

ASU Research Data Repository: Technical Infrastructure Snapshot

Arizona State University Library

Deirdre Kirmis
Rachel Fernandez

Arizona State University Library launched the ASU Library Research Data Repository using the Dataverse platform in 2020 to support research data sharing, preservation, and publication across the institution. Since its launch, the repository has grown to include multiple collections and datasets spanning a wide range of disciplines, including the social sciences, health sciences, engineering, and computing. The platform supports diverse research topics – from artificial intelligence and public health to education and the life sciences – reflecting the breadth of research activity at ASU.

Over time, the repository has also evolved from a primarily self-deposit model to a more service-oriented or self-mediated approach, with increased emphasis on consultation, curation, and review processes. This growth has shaped both the technical infrastructure and the operational practices used to support researchers and manage research data effectively.

The repository currently has 26 collections, 99 datasets, and over 12K files supporting diverse research activities.

Dataverse in Context

Core Design Principles of ASU Dataverse



Cloud-based
Built on AWS



Service-oriented
Modular and scalable



Sustainable
Reliable, secure, cost-aware

The ASU Library Research Data Repository is built on a cloud-based architecture using AWS managed services to support scalability, reliability, and operational efficiency. Traffic is filtered at the edge before entering a load-balanced application layer running on EC2 instances in private subnets.

The application interacts with a managed PostgreSQL database (RDS) and uses Amazon S3 for scalable, durable file storage. By separating compute, database, and storage

components, the system can scale efficiently and support long-term research data preservation and access.

This architecture allows the repository to adapt to growing data volumes and evolving research needs over time.

Risk Management and Security

- Layered, defense-in-depth approach across network, access, and system layers
- Controls applied at each stage to manage access, traffic, and system behavior
- Continuous vulnerability monitoring, patching, and remediation processes

Layered Security and Risk Management Approach

Network Isolation
Firewalls • Private Subnets

Access Control
IAM • MFA

Infrastructure Protection
Security Groups • WAF

Endpoint Security
HTTPD • Cloudflare • Crowdstrike

Vulnerability and Data Protection
Patching • CVEs • Backups

AI Bot Management

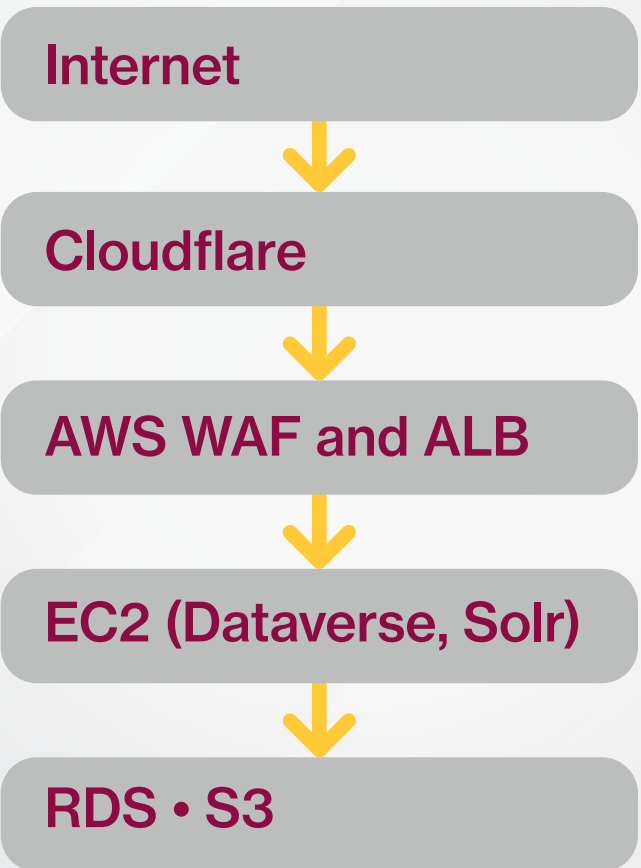
- High-volume automated traffic from bots and web crawlers
- Impacts performance, bandwidth usage, and operational workload
- Managed through layered controls across edge, AWS, and application layers

AI-Assisted Operations

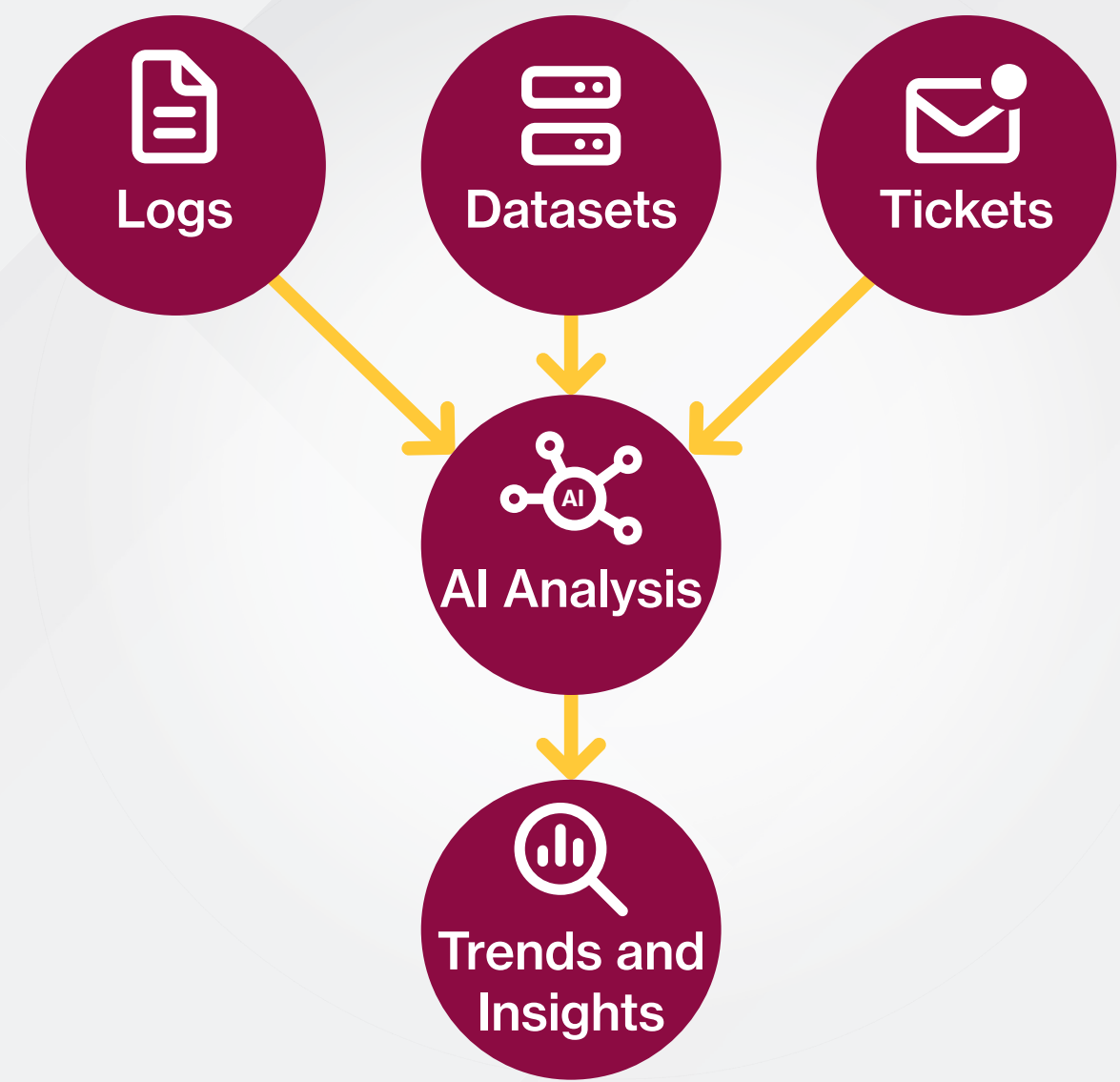
The ASU Library Research Support team is actively exploring the use of AI tools to support system operations and improve service workflows. These tools are used to analyze system activity and user interactions, helping identify patterns that may not be immediately visible through manual review.

We’re using AI in practical ways to understand and improve both system performance and user experience.

Managing Automated Traffic and Bot Activity



AI-Assisted Analysis of User Activity



AI in Operations

- Logs and traffic
- Anomaly detection
- Dataset analysis
- Support analysis

What it Enables

- Faster troubleshooting
- Better decisions
- Improved workflows
- Researcher support

From Infrastructure to Service

The technical infrastructure, operational practices, and security controls described here all serve a common purpose: supporting researchers. Decisions about architecture, storage, access, and performance directly influence and are influenced by how users deposit data, interact with the platform, and receive support throughout the research lifecycle.

In practice, there is significant overlap between technical operations and research and curatorial support. Tasks such as managing complex deposits, troubleshooting datasets, and assisting with large data uploads require close coordination between curatorial and technical teams. As a result, the technical and service layers are closely connected, with infrastructure enabling reliable, scalable, and user-focused research data support that can adapt to evolving research needs.