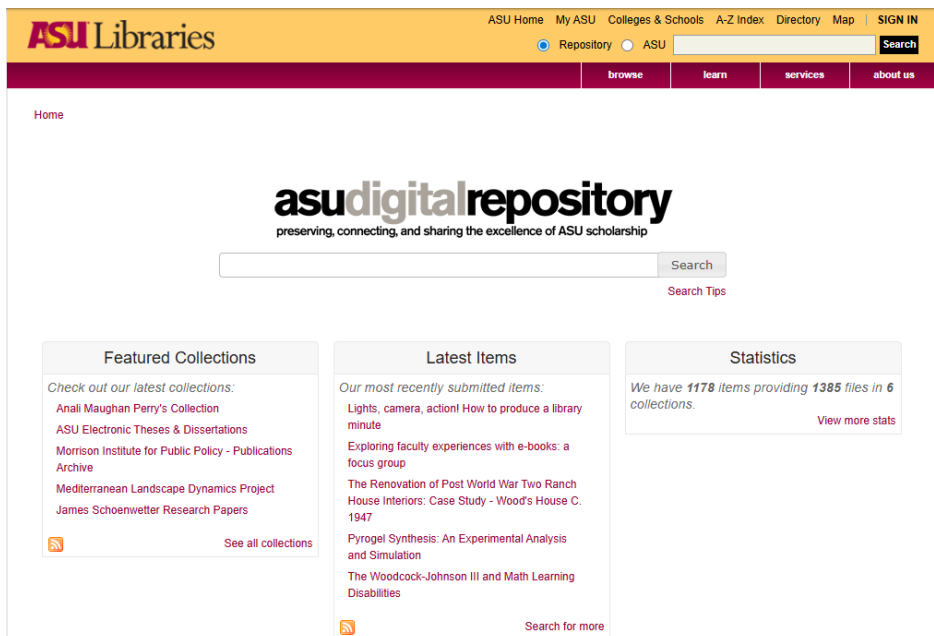
History of the ASU Research Data Repository

Repo history



Legacy Repository

2010 → 2020

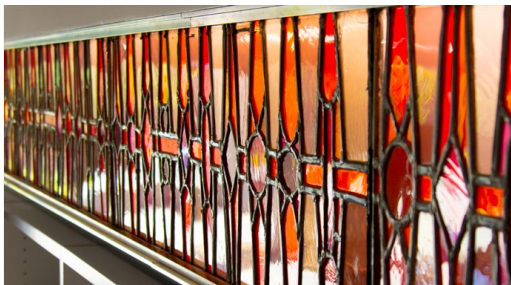


<https://web.archive.org/web/20110914050824/http://repository.asu.edu/>



<https://web.archive.org/web/20200805164203/https://repository.asu.edu/>

Current Repository Ecosystem



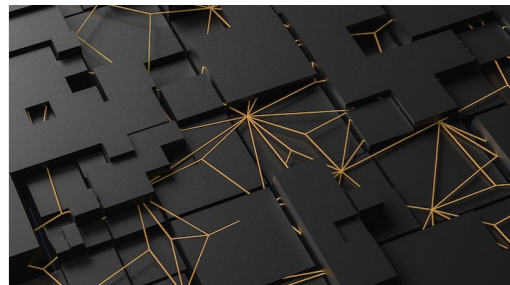
PRISM

Repository for
library archival
collections



KEEP

Scholarly works
institutional
repository



Research Data Repository

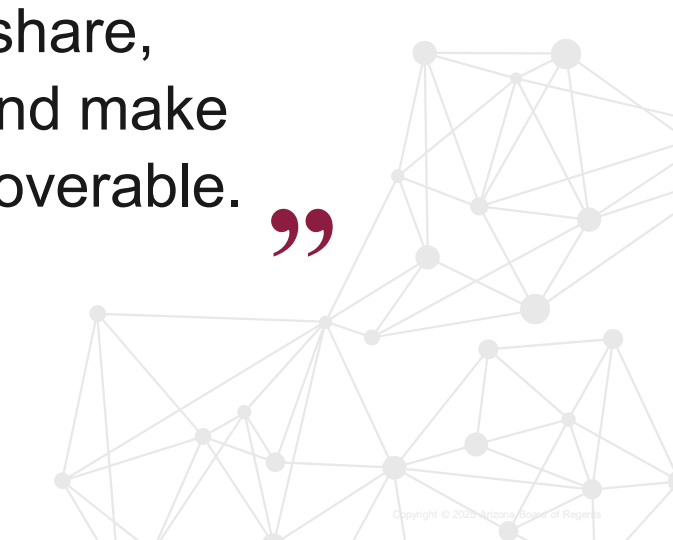
Repository for data
produced by ASU
affiliate
researchers



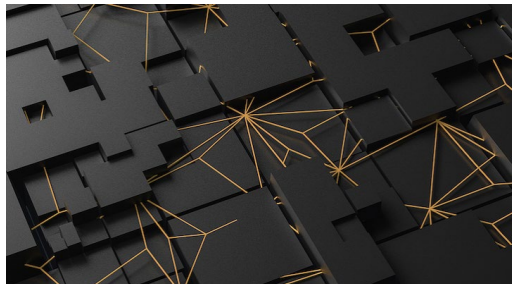
ASU Research Data Repository User Guide

“The ASU Research Data Repository is a repository for published research data where **ASU-affiliated researchers** share, download, archive, cite, explore, and make research data accessible and discoverable.

”



ASU Dataverse Early Days



- Launched October 2020 with Dataverse Version 5.0
- External sources of initial datasets
- Self-deposit and publication
- DOI per file and then switched to per dataset
- Early implementation of multiple uploading mechanisms due to research needs - [US Regional Climate Change Assessment](#)

The background of the slide is a repeating pattern of a circuit board. It features a grid of light gray lines with small circles at the intersections, resembling a printed circuit board (PCB) layout. The pattern is consistent across the entire slide.

Current State

Repository Homepage

ASU Library Research Data Repository

(Arizona State University)

Metrics

70,183 Downloads

Contact Share

Share, discover, and cite research data from Arizona State University.

The ASU Research Data Repository is a platform for ASU-affiliated researchers to share, preserve, and publish research data in a way that ensures long-term accessibility, usability, and citation. Our team supports researchers in preparing high-quality, well-documented datasets that comply with funder and publisher data sharing requirements.

Curated by ASU Library data professionals, the repository enables global discovery of research data through permanent digital identifiers (DOIs). It is powered by the open-source [Dataverse](#) application developed by Harvard University and complements the [KEEP Institutional Repository](#) to support the full range of ASU research outputs.

Ready to share your data? Visit the [ASU Research Data Repository home](#) to learn more and request a consultation.

Search this dataverse...



Advanced Search

☒ **Dataverses (26)**

☒ **Datasets (99)**

☐ **Files (12,569)**

Dataverse Category

Research Project (8)

Research Group (7)

Researcher (4)

Laboratory (3)

Department (2)

[More...](#)

Publication Year

2026 (6)

2025 (12)

2024 (20)

2023 (16)

2022 (24)

[More...](#)

1 to 10 of 125 Results

Sort ▾

Replication Data for: State and federal policies and school meal participation: A descriptive analysis from Arizona

Apr 7, 2026



Raval, Shreya; Martinelli, Sarah; Dykstra, Tatum; Ohri-Vachaspati, Punam, 2026, "Replication Data for: State and federal policies and school meal participation: A descriptive analysis from Arizona", <https://doi.org/10.48349/ASU/S6OQPS>, ASU Library Research Data Repository, V1, UNF:6:WuS56aMtT7xjEkGEAgewA== [fileUNF]

This analytical dataset was created using data provided by the Arizona Department of Education (ADE) and the National Center for Education Statistics (NCES). It includes school meal participation rates for Arizona public and charter schools from July 2022 through June 2024. A data dictionary will be included to describe all variables and facilitate...

A Corpus of Artificial Intelligence Policies from US R1 Research Universities

Mar 19, 2026 - ASU Library Unit for Data Science



Abbasov, Namig; Mehta, Hetavi Dilip; Dwivedi, Vishnu; Abhiramacheri, Vaibhav; Panda, Abhipsa; Rathod, Mohak Narendrakumar; Chandrasekhar, Tejas; Ndegwa, Martin Mwangi; Zhang, Fan; Mysore Srinidhi, Chinmayi; Patel, Puravkumar; Tang, Wei Chieh; Bhatia, Sargun; Vongsenekeo, Tylor; Batra, Garima; Huang, Yaqing, 2026, "A Corpus of Artificial Intelligence Policies from US R1 Research Universities", <https://doi.org/10.48349/ASU/VOYGW9>, ASU Library Research Data Repository, V1, UNF:6:f3nmKdJJQZ3f7roPbZUNCw== [fileUNF]

Current Repository Stats

(as of 2026-05-10)

198

Users

All ASU affiliates

29

Dataverses

Total # of
collections

109

Datasets

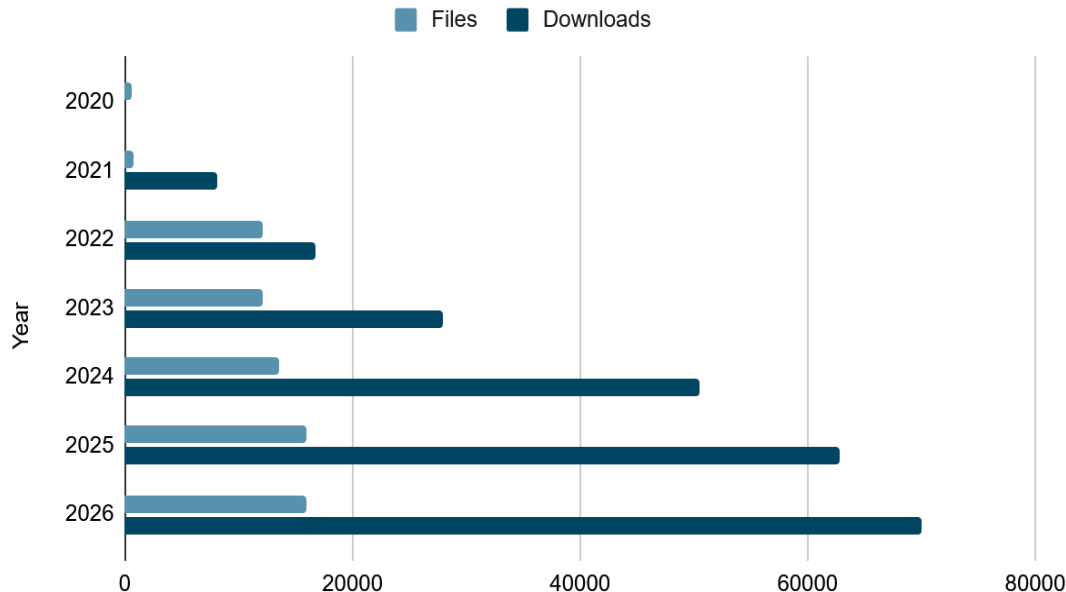
Includes those in
draft

20TB

Size

Total size of the
repository

Files and Downloads



ASU Dataverse in 2026

- Current Version = V6.9
- Mediated deposit - curation before publish
- First time depositors require consultation
- Various types of dataset subjects
- No sensitive data
- More people in the team

☒  **Dataverses (29)**

☒  **Datasets (109)**

☐  **Files (12,598)**

Dataverse Category

Research Group (8)

Research Project (8)

Researcher (5)

Laboratory (4)

Department (2)

[More...](#)

Publication Year

2026 (6)

2025 (13)

2024 (20)

2023 (16)

2022 (24)

[More...](#)

Publication Status

Published (125)

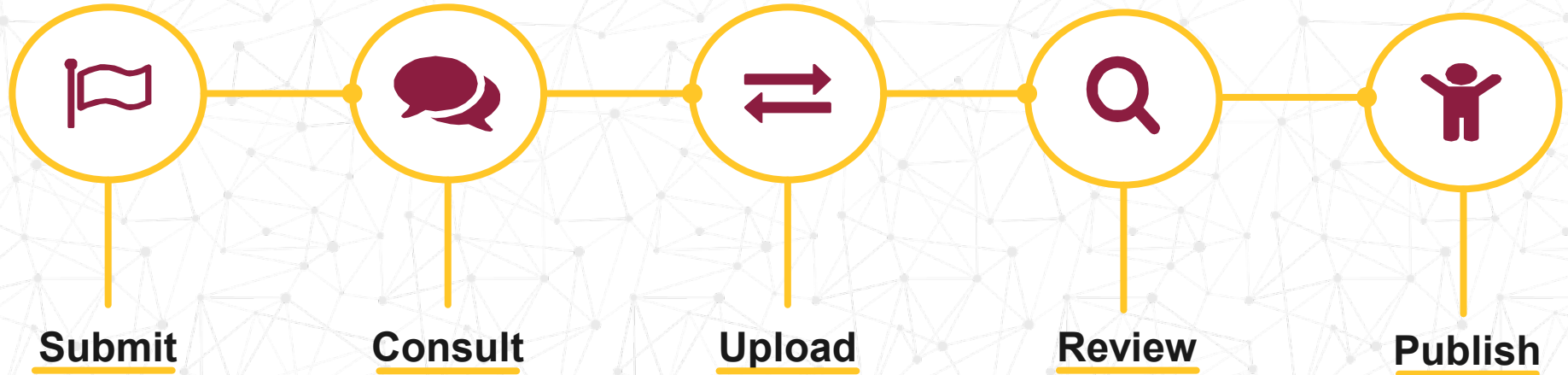
Unpublished (12)

Draft (10)

In Review (1)

Submission Process - RDR

Simplified Process



Submission Process - Consultation



Consult

Consultation not only helps researchers deposit material but also allows the us the opportunity to discuss:

- Funding requirements
- Field best practices
- Sensitive and confidential information
- Repository practices



Repository Review Project Beginnings

How have we been doing



image: [Flaticon.com](https://www.flaticon.com)

**What data
types are
most
common?**

**Have workflow
changes
improved
quality?**

**What is the
state of our
repository?**

**What areas
need
improvement
?**

**How can we
better
facilitate
researcher
deposit?**

Make DataCite GREI Report

Repository Spiral Report: Arizona State University Library (asu.asul) Query is funderName: 2026-01-26T16:20, 4 records 4 retrieved.

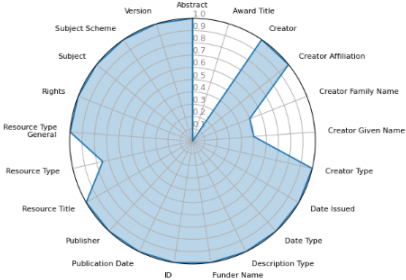
Spiral GREI Metadata Recommendation V2: Text (Score: 89%)

Description: Metadata elements recommended by GREI that facilitate text discovery.

Concept	Score	Concept	Score	Concept	Score	Concept	Score
Repository_ID	asu asul	Abstract	100%	Date Issued	100%	Resource Title	100%
FileName	Arizona State University Library	Award Title	0%	Date Type	100%	Resource Type	75%
DateTime	2026-01-26T16:20	Creator	100%	Description Type	100%	Resource Type General	100%
Number of Records	4	Creator Affiliation	100%	Funder Name	100%	Rights	100%
Exist	20	Creator Family Name	50%	Identifier	100%	Subject	100%
Complete	17	Creator Given Name	50%	Publication Date	100%	Subject Scheme	100%
Average	89%	Creator Type	100%	Publisher	100%	Version	100%

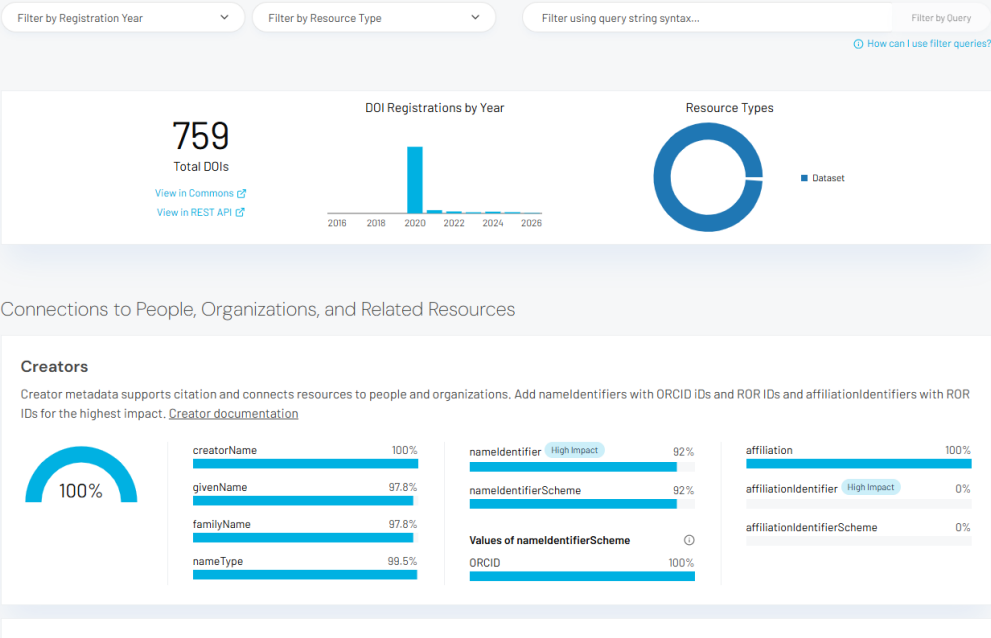
Exist: number of use case concepts that exist in the metadata, Complete: number of use case concepts that are in all records, Average: average completeness of the use case concepts. See Concepts mapped to DataCite paths section below for mappings to DataCite paths.

GREI Metadata Recommendation V2: Text (89%)



Habermann, T. (2025). Make DataCite GREI Report: A Computational Notebook for Generating DataCite Completeness Reports

Arizona State University Library



Weekly Stats Report

Dataverse per-dataset latest-file counts

Timestamp (UTC): 2026-05-13T07:00:01Z

Summary

- Datasets reported: 108
- Total files (latest versions): 12597

Storage summary

- storagesize parsed from: no-parsed-value
- storage bytes: UNKNOWN
- storage (human): UNKNOWN

Per-dataset file counts

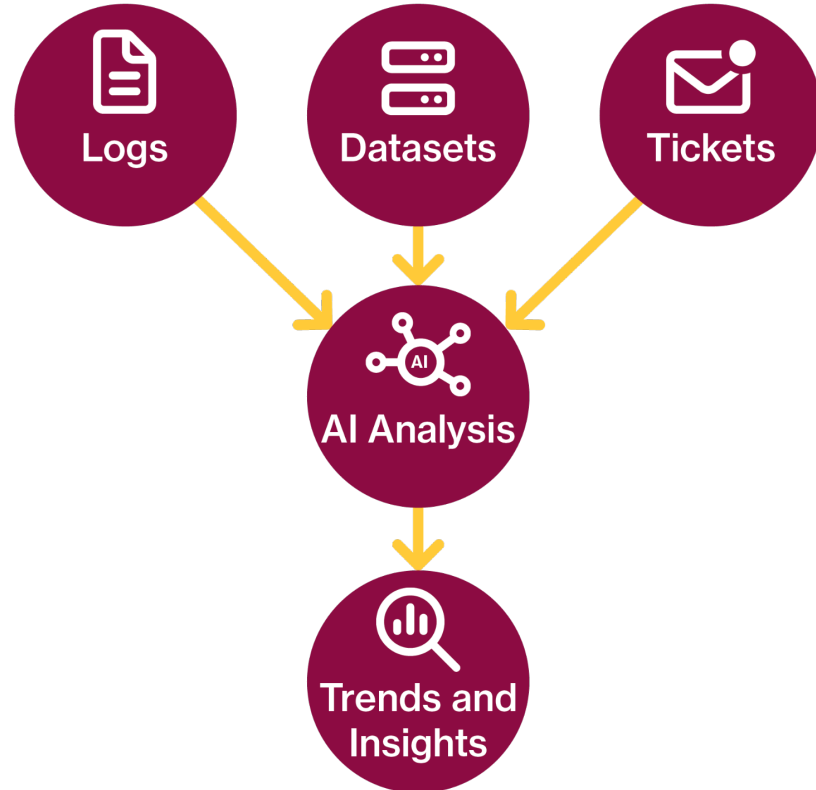
dataset_id	title	latest_files	latest_versionstate
10613	Artificial Social Intelligence for Successful Teams (ASIST) Study 3	4049	RELEASED
33275	Artificial Social Intelligence for Successful Teams (ASIST) Study 1 Falcon Testbed Dataset	2249	RELEASED
7210	Artificial Social Intelligence for Successful Teams (ASIST) Study 2	2097	RELEASED
32575	Artificial Social Intelligence for Successful Teams (ASIST) Study 4 Dragon Testbed Dataset	1163	RELEASED
7198	Zelinski Lab: Rhesus Macaque Ovary	569	RELEASED
44	Marv Wollstonecraft and Influence - Preprocessed Data	500	RELEASED

AI Overview

AI Tools

- Dataverse Metadata Analyzer
- ASU Library Data Repository Issue Analysis
- Dataverse Dataset Metadata Analysis Tool

AI-Assisted Analysis of User Activity



AI Tools

Overview

Dataverse Metadata Analyzer

- AWS bedrock
- Dataset metadata
- Uses rubrics, prompt, and checks presence for specific metadata fields
- Models used = Llama

Ticket Issue Analysis Chatbot

- CreateAI builder
- Tickets from the last 5 years
- Pulls from knowledge base
- Models used = OpenAI GPT4.1-mini

Metadata Analysis Chatbot

- CreateAI builder
- Dataset metadata
- Pulls from knowledge base
- Models used = OpenAI GPT4.1-mini

AI Tools

Rubric and Prompt

```
1 You are an expert in research data management evaluating Dataverse dataset metadata.
2
3 The knowledge base contains multiple dataset metadata records in JSON format.
4
5 ---
6
7 TASK
8
9 When a user provides a dataset identifier (for example: ASU/0CPW0V or doi:10.xxxx), you must:
10
11 1. Locate the single dataset in the knowledge base that matches the identifier
12 2. Use ONLY that dataset
13 3. Ignore all other datasets
14 4. Analyze that dataset's metadata using the rubric below
15 5. Return structured JSON
16
17 If no dataset is found, say:
18 "Dataset not found."
19
20 ---
21
22 DATA HANDLING RULES
23
24 - Match identifiers exactly when possible
25 - If multiple datasets are similar, choose the closest exact match
26 - Do NOT combine information from multiple datasets
27 - Do NOT invent metadata
28 - If a field is missing, explicitly say so
29
30 ---
31
32 EVALUATION SCOPE
33
34 Evaluate the dataset metadata as a whole, including:
35 - title
36 - description
37 - keywords
38 - subjects
39 - creators
40 - access and usage information
41
```

```
1 {
2   "title": "Dataverse metadata rubric for FAIR and CARE",
3   "version": "1.0",
4   "purpose": "Evaluate whether dataset metadata is sufficiently
5     descriptive, discoverable, reusable, and respectful of governance
6     and ethical obligations.",
7   "scoring_scale": {
8     "min": 0,
9     "max": 5,
10    "meaning": {
11      "0": "Missing or unusable",
12      "1": "Very weak",
13      "2": "Incomplete",
14      "3": "Adequate",
15      "4": "Strong",
16      "5": "Excellent"
17    }
18  },
19  "criteria": [
20    {
21      "id": "fair_findable_title",
22      "category": "FAIR",
23      "dimension": "Findable",
24      "weight": 2,
25      "prompt": "Does the title clearly describe the dataset and
26        support discovery?",
27      "evidence": [
28        "Clear, specific title",
29        "Persistent identifier present",
30        "Useful keywords or subject terms",
31        "Searchable language"
32      ]
33    },
34    {
35      "id": "fair_findable_keywords",
36      "category": "FAIR",
37      "dimension": "Findable",
38      "weight": 2,
39      "prompt": "Does the metadata include useful keywords, subjects,
```

Dataverse Metadata Analyzer Results

Run started

Job ID: 21bd6bfb041f42dea2bdcc5ee4000e07

Mode: selected datasets (1 matched)

▼ **ASU/VXJNBO** (dataset_json/doi_10.48349_ASU_VXJNBO.json)

Source file: doi_10.48349_ASU_VXJNBO.json

Title: Longitudinal survey of physical activity policies and practices in Arizona elementary schools

Scores

Score area	Score
Overall	3
Title	4
Description	3
Subjects	2
Keywords	4
Supporting metadata	3
CARE / governance	1
Completeness	3

Field presence

Field	Status	Value
title	present	Longitudinal survey of physical activity policies and practices in Arizona elementary schools
subtitle	missing	
description	present	A longitudinal survey of physical activity policies and practices in Arizona elementary schools administered via online questionnaire each fall since 2020.
subjects	present	Medicine, Health and Life Sciences
keywords	present	School children; Physical education and training; Education, Elementary; Longitudinal Studies; Arizona
creators	present	Poulos, Allison; Kulinna, Pamela
notes	missing	
language	present	English
publicationDate	present	2024-11-19
versionState	present	RELEASED
termsOfUse	missing	
accessToSources	missing	
software	present	{"softwareName": {"typeName": "softwareName", "multiple": false, "typeClass": "primitive", "value": "SPSS"}; {"softwareName": {"typeName": "softwareName", "multiple": false, "typeClass": "primitive", "value": "Microsoft Excel"}}
kindOfData	present	Online survey data

Field comments

Title: The title clearly describes the dataset.

Description: The description is brief but provides some context.

Subjects: Only one subject is listed, which may not fully capture the dataset's scope.

Keywords: The keywords are relevant and useful for discovery.

Supporting metadata: Some supporting metadata, such as creators and software, are present, but terms of use and access notes are missing.

CARE / governance: There is little evidence of consideration for community rights, governance, or ethical handling.

Strengths

- Clear title
- Relevant keywords
- Presence of creators and software information

Weaknesses

- Limited subject coverage
- Lack of terms of use and access notes
- Insufficient consideration for community rights and ethics

Missing elements

- Subtitle
- Notes
- Terms of use
- Access notes

Suggested changes

- Add more subjects to fully capture the dataset's scope
- Include terms of use and access notes
- Consider adding community rights and ethics information

Revised description

This longitudinal survey examines physical activity policies and practices in Arizona elementary schools, administered via online questionnaire since 2020. The dataset includes online survey data and was created using SPSS and Microsoft Excel.

Ticket Issue Analysis Chatbot



ASU Library Data Repository Issue Analysis

This project uses the current AskaLib tickets for the past year and gives responses to questions related to them.

Overall Statistics of the Data Repository

Which dataset subjects generate the most support?

What is the average and median time to first reply and time to close?

How many tickets are submitted per month?

What are the top 10 categories of requests based on Question + Details?

Which ASU Units or Affiliations submit the most tickets?

Metadata Analysis Chatbot

Is the metadata overall good for this repository



Metadata Analysis Chatbot

Is the metadata overall good for this repository



AI response

Based on the metadata records analyzed across multiple datasets in this repository, here is an assessment of overall metadata quality:

Strengths:

- Titles are generally clear, descriptive, and concise, aiding findability.
- Authors are well documented with names, affiliations, and ORCID identifiers, supporting authority and attribution.
- Publication dates and versioning information are consistently provided, supporting provenance and reusability.
- Licensing is explicitly stated, usually with clear open licenses like CC0 1.0 or CC BY 4.0, helping clarify reuse rights.
- Keywords and subject terms are frequently included, often from controlled vocabularies (e.g., LCSH), which enhances findability and interoperability.
- Dataset contacts are often provided with names, affiliations, and emails, supporting responsibility and contactability.
- Descriptions and related publications are included in many records, improving context and reusability.
- Metadata blocks seem standardized and structured, which helps interoperability and machine-readability.

Weaknesses / Gaps:

- Some datasets require file access requests, which may slightly reduce accessibility.
- Instrument- or software-specific details are limited overall, though one project cites specific R language and IDE versions.
- Certain metadata elements like methods, data collection instruments, or ethics statements are not prominently visible in the provided examples.
- Some datasets only include metadata records without direct data access, limiting immediate usability.
- The diversity of metadata completeness varies between datasets, with a few less detailed entries.

Summary: Overall, metadata quality in this repository appears solid with well-documented authorship, clear licensing, version control, and descriptive titles. There is good use of controlled vocabularies for keywords and subjects. However, some datasets have restricted access or limited detail on methodology, which could be improved for enhanced FAIRness and reusability. The repository shows good potential for supporting findability, accessibility (with some restrictions), interoperability, and reusability, especially with consistent metadata structure and ORCID adoption.

Metadata Analysis Chatbot

Summary: Overall, metadata quality in this repository appears solid with well-documented authorship, clear licensing, version control, and descriptive titles. There is good use of controlled vocabularies for keywords and subjects. However, some datasets have restricted access or limited detail on methodology, which could be improved for enhanced FAIRness and reusability. The repository shows good potential for supporting findability, accessibility (with some restrictions), interoperability, and reusability, especially with consistent metadata structure and ORCID adoption.



Results & Next Steps

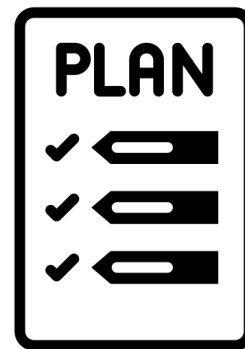
Results

Across all 3 AI tools

- Able to do multiple datasets at a time
- Currently only active datasets, but can pull from draft
- Rubric and Instructions customizable
- Results scoring not consistent
- Missing elements useful for remediation
- Works best with pre-form prompts
- Chatbot knowledge base needs to be kept up to date

What comes next?

- Continue evaluating
- Additional testing and editing of AI tools
- Explore incorporating it into current workflows
- How can this affect or assist with accessibility requirements





Questions?

Thank You!

Rachel Fernandez

Research Data Reproducibility Librarian

rachel.fernandez.1@asu.edu

Open Science and Scholarly Communication

ASU Library