

Digital Curation at the National Space Science Data Center

DigCCurr2007: Digital Curation In Practice
20 April 2007

Donald Sawyer/NASA/GSFC
Ed Grayzeck/NASA/GSFC

Overview

- NSSDC Requirements & Digital Curation
- NSSDC Holdings & Archival Services
- NSSDC Activities
 - Ingest
 - Archival Storage
 - Data Management
 - Access
 - Preservation Planning
 - Administration
- Key Staff Roles & Skills
- Conclusions

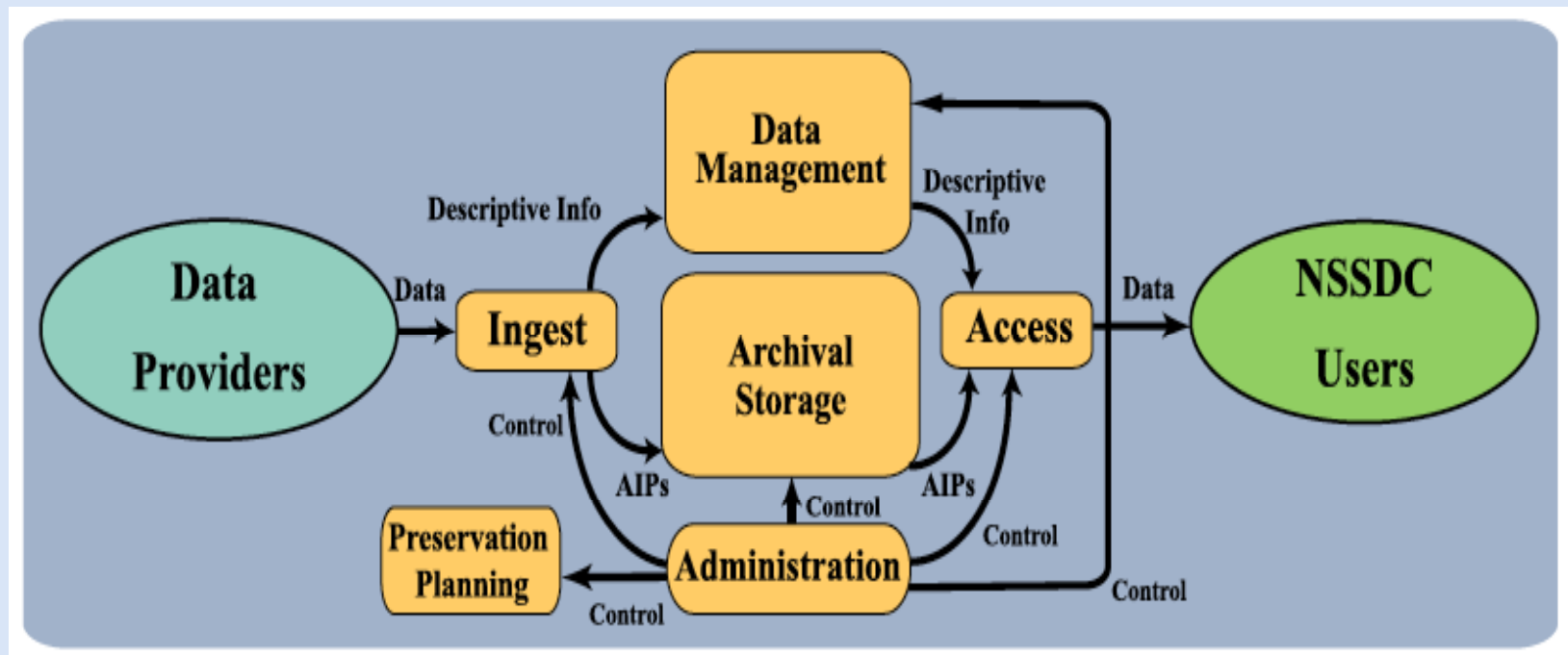
NSSDC Requirements

- **NSSDC functions as the space science permanent data/metadata repository.**
 - Work with discipline data systems, their repositories, missions, and investigators, to obtain data generated from missions
 - Hold it safely and securely for as long as it has significant value, and make it available to the research community WHEN it is not available elsewhere or when specified by other agreements.
- **NSSDC provides the space science community with data stewardship guidance and support.**
 - Data made available to the research community by various repositories should be well documented in order to support independent usability via, for example, virtual observatory access.
 - Independent usability is also critical in recognition of the reality that a large fraction of NASA and support contractor staff are likely to retire in the near future.
- **NSSDC, as a repository making unique data/metadata available, must participate in Virtual Observatory development efforts to assist in the practical evolution of those concepts.**

NSSDC Digital Curation

- Discharge NSSDC requirements for NASA's space science data
- Collective activity of entire staff
 - Many specialized skills and job positions
 - Informed by a preservation perspective
 - Cognizant of OAIS concepts and mapping to job functions

NSSDC Uses OAIS Concepts



NSSDC's Data Providers

- NASA's Space Science Active Archives (AAs), typically under written agreements (MOUs)
 - Heliophysics (2)
 - Astrophysics (6)
 - Planetary science (8 'nodes')
- Space Science Space Flight Projects
- Individual Investigators

NSSDC's Data Users

- NASA's Space Science Active Archives
- Space Science Projects
- Individual Researchers
 - Domestic
 - Foreign
- General Public
- NASA Headquarters

NSSDC's External Management

- NASA Headquarters, Science Mission Directorate (funding)
- GSFC Science and Exploration Directorate
 - Solar System Exploration Division (campus logistics and infrastructure)

NSSDC Digital Holdings

- Acquiring data for 40+ years
- Currently 47 TB, reaching 270 TB in 2010
 - Migrating 9 trk/3480 legacy tapes to superDLT
 - Over 50,000 media volumes in digital library
- 1300+ experiment from 375 US and international spacecraft
- Over 4400 data collections
 - Typically each with a large numbers of files in a single, but unique, format addressing a particular type of observation from an instrument

NSSDC Archival Services

- Three archival services defined
 - Permanent Archive
 - Long-term curation, uses AIP implementation
 - Data may be repackaged and/or transformed to maintain accessibility and usability
 - Second Archive
 - Data also held by another archive
 - Our holdings may be AIP form
 - Data may be repackaged and/or reversibly transformed
 - Backup Archive
 - Storage to support another archive's contingency plan
 - Data may or may not be AIP form
 - Data may be repackaged and/or reversibly transformed

Administration Activities

- External
 - Work MOUs with Active Archives (AAs) and missions
 - Respond to NASA Headquarter requests
 - Monitor progress of SAMPEX Resident Archive
- Internal
 - Oversee maintenance and modernization of infrastructure, including systems administration; e.g., low cost Linux
 - Manage personnel and physical space
 - Oversee refreshing of tapes in archive every 6 yrs (or less)
 - Oversee migration of legacy data from 9 trk/3480 tape archive into AIPs on superDLT

Ingest Activities

- Development
 - Develop, maintain, and enhance new AIP ingest software
 - Enhance remote Submission Information Package (SIP)/AIP creation software (MPGA) to support non-linux platforms, large SIPs, and reliable electronic delivery of SIPs
- Operations
 - Identify current/expected missions, collections; research and organize information; populate Data Management database
 - Conduct initial negotiations with data provider; help with Project Data Management Plans or MOUs
 - Review submission information to ensure adequate documentation and independent usability of data
 - Prepare for and carry out ingest for all service levels and media types; populate appropriate databases, recommend access & display options; verify/validate ingest completion

Archival Storage Activities

- Development
 - Develop upgrades to AIP storage manager
 - Develop provenance management system
 - Develop integrated document management/preservation system
- Operations
 - Manage media and AIPs for 3 service levels
 - Evaluate/integrate new data storage technology

Data Management Activities

- Development
 - Maintain and enhance Descriptive Information database to include photo searching & support automated ingest
 - Revise database to normalize and streamline infrastructure
 - Design/implement XML markup of metadata producing systems to enhance finding aids
- Operations
 - Continue/update data entry & content management
 - Perform routine interface maintenance & software migrations

Access Activities

- Development
 - Begin conversion of Descriptive Information database to web services
 - Create web services using Heliophysics data model
 - Create browse interfaces (photos, digital documents)
- Operations
 - Participate in appropriate registeries in Space Sciences, e.g., Heliophysics virtual observatories
 - Provide general request & access support, including by public

Preservation Planning Activities

- External
 - Continue participation/leadership in standards activities, e.g. Heliophysics data model
 - Monitor technology trends
 - Sponsor NASA-wide workshop on archiving & metadata standards
 - Provide curation guidance regarding documentation, database exports, preservation
- Internal
 - Recommend updates to NSSDC infrastructure, policies, procedures

Key Staff Roles/Skills

- Curation Scientists
 - PhD in space science discipline
 - Extensive data handling and analysis experience
 - Familiarity with OAIS concepts and terms
 - Supports primarily Ingest, Administration, Access
- Information Architect
 - Expertise in space science discipline
 - Extensive data handling and analysis experience
 - Thorough understanding of OAIS concepts and terms
 - Participates in development of data-related standards
 - Supports all functional areas

Key Staff Roles/Skills -2

- **Systems Engineers**
 - Extensive data handling experience
 - Extensive programming and systems development experience
 - Extensive experience with developing database applications
 - Familiarity with OAIS concepts and terms
 - Supports development of systems for all functional areas
- **Database Administrator**
 - Extensive database administration experience
 - Extensive experience in building science support database structures
 - May support all functional areas

Key Staff Roles/Skills-3

- Operations Manager
 - Operations experience
 - Familiar with computer applications
 - Basic understanding of OAIS concepts and terms
 - Supports Ingest, Archival Storage, Access
- Archive Head
 - PhD in space science discipline
 - Extensive data handling and analysis experience
 - Basic knowledge of OAIS concepts and terms
 - Supports all functional areas

Conclusions

- Need science discipline experts with curation training (curation scientists) for interfacing with data providers, data users
- Need computer professionals with curation training, working with curation scientists, for development and operation of internal systems, and to interact with similar personnel at data provider sites
- Desire data providers with ‘preservation understanding’ to make our job easier!