

# Digitization Activities: Who, Why, What, When, and Where



Mark Martin  
DOE/OSTI  
April 23, 2009



# The Scope of the Challenge

- One row in our vault
- Approximately 800,000 documents remained to be digitized
- Over 50,000,000 pages



# Digitization – Why and What is OSTI Doing

- Why?
  - In support of our notion that access to scientific results speeds scientific discovery
  - Increased transparency for the Department
- What is OSTI doing?
  - Demand driven digitization
  - Bulk digitization

# Demand Driven Digitization

- Product based requests:
  - Science Research Connection (SRC)
  - Adopt-A-Doc
  - OpenNet
- On demand requests - These are requests from either public, DOE Federal or DOE subcontractor personnel. These requests may be via phone, mail, email , etc.
- These sources yield approximately 150-200 new pieces of full text per month.

# Bulk Digitization

- OSTI Sponsored
  - Microfiche - ~38,000 digitized from early 1990s.
- Cost Recovery Projects
  - NE and NASA - ~11,000 documents digitized
  - Geothermal - ~6,000 documents digitized
- Initiatives with Partners
  - Originating Sites – Digital Document Access Project
  - INIS Project
  - Google/Microsoft proposed projects

# Digital Document Access

- At the STIP meeting in April 2006, OSTI encouraged coordination between OSTI and the DOE sites to actively participate in an exchange of digitized documents for which OSTI already has an electronic citation. OSTI may have a paper document that has never been digitized. Several sites have digitized a significant number of documents and expressed an interest in sharing those files with OSTI.
- Patty Simmons - OSTI Point of Contact
- 2009 DDA Survey
- Results to date
  - FNAL - 1 upload of 1,097 files in July 2007, 1 Upload of 6,460 in May 2008
  - LANL - 1 upload of 8,486 files in November 2008

# INIS Collaborative Digitization

- International Nuclear Information System
  - Program of International Atomic Energy Agency (IAEA)
- OSTI has provided over 180,000 documents to INIS in microfiche format
- Digitization of historic nuclear energy research from 1970 through the early 1990s

# INIS Collaborative Digitization

- Results

- 26,000 files received to date

- List of updated records have been provided to the sites

- 24,000 in process

- These include report prefixes:

- LBL\* ORNL\* PNL PPPL\* SLAC\* UCID UC
      - LA\* UCRL\*

- 130,000 in queue

# Conclusion, Questions, and Comments

- Much to be done and OSTI is actively pursuing many paths to achieve the goal
- Question or comments?