

Harnessing the Wisdom of the Crowd: The Citizen Archivist Program at the National Archives

2013 Digital Humanities Conference, University of Nebraska-Lincoln, 16 July 2013

There is a great confrontation scene in the movie, *Working Girl*, where Tess McGill has to explain how she came up with the Trask Industries merger deal for which her evil boss is taking credit. She sifts through scraps of paper, newspaper clippings, and tabloid pages to reconstruct her thinking. I had a similar out of body experience in preparing to talk with you today in terms of the origins of our Citizen Archivist Program.

As you heard in Ken's much too generous introduction, I have been in the research library business, connecting users with information, longer than many of you in this room have been alive! And in that time, I have learned so much from my users—undergraduates, graduate students, faculty, researchers, administrators, and the general public. The great joy of working in public service is getting involved in the research process—learning from the researcher about his or her topic and discipline and providing the researcher not only with the information they need but, more importantly, with the tools they need to be self sufficient in their research and personal lives.

And for me the greatest joy has been what I have learned from those researchers about their discoveries in the collections. The dirty little secret of the great research libraries of the world is that they are all in need of better information about their holdings. Everyone has a backlog. Until recently, we didn't talk about them. Now we called them our "hidden collections." Everyone has unprocessed collections. Everyone would die for item level control. And EVERYONE is heavily dependent upon those researchers to help us understand what we have for it is in the use of these collections that we learn so much more about the depth and richness of our material. So...across the world researchers are making wonderful use of collections, publishing articles and books, and leaving our libraries with intellectual capital that would be useful for the library to have to better describe those collections for access by other researchers.

In my own career I can point to examples of this engagement with researchers and what I learned from them:

At MIT, the birthplace of materials science and engineering, an archaeologist was trying to prove that the oil in oil paint made its way to Renaissance Italy by way of the Spice Trade route in the late 15th century. He was analyzing scrapings from a Titian painting hanging in Florence and started thinking about when and how oil was added to pigment and then where it came from.

At Duke, where a wonderful rare book and manuscript collection was assembled at a time when others couldn't afford to buy—The Great Depression—a graduate student was using one of those collections, the John Emory Bryant Papers, in a Civil War research project and discovered a trove of letters between Bryant and his wife, Emma, concerning her gynecological problems and examinations—letters which shed light on women's medical treatment during the 19th century.

At the New York Public Library, biographer T. J. Stiles mined the papers of the New York Central and Hudson River Railroad—a collection which had never been fully described or used—in his Pulitzer Prize–winning book on Cornelius Vanderbilt, *The First Tycoon*. And detail-oriented T.J. left behind an incredible item level inventory of the collection.

And at the National Archives a book being published in September tells the fascinating story of bootleggers off the coast of New York during Prohibition, a story told from the decklogs of U.S. Coast Guard vessels in our holdings.

Those are just four examples from the many hundreds in my own career which convinced me that we learn so much more about the richness of our collections every time a researcher makes use of them and that we need to figure out a way to take advantage of that intelligence.

Add to those experiences, the shift from intermediated searching to user driven searching (yes, kids, there was a time when users weren't allowed to do their own searching!), the publication of *The Wisdom of Crowds*, the Wikipedia phenomenon, and the whole social media evolution and revolution and you a sense of what is in my Tess McGill folder providing context my own conviction that we have a lot to learn from our users.

I was wooed away from a cushy job at the New York Public Library by a good salesman from the Obama Transition Team who carried the message that the new administration was convinced that the National Archives should play an important role in its Open Government Initiative.

In articulating his commitment to transparency, collaboration, and participation in his administration, President Obama said to his senior staff on his first day in office:

"Our commitment to openness means more than simply informing the American people about how decisions are made. It means recognizing that Government does not have all the answers, and that public officials need to draw on what citizens know. And that's why, as of today, I'm directing members of my administration to find new ways of tapping the knowledge and expertise of ordinary Americans..."

"...tapping the knowledge and expertise of ordinary Americans..." I have taken that message seriously and as I have assumed my new responsibilities, engaged my staff in fulfilling that mandate.

By way of context, let me tell you a little about the National Archives and Records Administration. Created by Franklin Roosevelt in 1934, we are an independent agency of the Executive Branch. We are not part of the Library of Congress or the Smithsonian Institution—which most people think!

We are responsible for the records of the government—275 Executive Branch agencies and departments, the White House, and the Supreme Court. We provide courtesy storage for the records of Congress.

We are 3300 strong in 44 facilities across the country from Anchorage, Alaska to Atlanta, Georgia, including 13 Presidential Libraries—from Herbert Hoover in West Branch, Iowa to George W. Bush in Dallas Texas. And 80 million civilian personnel and military service records in St. Louis.

The records start with the Oaths of Allegiance signed by George Washington and his troops at Valley Forge and go all the way up to the Tweets that are being created at the White House as I am speaking this evening.

It is a collection of 12 billion pieces of paper and parchment (1.5 million trees, circles the globe 84 times), 40 million photographs, miles and miles of film and video, and the fastest growing part of the collection—electronic records. We started collecting email during the Ronald Reagan administration in 1996. Between Reagan and Bush 41 we have 2.5 million email messages. 20 million from the Clinton White House. And 210 million from Bush 43.

And because we hold the records we have some additional responsibilities:

- (1) Established by Congress, the Information Security Oversight Office (ISOO) oversees the security classification systems in both government and industry. There are currently 2500 different classification guides at work in our government. The Public Information Declassification Board has presented a report to the President with 14 recommendations to improve the current classification system;
- (2) Also established by Congress, the Office of Government Information Services (OGIS) serves as the Freedom of Information Act Ombudsman, which provides mediate services to resolve disputes between individuals making FOIA requests and administrative agencies; and
- (3) The National Declassification Center, established by the President at the end of 2009 with the mandate to review 400m pages of classified records for release by the end of 2013. These are records that date back to the First World War.

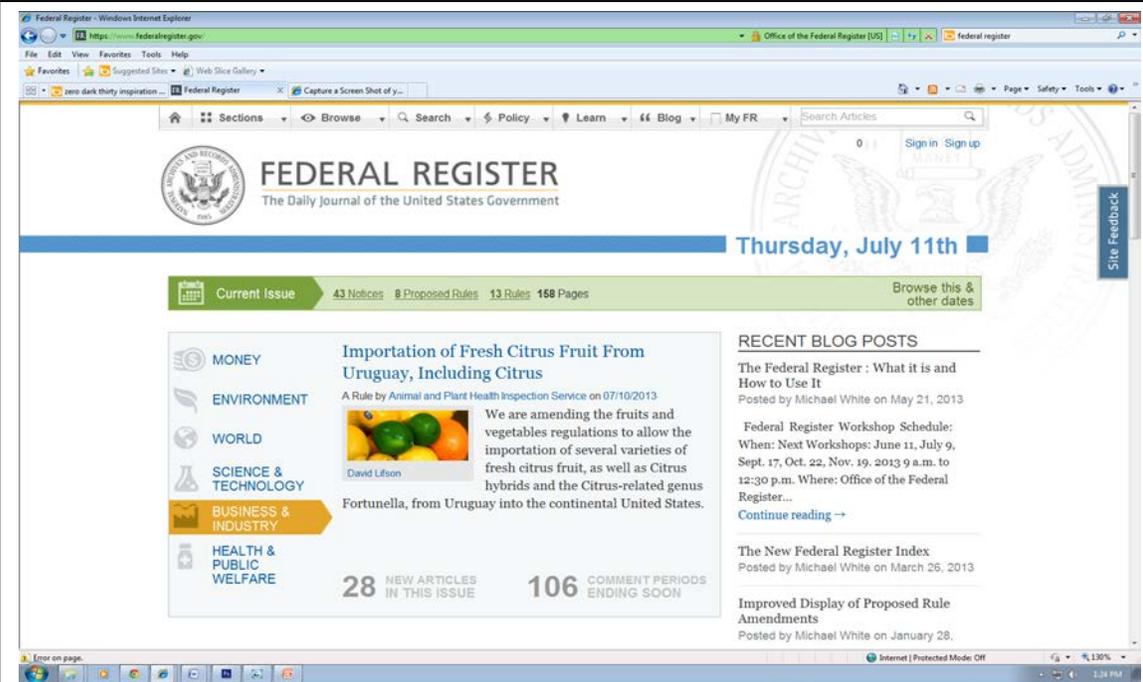


Page from the printed *Federal Register*

So, that's who we are and what we do.

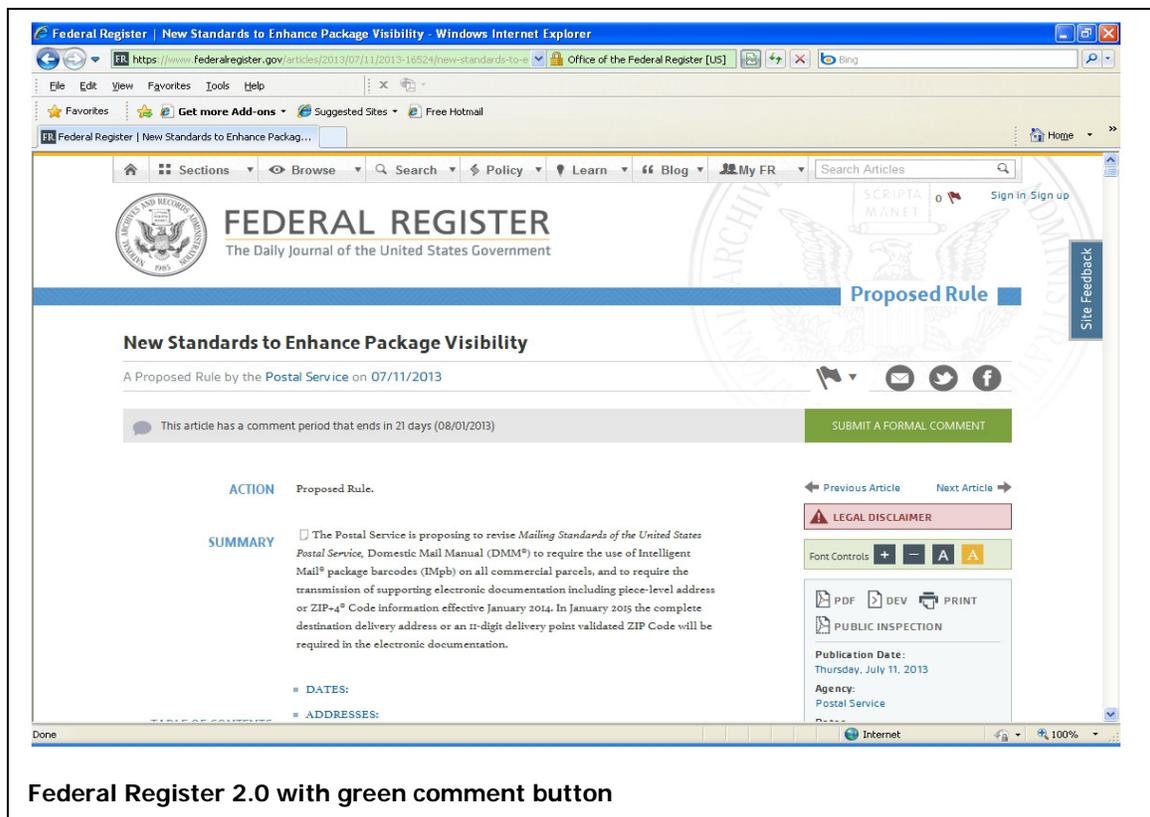
Let me now provide a glimpse of some of the things we are doing to engage the American public:

How many of you have ever used the Federal Register? Ever heard of it? It has been called the Government's Newspaper and is published by the National Archives in collaboration with the Government Printing Office. The Federal Register provides notices of public meetings, legislative hearings, grants and funding opportunities, and announcements of public interest. In addition, it publishes proposed

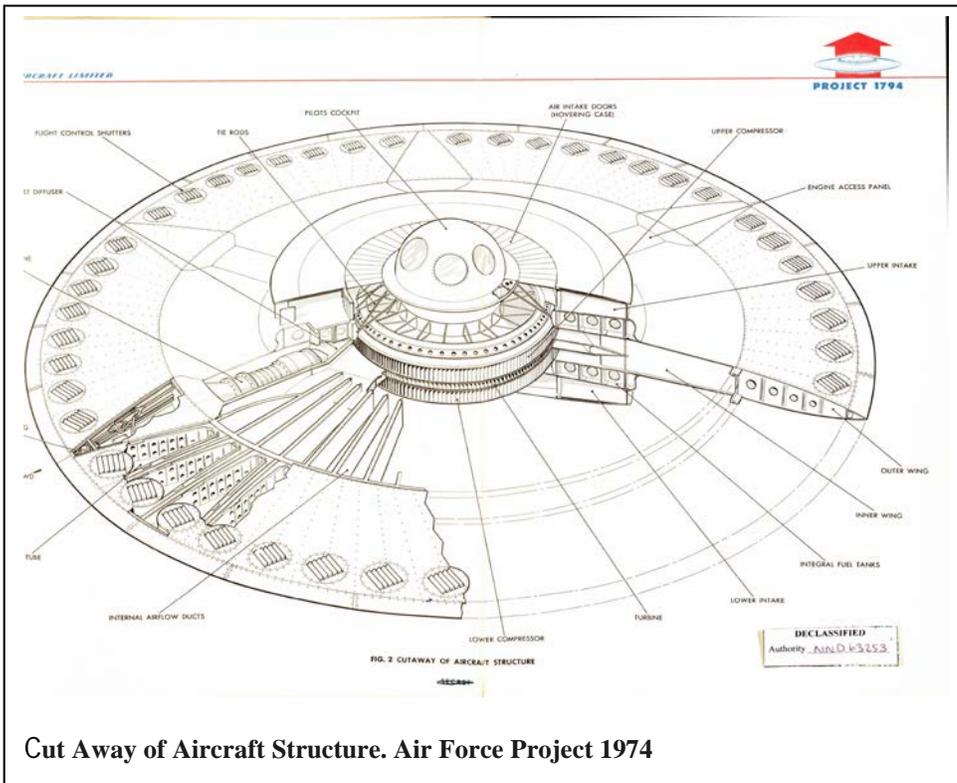


Federal Register 2.0

regulations and provides information about how to comment on these proposed regulations. On its 75th Anniversary in July of 2010 we launched Federal Register 2.0, exploiting social media tools to better connect the American public with their government. Highly graphic, clean and crisp, it is arranged in topical sections to meet user demand and interest: money, environment, world, science and technology, business and industry, and health and public welfare. The most important feature is the ability to immediately comment on proposed regulations.



Federal Register 2.0 with green comment button

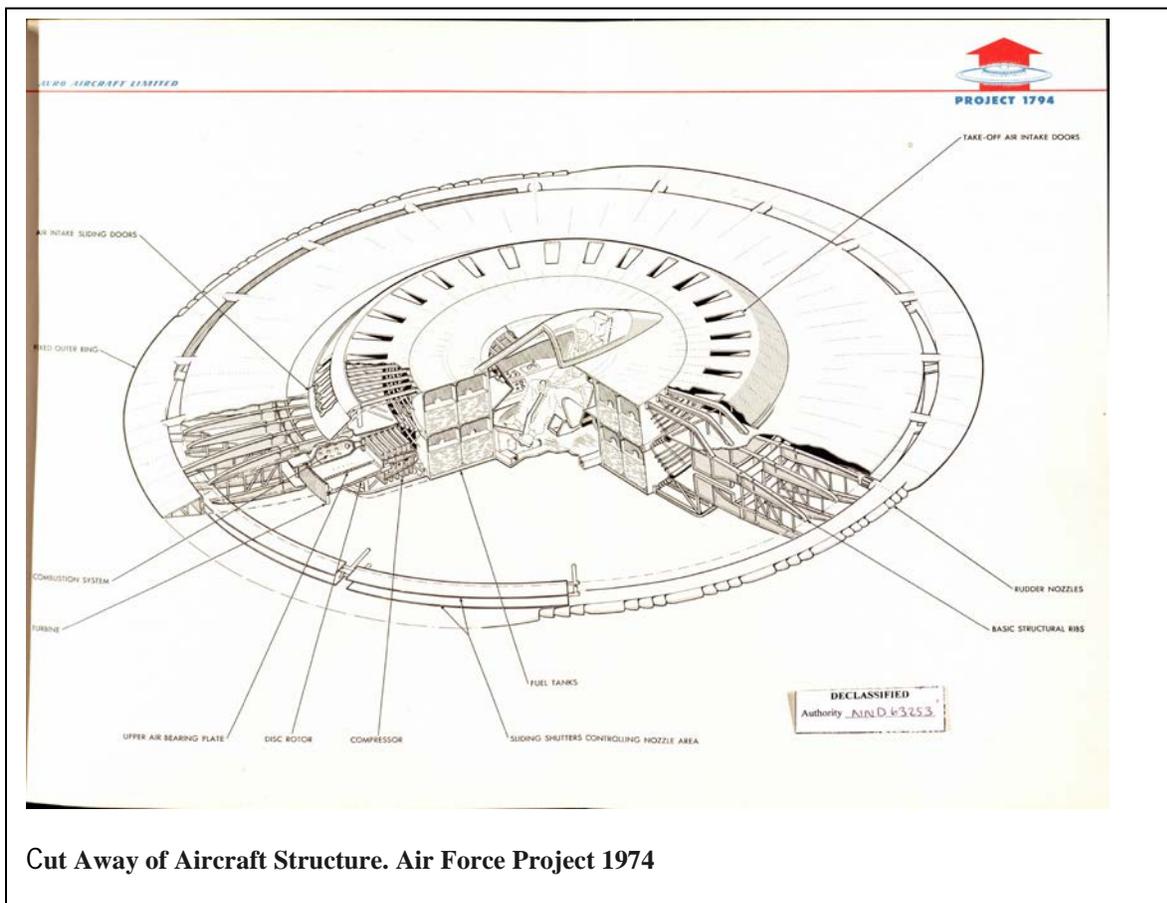


Cut Away of Aircraft Structure. Air Force Project 1974

I told you about the National Declassification Center. When we were setting it up we wanted to engage the public in letting us know which classified records were of most interest to them so that we could establish. We provided online opportunities and two public meetings in Washington which I hosted. If you go to the NDC site at archives.gov you'll find a list of those priorities based upon what we heard, mostly. The two meetings were eerily similar the room evenly split between those

interested in the Kennedy Assassination Conspiracy and those interested in UFOs. And I am so pleased that with the last couple of months we were able to deliver on at least one of these!

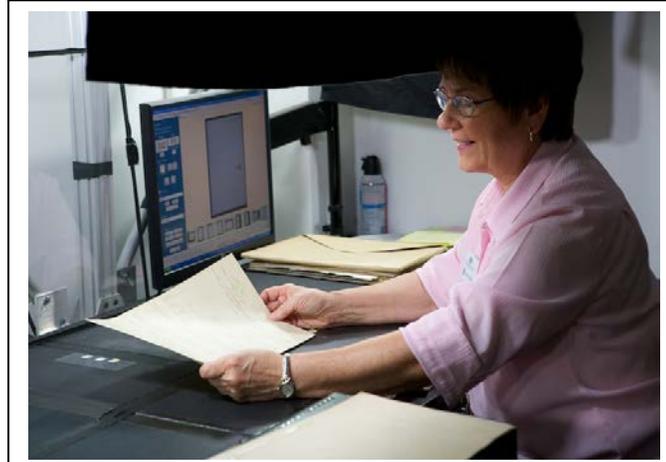
Air Force Project 1974.



Cut Away of Aircraft Structure. Air Force Project 1974



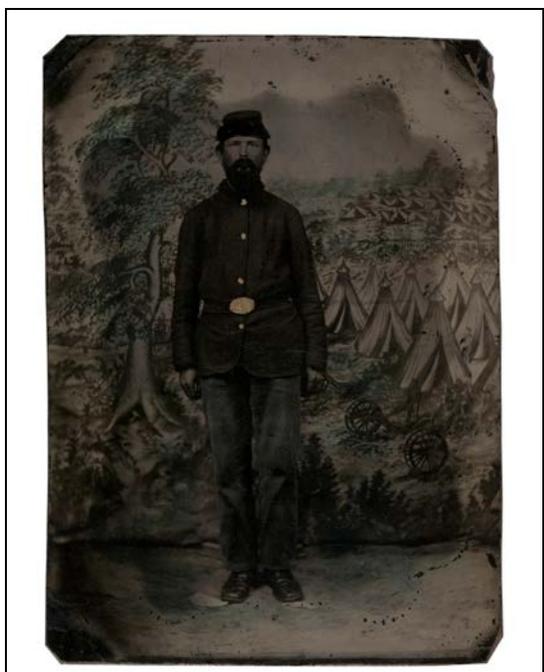
Camera operators working on the Civil War Widows' Pension Project.



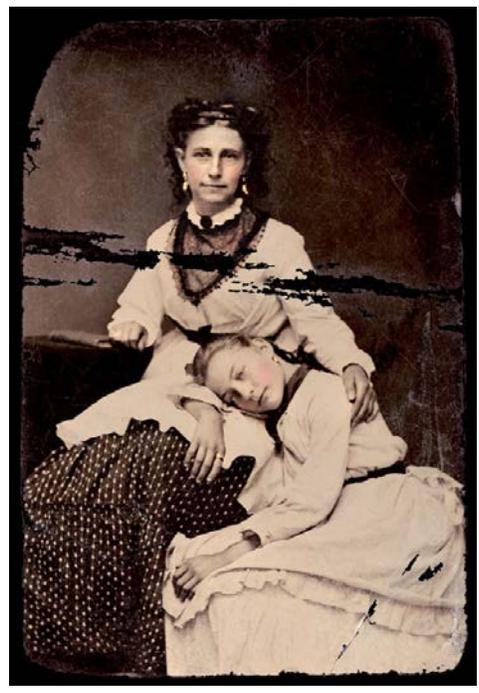
Camera operator working on the Civil War Widows' Pension Project.



William Renkel of the 17th Missouri Cavalry



William Carman of the 115th Pennsylvania Infantry

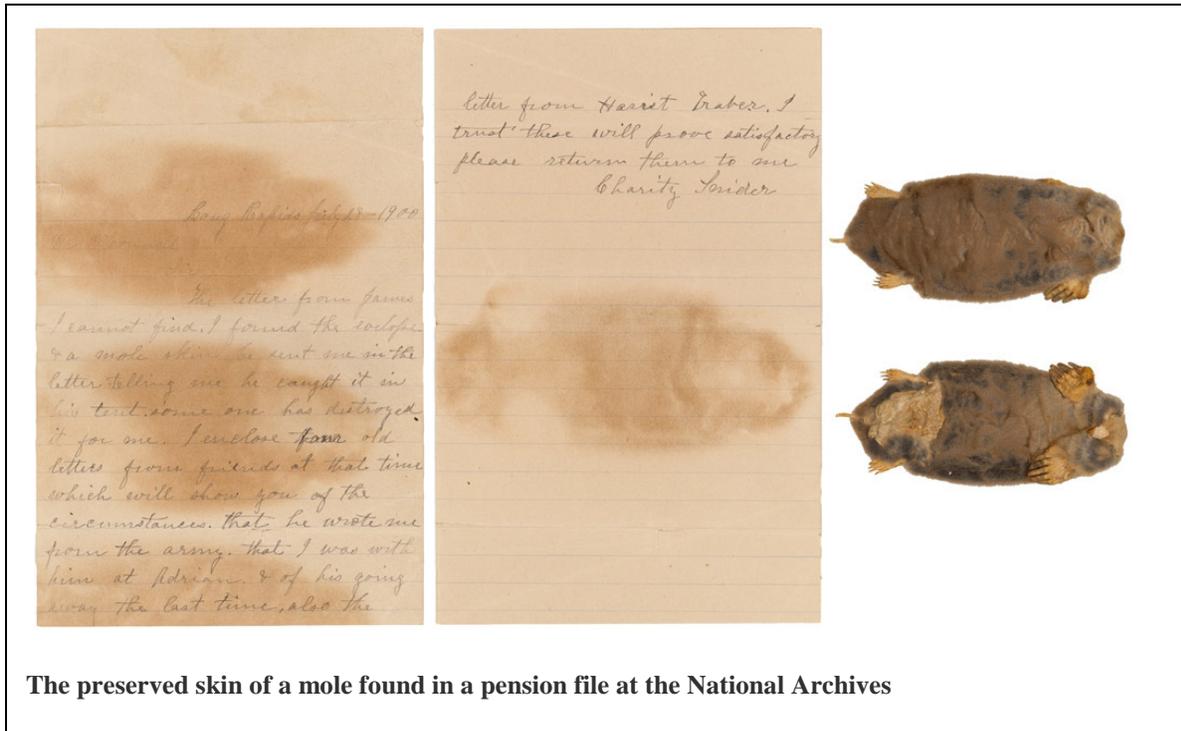


Adelia Springer's Pension file, deserted by husband

Because we have things like census records, immigration records, ships passenger lists, and pension files, genealogists comprise one of our largest user groups. Since 2007, a group of volunteers have been working on the Civil War Widow's Pension Certificate Digitization Project. Volunteers are responsible for prepping the pension records for scanning—gauging preservation needs, removing fasteners and tape, flattening, etc.—by

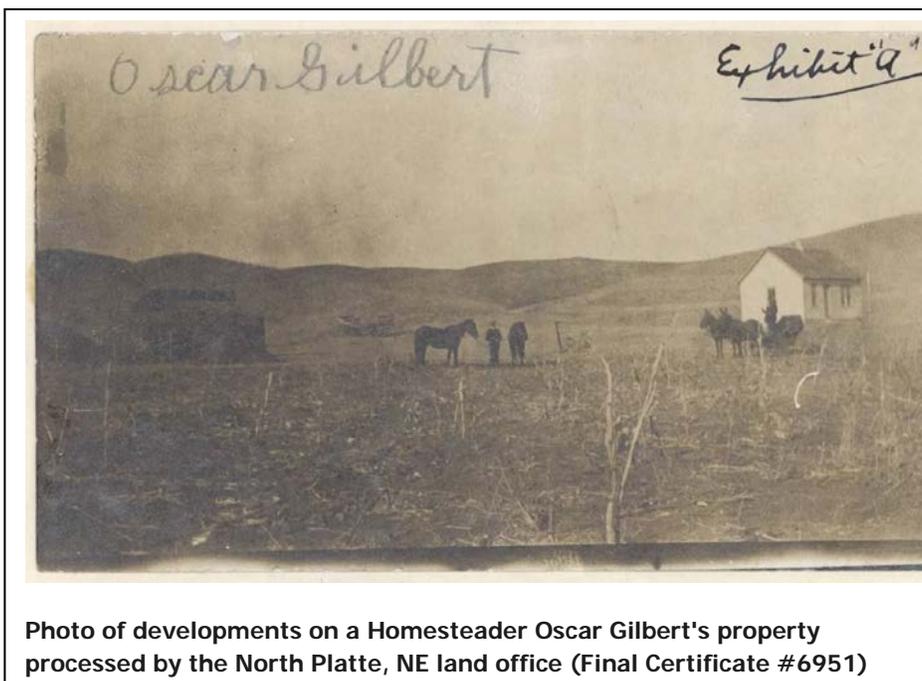
our commercial partner, Fold3. To date, 4.2 million images have been produced.

Many records contain surprises—tintypes of deceased soldiers and their families; and this interesting enclosure—a mole skin. Charity Snider applied for a widow's pension but had no marriage certificate. The only thing she had was the mole skin that her husband had sent her—captured in his tent. She got her pension!



Similar work is being done in collaboration with the National Park Service, Family Search, Fold3, and the University of Nebraska to digitize our Homestead Land Office records. Volunteers from the Mormon Church are preparing records for digitization. We started with Nebraska because

of the Homestead National Monument here. To date, 67,000 files have been digitized producing more than 1.5 million images searchable by land office, homesteader's name, and township on the Fold3 and National Archives sites. And I am very pleased that the University of Nebraska has students creating additional metadata to further our understanding of these records.





Homestead Final Certificate for William Merriam; Final Certificate #112, Dakota City, NE, land office

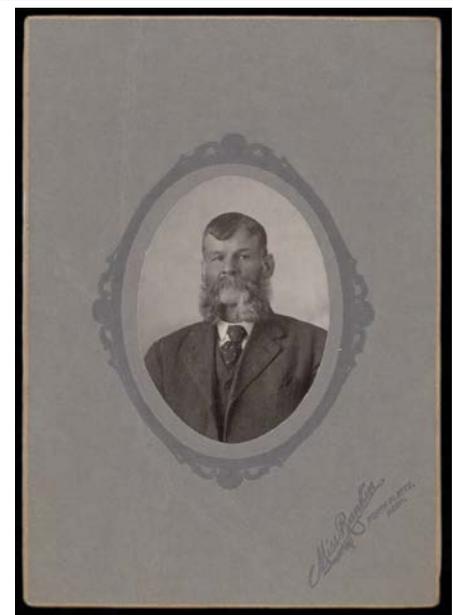
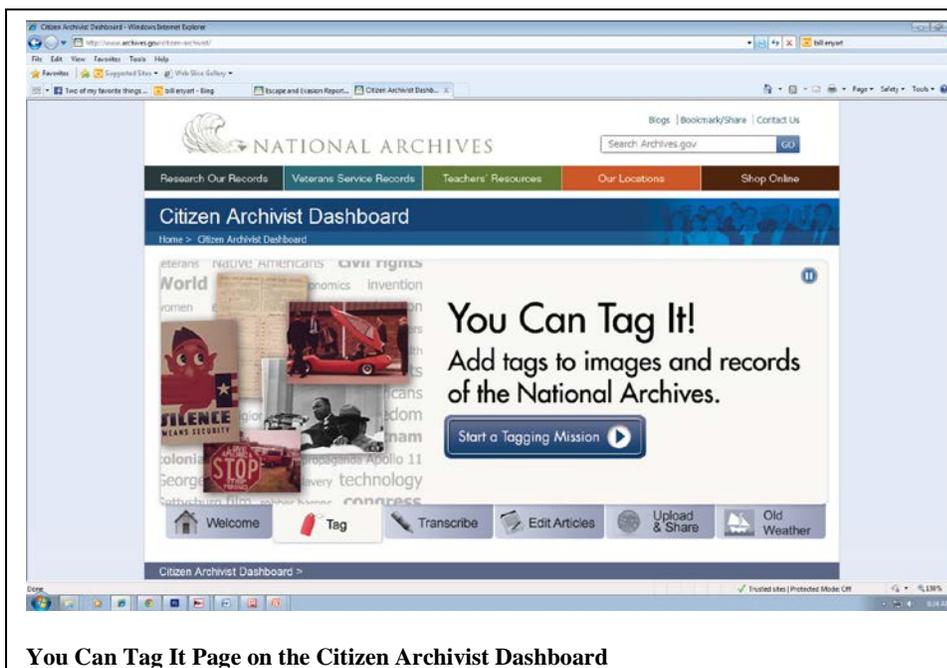


Photo of Benjamin O. Shedd—a homesteader in North Platte, Nebraska.

I mentioned the administration's Open Government focus. The Open Government Directive requires every agency to create an Open Government Plan and the flagship of my agency's plan is our Citizen Archivist Initiative envisioned to encourage substantive contributions like social metadata tags, transcripts, and digital images to improve access to our records. Every day in 44 research rooms across the country, individuals are capturing information from the records of our government that are held in trust by the National Archives. These users are scanning, writing books, conducting genealogy research and learning more about our records than we do. The initiative is our first digital attempt to harness this knowledge.

The Initiative is an adaptation of the long-standing tradition of crowdsourcing in science. Citizen science projects engage amateurs and nonprofessionals in scientific research, like bird sightings and categorizing galaxies. And we are using digital technology to engage recreational archivists.



You Can Tag It Page on the Citizen Archivist Dashboard

Our Citizen Archivist Dashboard will give you a sense of the range of opportunities we are providing for engagement:

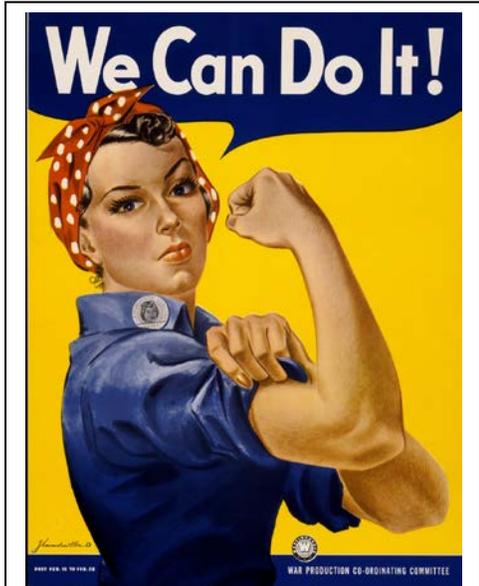
"You Can Tag It" is a call to action for tagging. The addition of small keywords improves search results and increases access to the records. This



"Someone Talked," World War II Poster



"He's Watching YOU," World War II Poster



"We Can Do It!" World War II Poster

type of social metadata can be fun and an easy way for contributions that cumulatively can make a big impact. To encourage the public to tag we've developed "tagging missions" that take users through our online catalog to tag particular topics, like World War II Posters, Photos of the Civil Rights March in Washington, DC, and EPA Photographs from the 1970s. Users have added more than 100k tags to our images on Flickr and we have incorporated more than 10k into our catalog.



A Young Civil Rights Demonstrator at the March on Washington, August 28, 1963



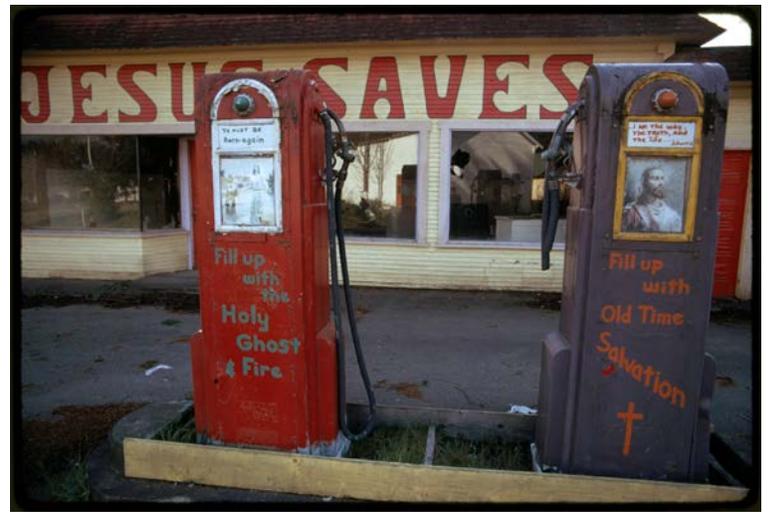
Civil rights leaders, including the Rev. Martin Luther King, Jr., at the March on Washington



View of marchers along the Mall in Washington, DC



Italian neighborhood, Chicago near Southwest side



Gasoline stations abandoned during the fuel crisis in the winter of 1973-74 were sometimes used for other purposes. This station at Potlatch, Washington, west of Olympia was turned into a religious meeting hall. Signs painted on the gas pumps proclaim "Fill up with the Holy Ghost ... and Salvation," April 1974



Inexpensive retirement hotels are a hallmark of the South Beach area. A favored place is the front porch, where residents sit and chat or watch the activities on the beach, June 1973

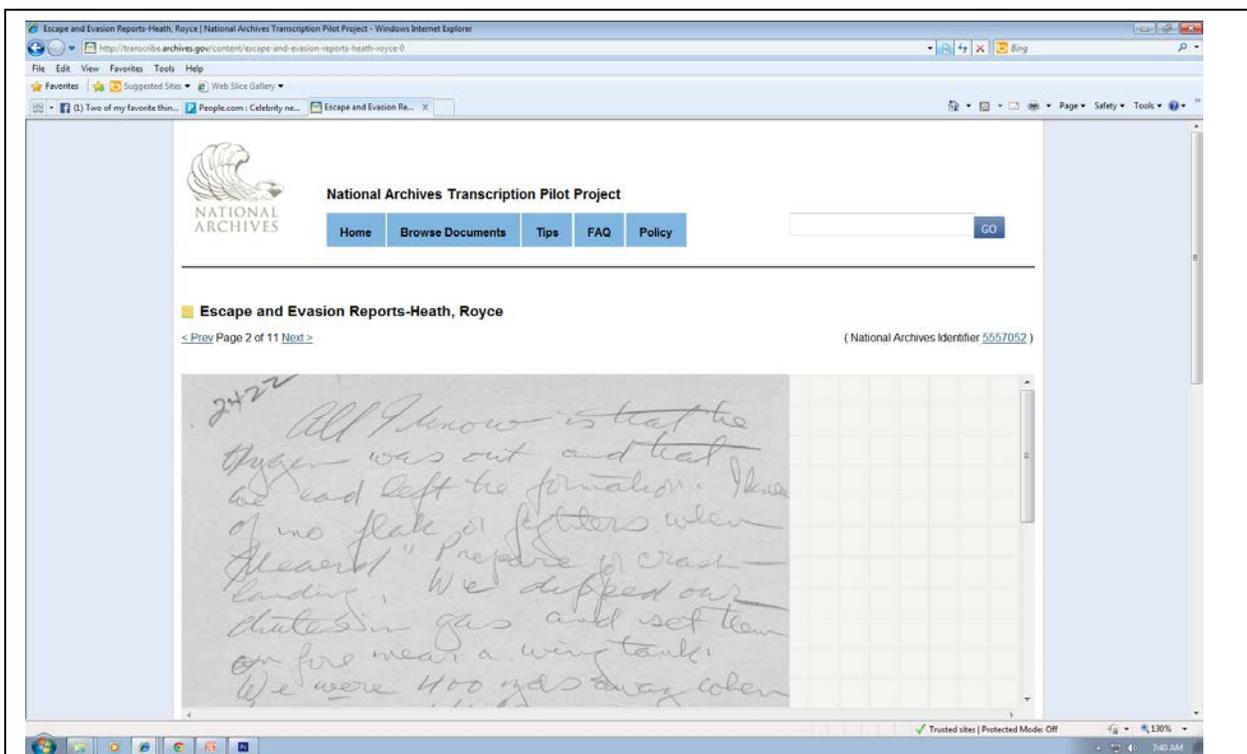
A large portion of the 12 billion records in our holdings are in cursive, and cursive has disappeared from kindergarten and elementary school curricula so we have a growing audience who can't read the records. In response we've developed a Transcription Pilot site on Drupal where we have loaded records to be transcribed. Our first 1000 pages were transcribed within two weeks.

As part of our Citizen Archivist Initiative we recruited a Wikipedian in Residence to work with us on project and fostering a collaboration with the Wikipedian community. We've held scan-a-thons (one of my favorites was State Department staff scanning State Department photographs) and events for Wikipedians in Washington, at our

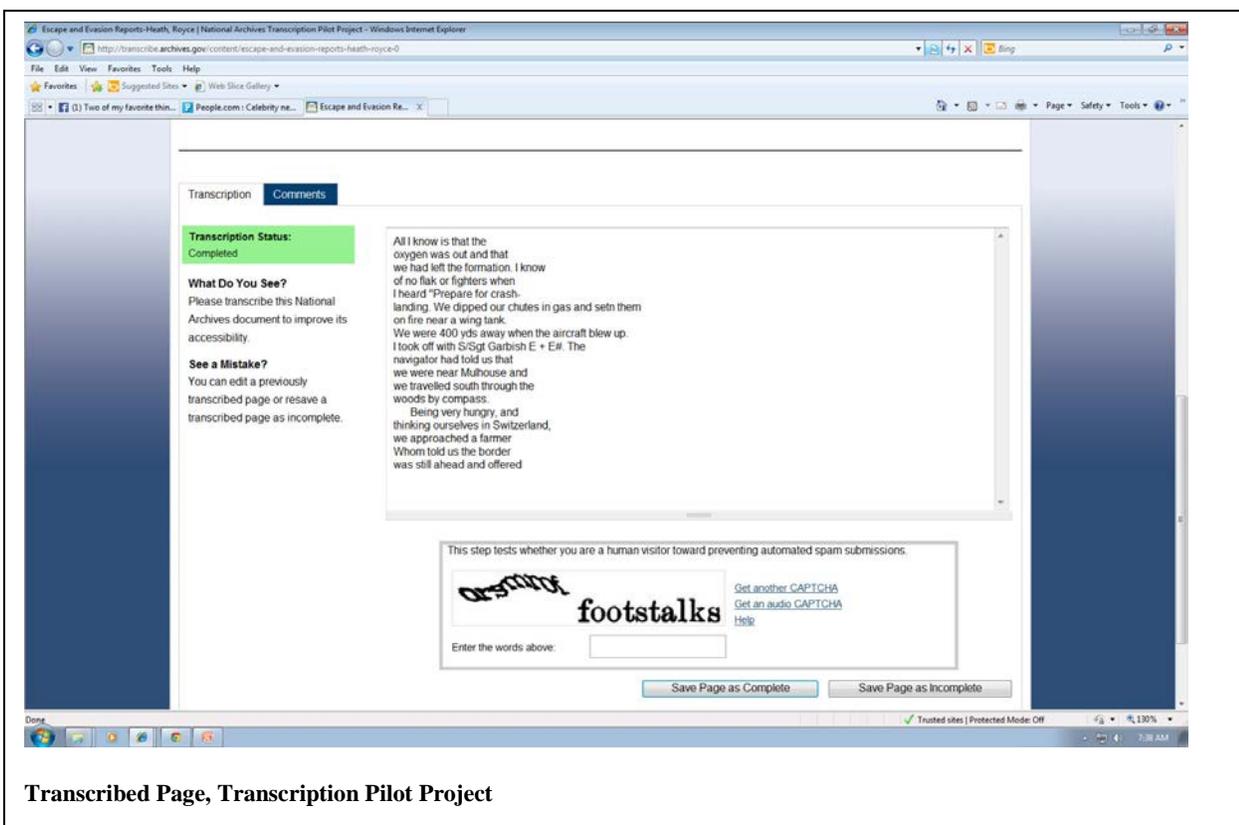
regional facilities, and at our Presidential Libraries. We have uploaded more than 100,000 digital images of National Archives records to Wikipedia Commons, and we have several projects for Wikipedians to incorporate these images into Wikipedia articles. In one month alone, the numbers of estimated views of



Citizen Archivist Dashboard Screen Shot for Transcribing Records



Page to be Transcribed, Transcription Pilot Project

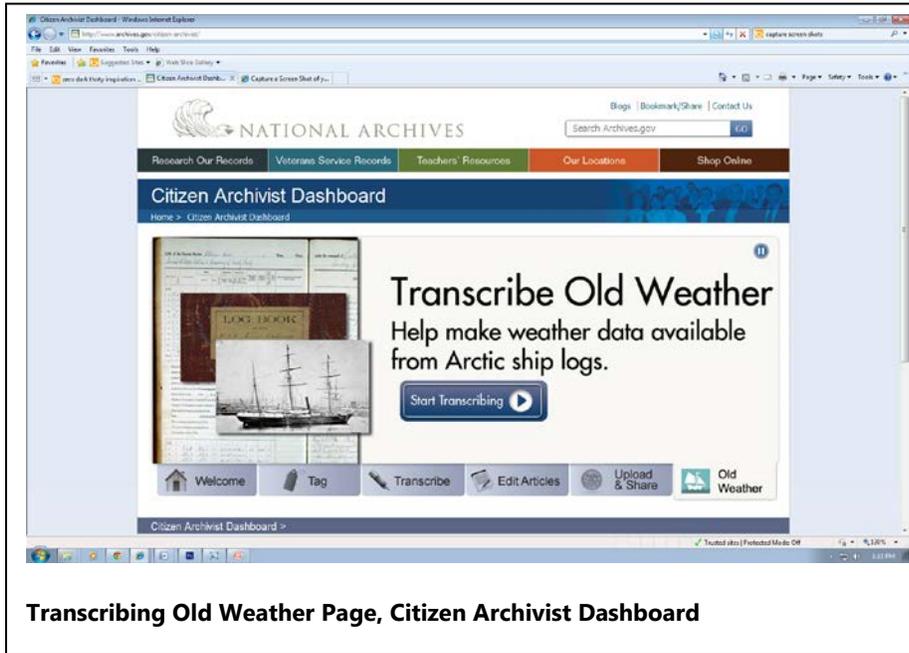


Transcribed Page, Transcription Pilot Project

Wikipedia articles with our content is estimated at more than 100m.

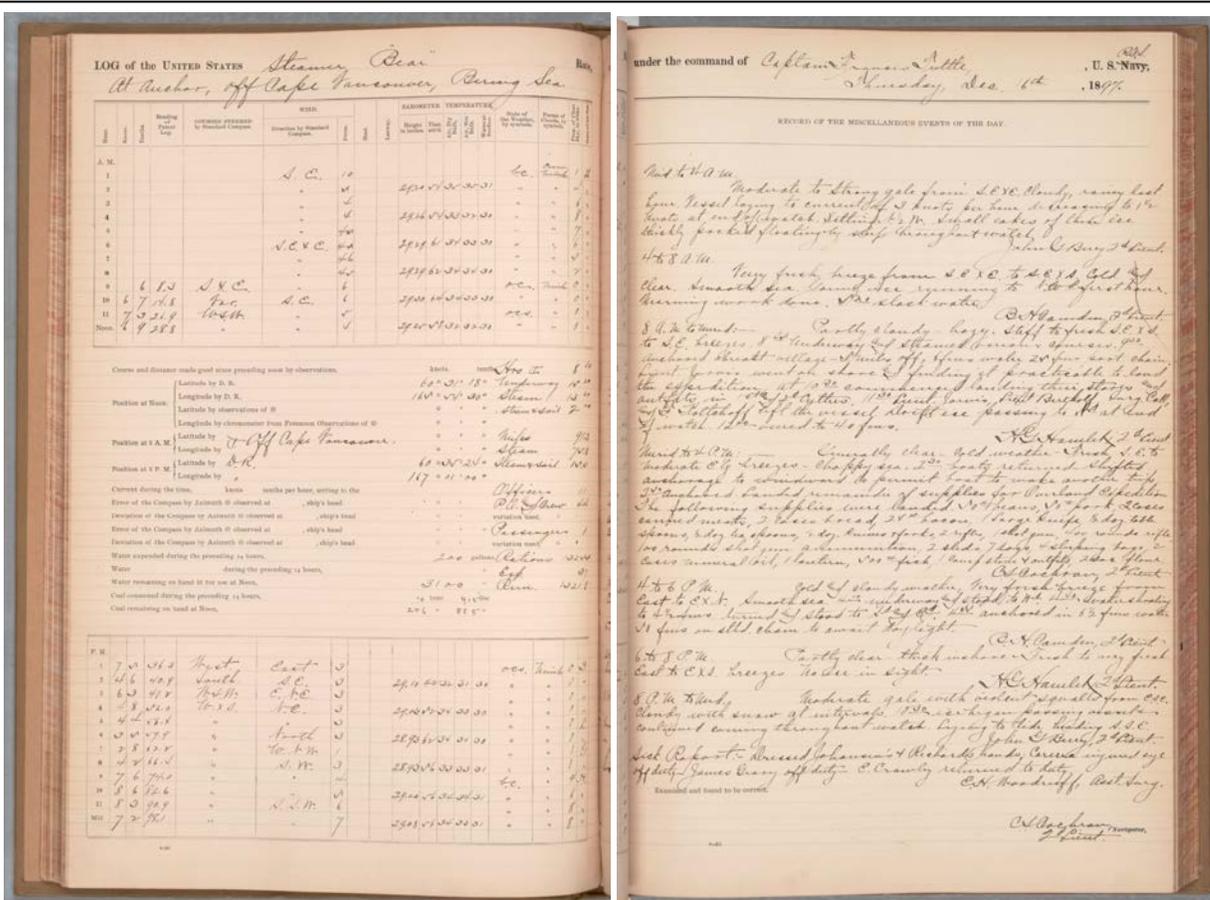
We are encouraging the public to upload photos of their own documents, working to expand this feature to include web uploads and mobile scanning. Researchers have already approached

us and are interested in donating thousands of images of records and metadata. Our challenge: make it simple.



Transcribing Old Weather Page, Citizen Archivist Dashboard

A terrific partnership with the National Oceanic and Atmospheric Administration is focused on the digitization of ship logs from the Navy, Coast Guard, and Revenue Cutters from the pre-Civil War period through World War II. These records are being featured on OldWeather which is encouraging people to transcribe weather information in the logs so that scientists can analyze



The log of the USRC Bear records the Overland Relief Expedition of the winter of 1897–98 to save the lives of 245 whalers trapped in the Arctic Ocean by ice around their ships near Point Barrow, Alaska.

weather observations made by United States ships.

In April of 2012 we released the 1940 census. The first digital release. And we partnered with FamilySearch on a community indexing project, using their indexing software. Working state by state, more than 170k volunteers created the name index—132m names in just five months.

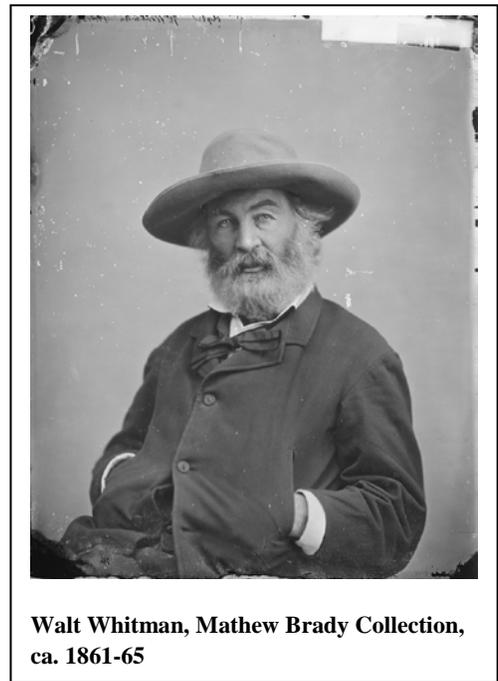
I am particularly proud of the fact that we were one of 20 Federal agencies participating in the National Day of Civil Hacking, expanding our reach beyond traditional researchers and engaging the community of “citizen developers.” We have made available more historical datasets for bulk downloads in our catalog and on Data.gov. We’ve developed several challenges around datasets, including mapping and visualizations of historical data related to World War II, the Korean War, and the Vietnam War. This data originated in Federal agencies and was collected in punch card computer systems. We’ve provided search access to these datasets (primarily for genealogical search) and now we’d like to expand this to include visualization and ways to provide access to aggregate data. And we have a challenge we hope the developer



Ken Price, citizen archivist

community will be interested in—a mobile app called “Pocket Archivist” making it easy for researchers to take images of records, record institutional-related metadata, and upload the images so they can be incorporated into our online catalog.

I would be remiss, in this setting, if I didn’t single out one of my own favorite Citizen Archivists. I have been fortunate to work in institutions with significant



Walt Whitman, Mathew Brady Collection, ca. 1861-65

Walt Whitman collections. Even MIT had a copy of Leaves of Grass! At both Duke and the New York Public Library I was pleased to be able to contribute to the work of Ken Price and Walt Whitman Digital Archive. Ken’s experience captures the essence of the citizen archivist

Concord Massachusetts,
10 January, 1863

Dear Sir,

Mr. Walt Whitman, of New York, writes me that he is seeking employment in the public service in Washington, & perhaps some application on his part has already been made to yourself.

Will you permit me to say

that he is known to me as a man of strong original genius, combining, with marked eccentricities, great powers & valuable traits of character: a self-relying large-hearted man, much beloved by his friends; entirely patriotic & benevolent in his theory, talks, & practice. If his

writings are in certain points open to criticism, they show extraordinary power, & are more deeply American, Democratic, & in the interest of political liberty, than those of any other poet.

A man of his talents & dispositions will quickly make himself useful, and, if the government has work that he can

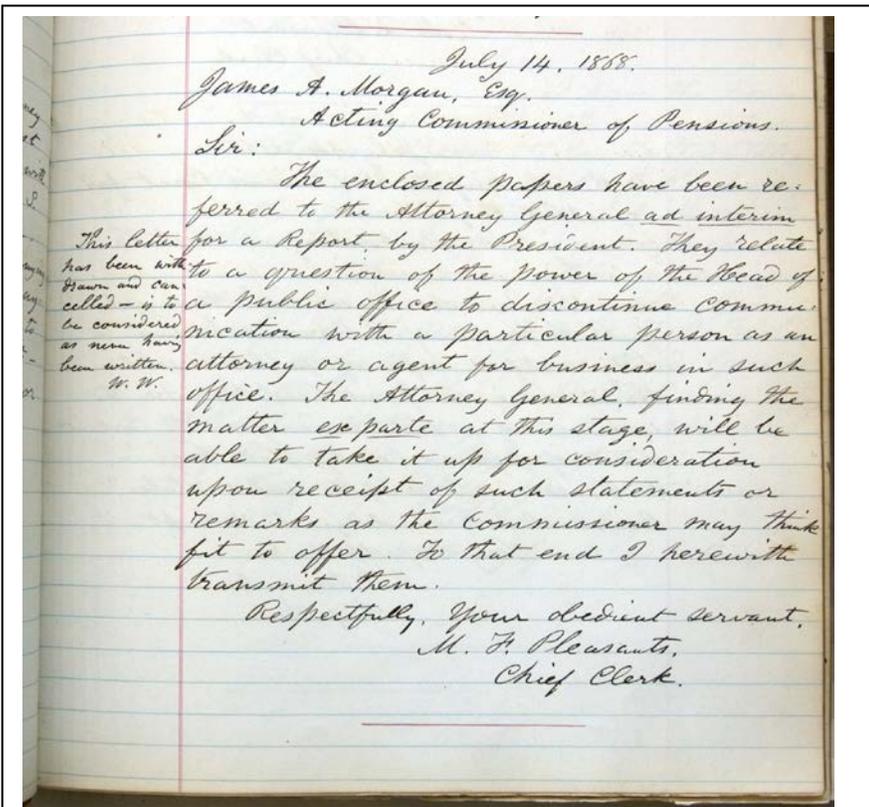
do, I think it may safely be said that it has called to its side more valuable aid than it bargained for.

With entire respect,
Your obedient servant,
R. W. Emerson.

Hon. Salmon P. Chase,
Secretary of the Treasury.

1863 Jan 10

Ralph Waldo Emerson’s recommendation letter in Walt Whitman’s personnel folder, Department of the Treasury



A letter from Chief Clerk M.F. Pleasants to James A. Morgan, dated July 14, 1868. The document is written in Whitman's hand and includes a note in the left-hand margin initialed "W.W." This is one of the few documents that is marked with his initials or signature.

program—a researcher who shares his discoveries so that we learning more about our own holdings. Knowing that Walt Whitman served as a clerk in the Attorney General's office after the Civil War, Ken went through those records at the National Archives and discovered 3000 documents in Whitman's hand. What a trove of new information about Whitman's life and the opportunity for scholars to mine this new information to discover influences on his later work. A wonderful example of the citizen archivist in action. Thanks, Ken.

As you can see, from our traditional users to new communities of technical developers, we are reaching out to the public—"tapping the knowledge and expertise of ordinary Americans"-- to help us with our most mission-critical efforts—providing access to the permanent records of the

government . I do, indeed, "hear America singing, what varied carols I hear" and what beautiful music we are making.