



# *National Archives and Records Administration*

8601 Adelphi Road  
College Park, Maryland 20740-6001

Date : April 23, 2010

Reply to

Attn of : Office of Inspector General (OIG)

Subject : Management Letter No. 10-10, Concerns with the Electronic Records Archives System's Ability to Conduct Full-Text Searches

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

The purpose of this management letter is to define our concerns as to the capacity and capability of the Electronic Records Archives (ERA) system to search the records which it will eventually store. The ERA Requirements Document (RD) dated October 17, 2005 defines the systems core requirements, and RD Sections 19 and 20 (attached) focus on the search and access capabilities of ERA. Based upon our interpretation, the RD calls for a system which at Final Operating Capability (FOC) would ingest, preserve and facilitate authorized user search, retrieval and access to all data which presided in each record maintained in ERA. For example, we believe ERA should be able to search the full text of an email's body, not just the subject line or delivery addresses. It has come to our attention NARA program officials may decide to limit ERA's searching functions short of full-text searches<sup>1</sup> due to the costs involved. With FOC looming in 2012, NARA has yet to make this crucial decision, nor have they alerted appropriators of the resource issues involved.

The ERA Program Director was asked if ERA users would have full-text search capability for all data ingested into ERA, and if that was the case, did NARA have the equipment and funding necessary either in place or requested to accomplish it. In response the ERA Program Director stated:

Full text search is a capability that is built-in with the Vivisimo Search Engine, the search engine selected for the ERA. The question is does it make sense to do full text searching on every record stored in ERA? Even if limited to a specific data type, such as text, this would be very resource intensive both in terms of the size of the index and in compute cycles to generate and maintain the indexes and to execute queries. This is a policy decision that will be made by NARA. According to our engineering staff, indexing for full text searching requires about a 3x increase in storage. These are implementation questions that will be answered during the next 8-12 months as the analysis, design and development of the production public access piece of ERA progresses.

---

<sup>1</sup> A search that compares every word in a document, as opposed to searching an abstract or a set of keywords associated with the document.

When directly asked if NARA did decide to do full text searching, was the funding already built into the current contract increment's development cost (increment 3) or would funding have to be built into the next increment (increment 4), the ERA Program Director stated:

The funding would come from increment 4 and/or increment 5. Beyond FOC, as the archive continues to grow, it will be necessary to add more storage as necessary.

Thus our understanding is that, while the ERA system will employ a publically-available search engine which should have full search capability, at this time no determination has been made as to whether this search engine will be able to actually search all of ERA's data. Rather, this overarching question is still on the table for perhaps another full year even though we have communicated to our stakeholders that FOC will arrive in March 2012. If the decision is made to enable full-text search and access capabilities, NARA would require significant additional funding and resources prior to FOC. NARA officials were unable to provide the OIG with any funding requests in this regard. This decision may have an impact on funding for increments 4 and 5, as well as out years, or may even result in the need for a follow-on contract. Lacking additional funding streams, per NARA's own definition, we would be unable to support full-text search of our electronic holdings. Rather, ERA users would be faced with the prospect of limited search capability and diminished capacity to navigate through NARA's vast holdings.

With FOC and the end of the ERA development contract looming, NARA management needs to make this critical decision in a timely manner and communicate it to our stakeholders. If you have any questions, or require additional information, please do not hesitate to contact me.



Paul Brachfeld  
Inspector General

cc: NH (M. Morphy)

Attachment:

ERA Requirements Document excerpt, dated October 17, 2005, 4 pages

## Attachment 1

### ERA Requirements Document Sections 19 and 20 Related to Search and Access Capabilities

#### **ERA19 The system shall provide the capability to search the assets it contains**

ERA19.1 The system shall provide the capability for the user to select the characteristics of a search against the assets it contains

ERA19.1.1 The system shall provide the capability for the user to enter the criteria for the search

ERA19.1.2 The system shall provide the capability to search by geographic information

ERA19.1.3 The system shall provide the capability to search by subject

ERA19.1.4 The system shall provide the capability to search by time period

ERA19.1.5 The system shall provide the capability to search by accession number

ERA19.1.6 The system shall provide the capability to search by transferring entity

ERA19.1.7 The system shall provide the capability to search by government function

ERA19.1.8 The system shall provide the capability to search by government line of business

ERA19.1.9 The system shall provide the capability to search by asset type

ERA19.1.10 The system shall provide the capability to search by geospatial identifiers

ERA19.1.11 The system shall provide the capability to search by any element defined in the asset's template

ERA19.1.12 The system shall provide the capability to search by media type

ERA19.1.13 The system shall provide the capability to search by record type

ERA19.1.14 The system shall provide the capability to search descriptions by description unique identifier

ERA19.1.15 The system shall provide the capability to search by title

ERA19.1.16 The system shall provide the capability for keyword searching

ERA19.1.17 The system shall provide the capability for exact phrase searching

ERA19.1.18 The system shall provide the capability for concept-based searching

ERA19.1.19 The system shall provide the capability for Boolean searching

ERA19.1.20 The system shall provide the capability for proximity searching

ERA19.1.21 The system shall provide the capability to search based on the frequency of access to assets by other researchers doing similar searches

ERA19.1.22 The system shall provide the capability for automated question-and-answer searching based on searches performed frequently

ERA19.1.23 The system shall provide the capability to search only for descriptions that refer to electronic records

ERA19.1.24 The system shall provide the capability to use wildcard characters in searches

ERA19.1.25 The system shall provide the capability for searching records lifecycle data

ERA19.1.26 The system shall provide the capability for searching authority sources for people

ERA19.1.27 The system shall provide the capability for searching authority sources for organizations

ERA19.1.28 The system shall provide the capability for the use of designated "stop words" that are disregarded during searches

ERA19.1.29 The system shall provide the capability to search by transfer

ERA19.2 The system shall provide the capability to search for assets based on their contents

ERA19.3 The system shall provide the capability to search through hierarchies of information

- ERA19.3.1 The system shall provide the capability to navigate from a description to an individual electronic record
- ERA19.3.2 The system shall provide the capability to navigate from an individual electronic record to an associated description
- ERA19.3.3 The system shall provide the capability to navigate from a description to a set of electronic records
- ERA19.3.4 The system shall provide the capability to navigate from a set of electronic records to a description of the set
- ERA19.3.5 The system shall provide the capability to navigate through all levels of records lifecycle data while searching
- ERA19.3.6 The system shall provide the capability to navigate through all levels of sets of records while searching
- ERA19.3.7 The system shall provide the capability to navigate from a description to a description of the description's creator
- ERA19.4 The system shall provide NARA-created default searches
  - ERA19.4.1 The system shall provide the capability for users to select a NARA default search from among available searches
  - ERA19.4.2 The system shall run the user-selected NARA default search
- ERA19.5 The system shall provide the capability for the user to select the search complexity level, from simple single-variable searches to multi-variable complex searches
- ERA19.6 The system shall provide the capability to control search run times
  - ERA19.6.1 The system shall limit search run times in a pre-emptive manner
  - ERA19.6.2 The system shall provide the capability for users to adjust search run time limits
- ERA19.7 The system shall provide information to the user while the search is in progress
  - ERA19.7.1 The system shall provide a search progress indicator
  - ERA19.7.2 The system shall display the search parameters selected by the user
  - ERA19.7.3 The system shall provide an estimate to the user of how long the search will take to execute
  - ERA19.7.4 The system shall notify the user when the search is complete
- ERA19.8 The system shall present the search results set to the user
  - ERA19.8.1 The system shall display a search results set that includes all assets meeting the search criteria
  - ERA19.8.2 The system shall display an explanation for the reason for withholding assets
  - ERA19.8.3 The system shall exclude from display assets whose existence cannot be disclosed to the requesting user
  - ERA19.8.4 The system shall display the total number of results in the result set returned by the search
  - ERA19.8.5 The system shall provide the capability for the user to select the quantity of search results to be presented in the results set
  - ERA19.8.6 The system shall provide the capability for users to select the order in which the result set is presented
  - ERA19.8.7 The system shall provide the capability to rank the results of the search by relevance
  - ERA19.8.8 The system shall present the search results set at user-selectable levels of detail
  - ERA19.8.9 The system shall indicate different versions of a record included in the search result set
- ERA19.9 The system shall provide the capability for a user to refine a search
  - ERA19.9.1 The system shall provide the capability to search within the result set returned by the initial search
    - ERA19.9.1.1 The system shall provide a "more like this" capability to refine a search for more

assets similar to those returned by the search

ERA19.9.1.2 The system shall provide the capability to refine a search using any search criteria available in the system

ERA19.9.2 The system shall provide the capability to stop a search in progress in order to refine the search

ERA19.10 The system shall provide the capability for the user to select the assets they wish to access from among the search results set

ERA19.11 The system shall provide the capability to save a search

ERA19.11.1 The system shall provide the capability for the user to select a saved search from their saved searches

ERA19.11.2 The system shall provide the capability to run saved searches

ERA19.12 The system shall provide the capability for users to store results sets over time

ERA19.12.1 The system shall provide the capability to store search results

ERA19.12.2 The system shall provide the capability to save selected portions of results sets

ERA19.12.3 The system shall maintain a search results set for a specified period of time

ERA19.13 The system shall manage mediated searches

ERA19.13.1 The system shall provide the capability to request a mediated search

ERA19.13.2 The system shall provide the capability for mediated searchers to dialog with search requestors about their mediated search

ERA19.13.3 The system shall provide the capability to manage mediated search request responses

ERA19.13.4 The system shall provide the capability to prioritize mediated searches

ERA20 The system shall provide access to the assets it contains

ERA20.1 The system shall provide the capability to electronically present all electronic record types

ERA20.2 The system shall provide the capability for users to request copies of assets

ERA20.3 The system shall provide the capability to output copies of all assets

ERA20.3.1 The system shall provide the capability to output all assets to media

ERA20.3.1.1 The system shall print address labels for media orders

ERA20.3.1.2 The system shall print packing lists for media orders

ERA20.3.2 The system shall provide the capability to print all printable assets

ERA20.4 The system shall provide access to assets independently of the hardware with which they were created

ERA20.4.1 The system shall provide the capability to output assets independently of the hardware with which they were created

ERA20.4.2 The system shall provide the capability to electronically present assets independently of the hardware with which they were created

ERA20.5 The system shall provide access to assets independently of the software with which they were created

ERA20.5.1 The system shall provide the capability to output assets independently of the software with which they were created

ERA20.5.2 The system shall provide the capability to electronically present assets independently of the software with which they were created

ERA20.6 The system shall provide the capability to access an entire electronic record

ERA20.7 The system shall provide the capability to access a set of electronic records

ERA20.8 The system shall provide the capability to access a portion of an electronic record

ERA20.9 The system shall provide the capability to access all digital components of an electronic record

ERA20.10 The system shall provide the capability to output assets in formats selected by the

user from available choices

ERA20.10.1 The system shall provide the capability for users to select the output format of selected assets from among available formats

ERA20.10.2 The system shall output certified copies of electronic records in formats selectable by the user from available choices

ERA20.10.3 The system shall output certified copies of electronic records on media selectable by the user from available choices

ERA20.10.4 The system shall provide the capability to output selected asset formats via telecommunications

ERA20.11 The system shall maintain the authenticity of an electronic record during access

ERA20.11.1 The system shall maintain electronic record content during access

ERA20.11.2 The system shall maintain electronic record specified behavior during access

ERA20.11.3 The system shall maintain electronic record context during access

ERA20.11.4 The system shall maintain electronic record structure during access

ERA20.11.5 The system shall maintain electronic record presentation during access

ERA20.11.6 The system shall provide the capability to present digital components of electronic records individually

ERA20.11.7 The system shall provide the capability to output digital components of electronic records individually

ERA20.11.8 The system shall provide the capability to present electronic records composed of multiple digital components

ERA20.11.9 The system shall provide the capability to output electronic records composed of multiple digital components