



Interagency Working Group  
on Digital Data

## **Interagency Working Group on Digital Data**

# **Status of Digital Data Management at the National Archives and Records Administration**

**September 21, 2006**

**Ken Thibodeau**

# Outline

1. **Types of Digital Data/Databases produced**
2. **Types of Digital Data/Databases used**
3. **Overview of digital data management process (DDMP) (if one exists) for research funded or performed**
  1. Including key standards referenced by the DDMP
4. **Most significant digital data issues**
5. **Future plans for digital data management**
6. **Comments on the IWGDD Charter**

**NARA**  
**Types of Digital Databases/Data**  
**PRODUCED OR USED**

**NONE**

**NARA**  
**Types of Digital Databases/Data**  
**PRESERVED & MANAGED**

- ✓ **Publications**
  - ✓ Papers, reports, books, etc...
- ✓ **Photographs**
- ✓ **Audio**
- ✓ **Video**
- ✓ **Measurements**
  - ✓ Physical, Biological, Chemical, Clinical etc...
- ✓ **Personnel**
- ✓ **Surveys**
- ✓ **Financial**

# NARA

## Overview of digital data management process

### DDMP

- **Scope: digital data that are or are contained in federal records**
  - Federal record, type 2: an object made or received by a federal agency in the course of its activities that contains information which has sufficient value to be kept for some time
  - Records differentiated from other types of information objects by fundamental emphasis on contextual information about provenance and lifecycle management for proper interpretation.
- **Management functions**
  - Guidance to agencies on lifecycle management of federal records
  - Preservation and access to records with long term value
    - National Archives = (def.) federal records transferred to NARA's legal custody for long-term preservation and access
    - Includes records in the physical custody of other organizations
    - Long-term records may also remain in the legal custody of originating agency
- **Key Standards**
  - OAIS and derivative products, ISO std. on Records Management,
  - Content standards; e.g., FGDC standards,
  - Various standards for encoding data; e.g., XML, PDF-A, ODF

**NARA**  
**Most significant digital data issues**  
**(relevant to IWGDD)**

- **Responsible stewardship: effective recognition of the issue of sustained preservation & access**
- **Scalable, extensible and evolvable solutions for building and operating digital repositories in a networked world**
- **Trustworthy, verifiable methods for overcoming format obsolescence & providing sustained access to authentic electronic records**
- **Resources for digital preservation**

## **Future plans for digital data management, 1**

### ● **The Electronic Records Archives**

- NARA's strategic initiative for a scalable, extensible and evolvable system to preserve and provide sustained access to any type of electronic record, including scientific & engineering data
- Includes business modernization for governmentwide lifecycle management of all types of records
- Initial Operating Capability (ingest & bitfile preservation): 9/2007
- Incremental deployments annually
- Full Operating Capability: 2011

## NARA

# Future plans for digital data management, 2

- **Partnerships**

- Affiliated Archives; i.e. another institution which preserves federal records which are legally part of the National Archives
  - Current: military academies, Yellowstone NP, GPO, others
- Other arrangements
  - Current:
    - NSF/San Diego Supercomputer Center (SDSC): collaboration on digital preservation research; preservation at SDSC of scientific data records generated under NSF funding. National Archives as fall-back repository. Possible parallel agreements with other agencies.
    - Permanent preservation of scientific records in agencies; e.g. USGS seismology data: permanent in agency; EROS Data Center: statutory authority; NOAA data centers: responsible stewardship.
    - Naval Oceanographic Office: exploring shared responsibilities for preservation & access to oceanographic data records.
  - Future: ?

## NARA

### Future plans for digital data management, 3

- **Collaboration in R&D & standards efforts to long-term preservation, access, and management.**
  - Current:
    - Group Efforts: NITR&D IWG, FGDC, Internet 2, OMG, PDES, CCSDS, InterPARES project, DLF
    - Research co-sponsorship
      - NSF
        - Prototype Federated Archives: SDSC, UMD, MIT, SLAC
        - Scalable architectures for preservation of complex objects, empirical research on data management using innovative architectures: NCSA
      - ARL: research on information assurance, GTRI
      - NNSA, Navy & Defense Logistics Agency: preservation of CAD/CAM/CAE records for operational use
      - NIST: text retrieval, robust algorithmic identification of data objects.
  - Future: More of the same & ?