



NATIONAL
ARCHIVES

OFFICE *of*
INSPECTOR GENERAL

Date : February 21, 2012

Reply to

Attn of : Office of Inspector General (OIG)

Subject : Management Letter No. 12-06, Access to Records in the Base Electronic Records
Archive System

To : David S. Ferriero, Archivist of the United States (N)

On January 19, 2012, the OIG office met with NARA officials to discuss access to unstructured records stored in the Electronic Records Archives (ERA). The purpose of this memorandum is to formally bring to your attention conditions which impact NARA's ability to provide access to the essential evidence documenting the rights of citizens, the actions of Federal officials, and the national experience. These conditions are further amplified by the Presidential Memorandum on Managing Government Records which highlights the importance of electronic records, and the need for their effective management.

The ERA system was to enable NARA to realize its strategic mission with regards to electronic records, and business transactions associated with NARA records. The NARA OIG has issued numerous reports on the ERA program which identified cost overruns, failed deliverables, and functional deficiencies. These prior reports were issued while NARA was in a developmental project phase with the prime contractor, Lockheed Martin (LM). The ERA program is now in an Operations & Maintenance (O&M) phase with constrained funding. Specific to the base ERA system (excluding the Presidential and Congressional "instances"), what was developed at a cost of over \$383 million is a shell of that which was envisioned at program inception, and one that fails to meet even the most basic of requirements initially defined in the contract.

The following outlines some of the primary functional failings or deficiencies existing in ERA:

- There are no reliable automated tools, or staff, available to validate that Federal records transferred by originating agencies to be ingested into base ERA are free of national security classified content. NARA officials involved in this process are aware this will create an unmanageable backlog of records which are not even ingested into ERA. Indeed, it is assumed a large volume of records will have to be returned to the originating agency for review when the potential for classified content is detected (even if it is a false-positive). NARA officials acknowledge this may delay public access for many years.

- The same problems related to screening for classified information exist for personally identifiable information (PII). There is no reliable automated tool for screening for PII, and the manual process proposed is too labor intensive to be feasible. Even if minimal PII is found in a record, base ERA has no redaction capability to provide the public a releasable version of the record.
- Unstructured records ingested into base ERA would be subject to archival treatment consistent with that afforded existing paper records. This includes such things as formulating records descriptions, finding aids, etc. It would be manual and labor intensive. However, no meaningful additional staff resources are, or will be, available to NARA in the foreseeable future to accomplish this work.
- The public will only be able to search for records in base ERA by contacting an Archivist (thus replicating a paper process in this the digital age), unless special treatment has been afforded to a record by having it moved from base ERA to the On-Line Public Access (OPA) system. OPA is not a part of ERA and currently houses just a test sample of 28 files of base ERA data. There is currently no means nor process to ingest or transfer meaningful quantities of records from base ERA to OPA.
- The public will not have efficient access to the unstructured records, such as emails or word processing files, housed in base ERA. The vast majority of records will not be afforded special treatment and copied into OPA. Those files which remain in base ERA are not full-text searchable, even by NARA archivists. Thus, even if a researcher knows an exact phrase used in a Word document, they will still have to use a search process replicating the paper system to retrieve many possible records and manually search them.
- Notwithstanding the previous bullets, should an unstructured base ERA record be identified and made available to a requester, the record would be in the format or language it was created in, not as envisioned as an “asset” independent of the hardware or software in which it was created. Thus the recipient may not be able to readily view the document as it may be in a computer code or language no longer readily available.
- An over the horizon system which would afford users direct access from any location is now one in which (unless the file can be downloaded) a NARA archivist must be contacted to search for a record, and a fee structure or “cost recovery fee” for born-digital records has been defined. The current fees are set at \$15 per file when the order contains 10 or fewer files, and \$13 per file for orders of 11 or more files.

In plain text, the ERA system at “full operating capability” and at the OMB mandated end of the developmental phase (September 30, 2011) is neither functional nor operational in meeting the basic needs and requirements of our users, the American people. Records that are “clean” and scheduled to afford the public access cannot be readily identified, located and presented to interested parties.

We believe it is imperative that NARA's stakeholders be provided clear and unfiltered information as to the status of ERA. Per the ERA Base User Liaison, the last revision of the "ERA Misconceptions and Facts" page on Archives.gov was in April 2011. On January 9, 2012 eight questions were transmitted by the Inspector General to the Chief Information Officer (CIO). In the correspondence, the CIO was invited to "share these questions with your colleagues as necessary. Please note for all of these questions we are only referring to records the public has the right to view with appropriate redactions (PII, privacy, etc.)." The responses to the eight questions are included in their entirety in Attachment A.

The OIG plans to perform a series of audits on the life cycle of electronic records in base ERA, from their ingestion through preservation and dissemination to customers, during FY 2012 which will focus on the conditions mentioned above. If you have any questions concerning the information presented in this Management Letter, please contact me at (301) 837-1532.



Paul Brachfeld
Inspector General

Attachment A

Question 1.

I am interested in unstructured records (word processing files, email files, etc.) which are being initially ingested into base ERA. I am not referencing unstructured records which were incorporated into AAD, ARC or other legacy systems first before being transferred over to base ERA; or those records selected from base ERA and copied back over to legacy systems so that OPA can search them. How can researchers access and explore unstructured records which have only been ingested into base ERA at the present time, and do not reside in any other form or system?

Response: We never envisioned, nor do we ever intend, that public users would access records in Base. Because many records in Base contain sensitive information exempt under FOIA, public access is not permissible. Rather, following archival screening and any necessary withdrawal or redaction, public access to ERA records will be through OPA. We view the ERA Base repository as analogous to the stacks, where researchers cannot roam freely to access what they choose. As with all archival research, researchers start with a search of record descriptions in ARC/OPA. If the researcher finds descriptions of records of interest but the records themselves are not available online as part of the description, the researcher contacts the electronic records reference staff to request access (including obtaining copies of) the records in question. This process is parallel to the process researchers use to get any records in physical format. Therefore, if a researcher is interested in a series containing thousands of PDFs and there is no index to documents, the user could obtain copies of all of the PDFs (at a fee) to browse. In addition, a researcher will be able to "explore" any set of unstructured electronic records online once the records have been transferred into OPA.

Question 2.

If researchers can access and explore these files, what is the methodology a researcher would follow? I know for specific structured data-files requested from base a NARA archivist can populate media and transmit it to a requestor? Thus if a researcher already knows a specific file was ingested into ERA, and knows accession number or other unique identifier, then the file can be pulled. What I want to know is what does a researcher do who only knows they are interested in a topic and is looking for files on that topic do? If there is a methodology, please provide an illustrative case where this has occurred (remember I am interested in unstructured records that did not already reside in one of legacy systems and is not from the Presidential or Congressional instance).

Response: See response to Question 1.

Question 3.

Currently, there is a format and structured followed for researchers to access paper records which happens all the time in our research rooms across the country. How does/will that process play

out in the very near future with records ingested into base ERA that have yet to be uploaded into the public interface (OPA)? Is that process and capability defined and can you provide it to us?

Response: As with all archival research, researchers start with a search of record descriptions in ARC/OPA. If the researcher finds descriptions of records of interest but the records themselves are not available online as part of the description, the researcher contacts the electronic records reference staff to access (including obtaining copies of) the records in question. This process is parallel to the process researchers use to get any records in physical formats. If a researcher is interested in a series containing thousands of PDFs and there is no index to documents, the user could obtain copies of all of the PDFs (at a fee) to browse. The process of providing reference service for electronic records is defined in SOPs which we will be happy to share.

Question 4.

- A) What is the volume of unstructured records expected to be ingested into base ERA over the next few months, years etc., and
- B) do we have a definitive described process and capability to respond to researchers requests for these digital records?
- C) Do we have an estimate as to how many requests we are to have for unstructured records that are coming to the base ERA system from Federal agencies?

Response:

A) We do not have specific volume numbers for future transfers of unstructured records. We have estimates of e-records in any format that we expect to be transferred over the short term but these are not broken down by structured and unstructured and we do not know exactly when they will come in.

B) As with all archival research, researchers start with a search of record descriptions in ARC/OPA. If the researcher finds descriptions of records of interest but the records themselves are not available online as part of the description, the researcher contacts the electronic records reference staff to access (including obtaining copies of) the records in question. This process is parallel to the process researchers use to get any records in physical formats. The electronic records reference unit has SOPs for responding to requests, which they would be happy to share.

C) We cannot estimate future reference demand. Based on Electronic Records Section staff experience, since fiscal year 2011, out of the approximately 2000 research inquiries received, there have been maybe a dozen at most inquiries for unstructured electronic records (PDFs, emails). As best as staff can recollect, only one inquiry was for unstructured electronic records that were only in Base ERA

Question 5.

If additional resources and funds will be required to facilitate the public's access to records ingested into base ERA has this been defined and communicated to our oversight committees and OMB? If so, please provide.

Response: We've talked to OMB about our needs through 2013 and presented projections for the exhibit 300. Project level details for public access have not been provided since public access through OPA is still our planned approach and there will be no separate method of public access

to Base records. The FY 2013 request is under an OMB embargo until it is delivered to Congress on February 6.

The recently signed FY 2012 budget marked the first time that ERA funding was provided as an operations and maintenance (O&M) line within NARA's OE account. In the course of discussions that led to passage of that budget, House and Senate Appropriations Committees staffs were briefed on the O&M contract. There were no discussions about additional funds above the FY 2012 request; rather, the most recent discussions dealt with how much the Committees could cut ERA O&M below the request without significant disruptions to the program.

Question 6.

If someone requests access to an unstructured base ERA record, if we can provide them the records, will it be in original language/format or restructured in some manner?

Response: At this point in time we can only provide exact copies of unstructured electronic records (PDFs, HTML, Word, etc). The copies will be in the format as transferred by the agency. Format migration is a desired enhancement expected to be developed at some point in the future.

Question 7.

Is there a charge for researchers who obtain structured or unstructured data from base ERA? If so, how is that charge defined?

Response: Normal fees would apply. The current schedule for reproduction orders, including copies of digital files, is available here: <http://www.archives.gov/research/order/fees.html>. The National Archives Trust Fund has set the cost-recovery fee for "born-digital" records, whether from Base ERA or legacy systems, at a \$15 fee per file when the order contains 10 or fewer files and for orders of 11 or more files, the fee is \$13 per file. There are no charges for electronic records available for download via ARC/OPA or for searching via AAD.

Question 8.

Behind all the complexity I simply want to know the answer to a fundamental question I have been asking for a decade which goes something like this. If a person wants to access a digital record that has been ingested into base ERA (and only resides there) how will they be able to locate and access the record from their desktop or portable device?

Response: See response to Question 1.