

Screened	By:
Joseph Scanlon	Date:
02-02-2011	DOCID:

January 10, 2001

USAGE OF CIA SCANNED IMAGES AT NARA

The CIA collection of three million scanned images has been available to the public at NARA since October 1, 2000. The scanned images are contained in a stand-alone computer located in the NARA Library. The collection is accompanied by a computer-generated index. The metadata for each document includes the document, folder, box and CIA archives accession number. The documents include about one million photograph caption cards, scientific articles and abstracts, and more than one million pages reflecting systematic declassification review of a wide range of agency records.

The collection has been available on a self-service basis with very little assistance required from the NARA Library staff. Although some individual use sessions have lasted up to several hours NARA has not had to impose any time limitations. On occasion a second user has come to use the terminal when it was already in use but has agreed to conduct other research and return later to use the scanned image collection. Approximately 35 researchers used the collection in the first quarter of Fiscal Year 2001. Virtually all use has been onsite with only one offsite request for staff to search for items for a user. That search was inconclusive.

Most users learned of the collection before coming to the Library. Some saw the press release. Some heard about it from other researchers. Some learned from NARA staff, especially John Taylor, NARA's intelligence subject area specialist, who brought them to the Library.

The 35+ users have printed approximately 7000 pages total for an average of 200 pages per user. Per CIA instruction, NARA has not yet imposed any print page limitations.

The National Security Archive requested a full copy of the database and the index. The Library staff referred them to the CIA.

User Observations:

1. The scanned images are generally of high quality.
2. The OCR quality is low. CIA did not attempt to improve the OCR, hence the quality will be low.
3. Researchers do not understand the construction of the index and are uncertain as to its utility.
4. Researchers can supplement the index with key word searches. Such searches can be inconclusive since they are based on the OCR text. Uncorrected OCR can impact the number of times a search term appears and thus affect the number of "hits" and the resulting relevance ranking.
5. Users are not confident that they have found all relevant documents. Nonetheless, they are pleased to have access to what they do find.
6. Withdrawn and/or redacted items are appropriately marked.
7. The TIFF headers that contain the CIA location number for the image sometimes overlap the scanned image and obscure some of the document. CIA was more concerned with being able to locate the original document in the future than with obscuring a small portion of a scanned document.

8. CIA policy requires redaction of any individual's cursive signature even if their name is printed below that signature to prevent fraudulent use of a signature.

Prepared by Bruce Ambacher and Jeff Hartley, NARA