

Long Term Digital Preservation of Earth Science Data with Fedora

Robert R. Downs (rdowns@ciesin.columbia.edu), Alex de Sherbinin, Robert S. Chen

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

Prepared for Presentation by Alex de Sherbinin to the NASA Data Center Managers Meeting, May 4-6, 2010

Abstract: Fedora architecture enables the long-term preservation of digital scientific data. This poster presents Fedora and explores the utility of Fedora for digital preservation among the NASA Earth Science Data Centers.

Traditional Data Stewardship Practices

Each version of a data set is treated separately

Relations between data sets are manually maintained by location and directory

Manual backup, storage, and recovery is completed with people intensive processes and procedures

Metadata is manually gathered and linked to the data

People replicate data to provide new applications and services

Data Stewardship Using Digital Asset Management Systems

Manage multiple versions and complex relationships among data sets that evolve over time

Heterogeneous information contained in a set of objects and other information

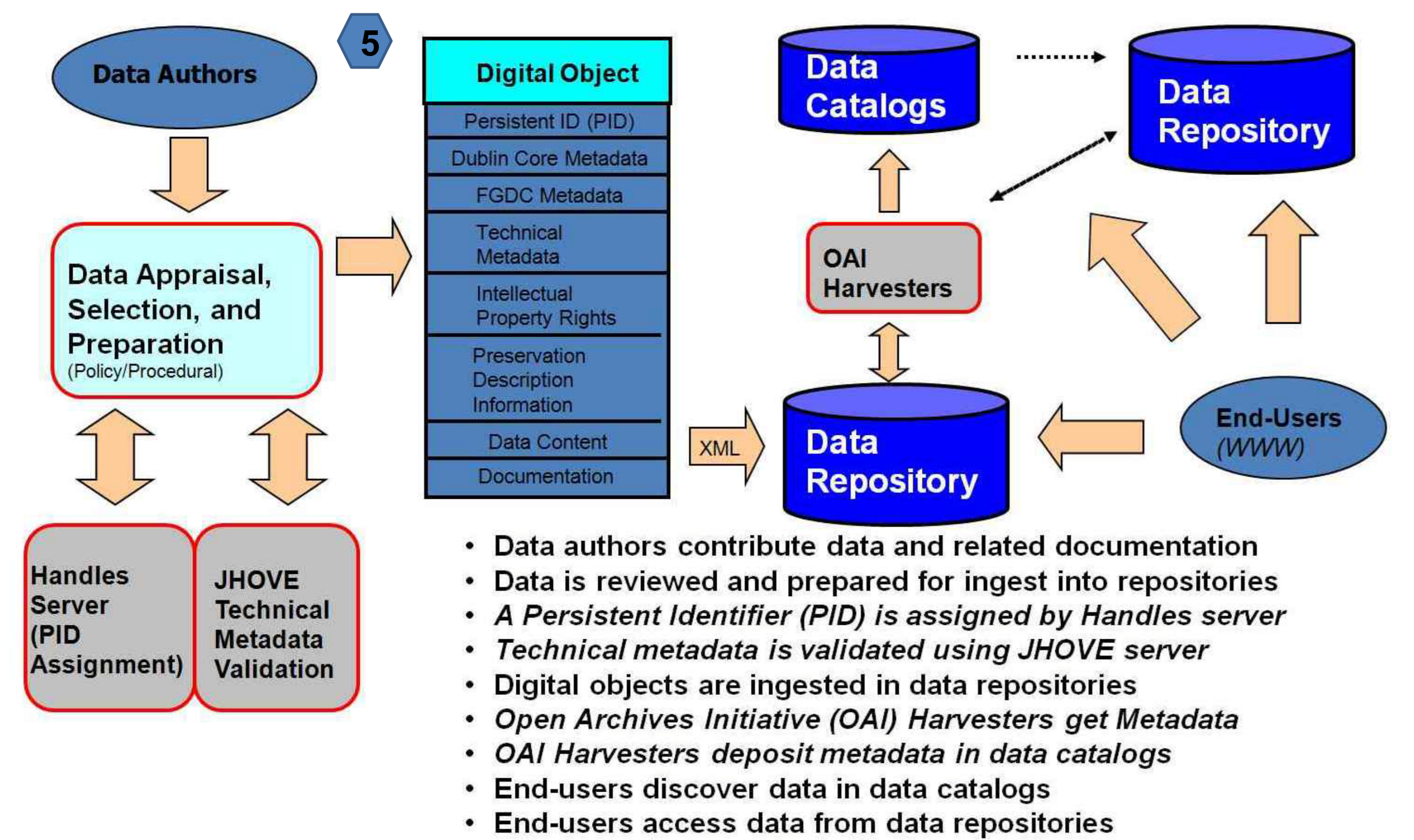
Automate backup, distributed storage, recovery, and transfer between archives

Support data deposits, descriptions, and collection of long-term preservation metadata

Reuse data flexibly to support use in different contexts, applications, and services.

- Store descriptions of relationships among digital objects
- Organized to reflect relationships with associated collections and objects
- Efficiently maintain and modify backup and recovery strategies
- Automate ingest of metadata and relevant information from community
- Open Source
- XML encapsulated objects
- Support for multiple schemas and transformations as schemas evolve
- Tie to front-end applications and services

Older data withdrawn from catalogs such as GCMD → Enables harvesting of metadata into multiple catalogs to facilitate discovery



- Data authors contribute data and related documentation
- Data is reviewed and prepared for ingest into repositories
- A *Persistent Identifier (PID)* is assigned by *Handles server*
- *Technical metadata is validated using JHOVE server*
- Digital objects are ingested in data repositories
- *Open Archives Initiative (OAI) Harvesters get Metadata*
- *OAI Harvesters deposit metadata in data catalogs*
- End-users discover data in data catalogs
- End-users access data from data repositories

N.B.: Italics indicates machine-to-machine, automated or semi-automated
Source: Downs & Chen (2008) 21st International CODATA Conference, Kyiv, Ukraine

The authors appreciate support received from NASA for the Socioeconomic Data and Applications Center (SEDAC). SEDAC is one of the Earth Science Data Centers in the NASA Earth Observing System Data and Information System (EOSDIS) and is operated by CIESIN of Columbia University under Contract NNG08HZ11C.