Selection of FEDORA with VITAL as a Digital Repository Platform for Preserving Scientific Data

Robert R. Downs[1]

[1] *rdowns@ciesin.columbia.edu*

NASA Socioeconomic Data and Applications Center (SEDAC)
Center for International Earth Science Information Network (CIESIN)
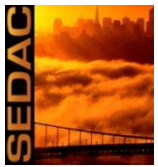The Earth Institute, Columbia University

Research Data Alliance Plenary 9
Barcelona, Spain; 5-7 April 2017

Session: Repository Platforms for Research Data IG:
Experiences with Research Data Repositories in Different Institutions and Domains
Friday, 7 April 2017; 11:00 a.m. - 12:30 p.m.

Abstract

- The Fedora-based VITAL digital asset management system is being used to preserve scientific data at the NASA Socioeconomic Data and Applications Center, SEDAC. After reviewing the digital repository platforms that were currently available, SEDAC installed and implemented the FEDORA digital repository platform in 2006. Based on its experience with FEDORA and its review of the systems and services available for preserving digital objects using FEDORA, SEDAC acquired the VITAL product and services for managing its scientific data. The decision-making process and criteria used for selecting FEDORA and VITAL are described.
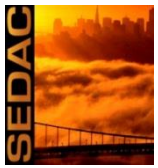
Primary Initial Pre-Fedora Selection Criteria

- Interfaces for depositors, administrators, and users
    - Access controls based on collections and roles
    - Data ingest, metadata entry, review, and revisions
- Process automation, workflow facilitation, and tracking
    - Robust, but simple to use for each population of users
    - Metrics on collections, access, and use
- Support Data Model for Long-Term Preservation
    - Archival Information Package and Dissemination Information Package for each object
    - Collections, objects, files, metadata
- Extendable to meet evolving requirements
    - Adaptable to meet new user requirements for each interface
   - Increasing requirements for routine security audits

Early Fedora Experience
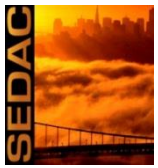
- Conducted preliminary tests with Fedora in 2005
  - Determined that implementation was feasible
- Implemented Fedora version 2.11 in 2006
  - Strong community support
  - Active developer community
- Resources needed for further development and support
  - Integration of new interface modules
  - Process support capabilities
  - Data model enhancement
  - Security issues prohibited public access

Adoption of Vital with Fedora
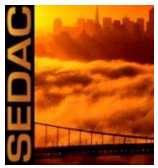
- Purchased Vital License, Implementation, and Maintenance
  - Including Vital and Handle servers
  - Obtained training for archivists and information technology staff
- Installed and maintained Vital on-site
  - Running MySQL Relational Database Management System
  - Used internally for Long-Term Archive and documents
- Continued Maintenance contract
  - Upgraded periodically from Vital 3.0 to 6.3 (Fedora 3.3)
  - Support requests have decreased in frequency with recent versions

Exploring Services to Resolve Identified Gaps
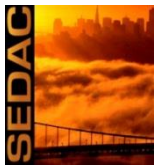
- Hosted Services
- Reduced costs
- Security for public access
- Pre-ingest services
- Enhanced provenance capture
- Scheduled recurrent integrity verification
- Integration with website and other services
- Data download user interface customization
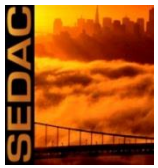- Batch transfer path to future systems

Desired Hosted Repository Services Capabilities

- ## Discovery and Access
  - Security, Discovery, Data Landing Page, User Access and Analysis

- ## Administration
  - Administration Authorization, Submission, Pre-Ingest Review, Administrative Ingest, Digital Object Identifier, File Identifier, and Administrative Support

- ## Geospatial Data Services
  - Provision of Open Geospatial Consortium (OGC) web services (Web Map Services, Web Feature Services, Web Coverage Services, and Web Map Tile Services), and Service Discovery capabilities