THE NATIONAL ARCHIVES AND RECORDS ADMINISTRATION

About The National Archives

The National Archives and Records Administration (NARA) is the nation's record keeper. Of all documents and materials created in the course of business conducted by the United States Federal government, only 1%-3% are so important for legal or historical reasons that they are kept by us forever.

Our Mission: NARA serves American democracy by safeguarding and preserving the records of our Government, ensuring that the people can discover, use, and learn from this documentary heritage. We ensure continuing access to the essential documentation of the rights of American citizens and the actions of their government. We support democracy, promote civic education, and facilitate historical understanding of our national experience.

The President of the United States issued a Memorandum of November 28, 2011 to the Heads of Executive Departments and Agencies, stating, "*Records transferred to [NARA] provide the prism through which future generations will understand and learn from our actions and decisions*."

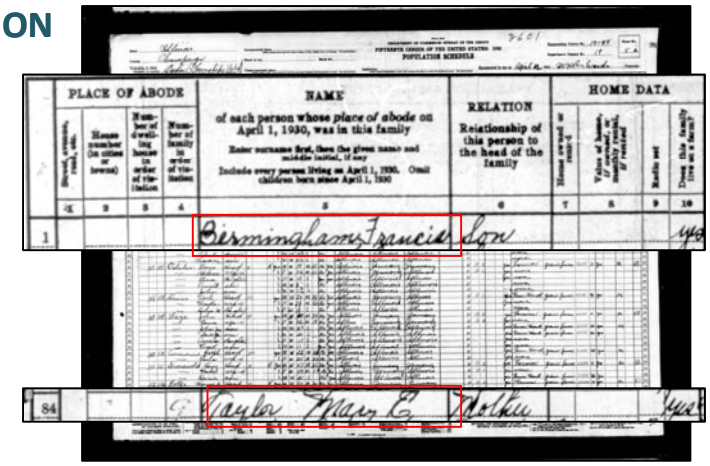
NARA is a "*Key Player in the Area of Digital Democracy*"

In December 2010, the President's Council of Advisors on Science and Technology (PCAST) released its report, *"Designing a Digital Future: Federally Funded Research and Development in Networking and Information Technology,"* specifically naming **NARA as a key player in the area of digital democracy** – primarily as a direct result of our involvement in the NITRD program. http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-nitrd-report-2010.pdf Through our continued partnership with NITRD, we can support our mission of serving American democracy through the safeguarding, preservation, and continued access to the records of our Government. NARA continues to participate in the **Inter-agency Working Group on Digital Data (IWGDD) to the National Science and Technology Council (NSTC),** including contributions to January 2009 report, "*Harnessing the Power of Digital Data for Science and Society*." http://www.nitrd.gov/About/Harnessing_Power.aspx

NARA's NITRD 2012 Symposium Demonstration

Featured NARA-NITRD Research Partner: The Image, Spatial, and Data Analysis (ISDA) Group at the National Center for Supercomputing Applications (NCSA)

The Image, Spatial, and Data Analysis Group at NCSA has been investigating the use of Word Spotting as a scalable means of searching images of handwritten forms for the 1930s and 1940s Census. Typically for this information to be searchable it must be manually transcribed, a long costly and tedious process. *Word Spotting*, on the other hand, is an automated computer vision technique that works directly from the images and attempts to rank images within a database according to how similar their contained content is to a given query image. We will discuss and demonstrate findings thus far.



The 1940 Census will be released by NARA on April 2, 2012 as scanned images available online. NCSA will demonstrate the use of *Word Spotting* to search images of handwritten words, using the 1930 and 1940 census forms.

The National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign, provides powerful computers and expert support that help thousands of scientists and engineers across the country improve our world. With the computing power available at NCSA, researchers simulate how galaxies collide and merge, how proteins fold and how molecules move through the wall of a cell, how tornadoes and hurricanes form, and other complex natural and engineered phenomena. See reverse side for additional NCSA projects and tools.

National Archives Applied Research, 8601 Adelphi Road, Rm 3311, College Park, MD 20740-6001 / Phone: 301-837-0365 / Fax:301-837-3522
E-mail: AppliedResearch@nara.gov **/Web site** www.archives.gov/applied-research/ **Facebook:** www.facebook.com/naracast Jan 2012

Archivist of the United States, David Ferriero welcomes attendees and honored guests at the NITRD Open Government R&D Summit, March 21-22, 2011 hosted by NARA.

NARA NITRD Highlights

Open Government R&D Summit, March 21-22, 2011 – Hosted in NARA's McGowan Theater, panelists discussed foundations for a robust R&D agenda that ensures the benefits of open government, in the context of Education, Health, Economic Policy, and International Open Government. http://www.nitrd.gov/openggov/

Media Access to Government Information Conference (MAGIC), April 12, 2011 – A collaborative, one-day conference sponsored by NARA and Duke University's DeWitt Wallace Center for Media and Democracy (featuring NARA NITRD research partners) that focused on exploring obstacles and opportunities for access to government information by journalists and others writing about public affairs in the digital age, and using technology to provide better access to government information. http://www.archives.gov/applied-research/events/magic.html

2006 Internet2 IDEA Award – The Transcontinental Persistent Archives Prototype (TPAP) was selected as one of four inaugural exemplary applications of advanced networking technologies. http://www.internet2.edu/idea/2006/transcontinental_persistent_archives_prototypes.html

NARA NITRD Research Partners and Project Highlights

Georgia Tech Research Institute (GTRI)

PERPOS – A suite of tools that support archival processes for Accessioning, Preservation, Arrangement, Review/Redaction, and Description of electronic records; PERPOS was initially designed to process presidential records. http://perpos.gtri.gatech.edu/publications/index.htm

Robust Digital File Format Identification – Tools and improved methods for automated file type identification and technology transfer to the National Archives of the UK for incorporation into PRONOM/DROID. http://www.gtri.gatech.edu/casestudy/identifying-files-uk-archives

The National Center for Supercomputing Applications (NCSA) University of Illinois at Urbana-Champaign

http://isda.ncsa.uiuc.edu/NARA/naraProjects.html Download software at: http://isda.ncsa.illinois.edu/drupal/software

Doc2Learn - Allows archivists to compare the contents of documents containing text, images, and vector graphics even if the documents are stored in different file formats.

File2Learn - Enables the discovery of relationships among records in large collections; establishes relationships between two dimensional engineering drawings and three dimensional CAD models in large collections of engineering records.

Polyglot - Provides an extensible, scalable, and quantifiable means of converting between file formats: Extensible: easily incorporates new conversion software; Scalable: distributes work load among parallel machines); and Quantifiable: built-in framework measures information loss across conversions.

The Renaissance Computing Institute (RENCI) University of North Carolina at Chapel Hill

CI-BER – *The Cyberinfrastructure for Billions of Electronic Records* is NARA's electronic records testbed, enabling the examination of relevant preservation/access technologies and processes for large volumes of diverse and complex electronic records that may be stored in various repositories. CI-BER includes tools supporting search of large digital records collections from tablets or smart phones based on geospatial attributes mined from individual files in the collections. http://blogs.archives.gov/online-public-access/?p=6785

iRODS – The Integrated Rule Oriented Data System is an open source data grid that helps in the organization and management of large collections of distributed digital data. http://www.dice.unc.edu

Texas Advanced Computing Center (TACC), University of Texas

Visualization Framework – A scalable solution that combines different data analysis methods into a visualization framework that facilitates understanding of large-scale collections of electronic records. http://www.tacc.utexas.edu/news/feature-stories/2011/a-window-on-the-archives-of-the-future/

National Archives Applied Research, 8601 Adelphi Road, Rm 3311, College Park, MD 20740-6001 / Phone: 301-837-0365 / Fax:301-837-3522
E-mail: AppliedResearch@nara.gov **/Web site** www.archives.gov/applied-research/ **Facebook:** www.facebook.com/naracast