IN31D-1789 Expanding Access to NCAR's Digital Assets: Towards a Unified Scientific Data Management System

Don Stott and the NCAR Data Stewardship Engineering Team
National Center for Atmospheric Research, Boulder CO, USA, stott@ucar.edu

Data Stewardship Engineering Team

The Directorate of the National Center for Atmospheric Research (NCAR) created the Data Stewardship Engineering Team (DSET) to plan and implement the strategic vision of a comprehensive data discovery and access service across the organization.

The DSET is focused on improving the quality of users' experiences in finding and using NCAR's assets across all laboratories, the library, and UCAR Community Programs (UCP). Assets include data files, metadata, publications and reports, software, models and related services.

The DSET has been tasked with achieving uniform digital asset discovery across NCAR, and establishing the foundation for interoperable access with other federated systems.

Why make this change?

Users' Perspective

- Simplify and assure user data discovery
- Share metadata with outside entities
  - Export ISO standard metadata
  - Federate data discovery with others
- Provide organization-wide consulting
- Mitigate challenges like:
  - "I got lost looking for the data at NCAR"
  - "I don't know who to ask"

➢ ***Too many ad hoc systems for field projects and facilities***

➢ ***Too many orphaned datasets not in managed repositories***

Organization Perspective

- Gain functional efficiency
- Cross-organizational interconnectivity
- Improved data services by sharing:
  - Data management expertise
  - Software expertise
  - Infrastructure where appropriate
- Reduce duplication of effort
- Cost savings across the board

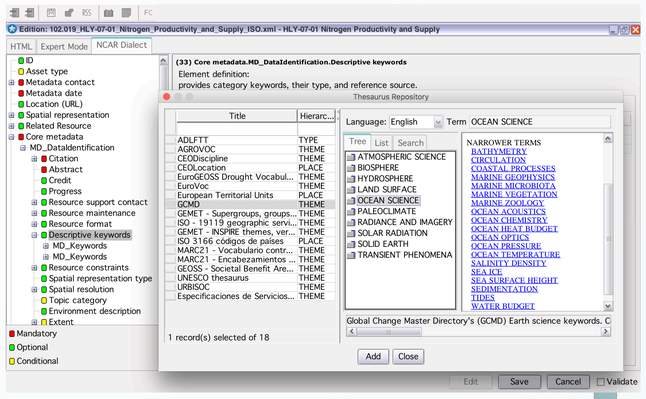
Policy and Political Perspective

- Federal
  - 2013 White House Directive, "Opening Up Access to Scientific Research"
  - 2014 White House "Climate Data Initiative"
- NSF
  - 2011 NSF data management plan requirement
  - EarthCube: Funding and collaboration opportunities
- Publication Community
  - AGU: 2013 Publications Data Policy
  - AMS: 2013 Statement on "Full and Open Access to Data"
- Internal
  - UCAR commitment for "Publication and Information Dissemination"

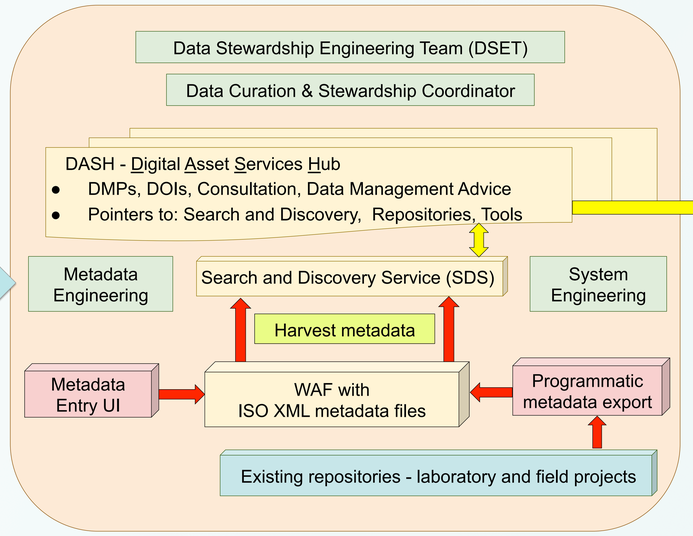
Identifying Metadata for NCAR Assets

- Survey across all NCAR labs
  - 97 respondents
  - Identified 68 persons responsible for more than 3 petabytes of digital data
- Inventory assets
  - *UK Digital Curation Centre* Data Asset Framework
  - Asset types include files and metadata, publication, images, services, and software (visualization, analysis, model codes)
  - 1/4 of the inventory respondents noted that they do not provide user support for their assets
  - Only 76% of digital assets are currently publically available
- User story sessions with each laboratory
  - Identified and ranked desired features for a Scientific Data Management System
  - User personas: asset consumers and asset providers
- Organization-wide assessment of metadata
  - conducted by the HDF Group
  - Individual lab meetings to identify key concepts
- Developed the NCAR Dialect
  - Leverages DataCite and ISO 19115 standards
  - Minimum Required and Enhanced Capabilities metadata
  - Mapped to the ISO 19139 and 19115-1 schemas

N-MD1: NCAR Metadata input form



Metadata Workflow



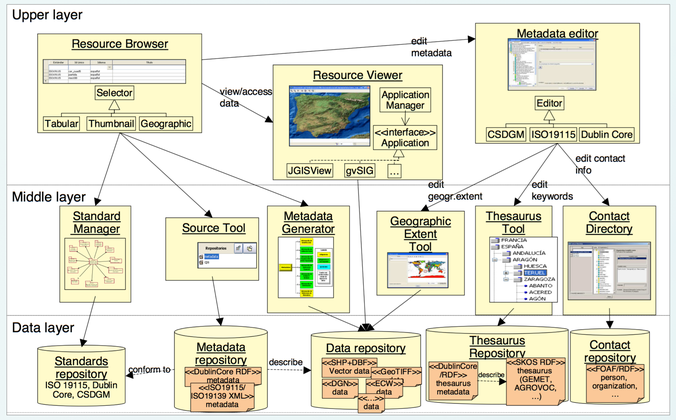
Guiding Principles of the DSET effort

- Document decisions made, the work process, technical and organizational achievements.
- Enable true cross-organizational participation by keeping a close relationship with the NCAR executive leadership and including all NCAR laboratories in the process.
- Ensure the committee includes strong representation from scientists and technical experts.

The Development System

- Search and Discovery system candidates were installed and tested in Open Lab sessions
  - Candidates: GeoPortal, CKAN, FigShare
    - CKAN (http://ckan.org) selected for the Search and Discovery system
      - In use by national research institutions, including NOAA's Data Catalog Service
      - CKAN harvests metadata files to support catalog search and discovery
    - Metadata files are stored in a Web Accessible Folder (WAF) for harvest by Search and Discovery
  - Labs produce the ISO XML metadata files for harvest
    - Labs with repositories and metadata databases can export metadata to the WAF
    - Metadata records can also be produced manually
      - CatMDEdit metadata editor (http://catmdedit.sourceforge.net) chosen as the basis for the interim manual metadata entry user interface
        - An open source software metadata editor popular in Europe
        - Produces ISO 19115 metadata for geographic datasets
      - N-MD1: an NCAR modified metadata input tool based on CatMDEdit and the NCAR Metadata Dialect
- Future directions
  - Further customization of open source software tools
  - Develop a web based metadata entry user interface to create ISO records for the WAF

CatMDEdit software Architecture



DSET members: **Steven Worley** (Computational & Information Systems Lab, CISL), **Eric Nienhouse** (CISL), **Matt Mayernik** (NCAR Library), **Steven Williams** (Earth Observing Laboratory, EOL), **Linda Cully** (EOL), **Abby Jaye** (Mesoscale & Microscale Meteorology Laboratory, MMM), **Dave Schneider** (Climate & Global Dynamics, CGD), **Ryan May** (UCAR Community Programs, Unidata), **Don Kolinski** (High Altitude Observatory, HAO), **Rebecca Centeno Elliot** (HAO), **Louisa Emmons** (Atmospheric Chemistry Observations & Modeling, ACOM), **Dan Ziskin** (ACOM), **Olga Wilhelmi** (Research Applications Laboratory, RAL), **Tor Mohling** (RAL).

DSET is funded by the Directorate of the National Center for Atmospheric Research
NCAR is supported by the US National Science Foundation.