

**Georgia
Tech**



**Research
Institute**



**Results of Pilot Testing
of FOIA Processing Using PERPOS**

Debbie Carter

Brooke Clement

Bush Presidential Library and Museum

College Station, Texas

and

Sandra Laib

William Underwood

Information Technology and Telecommunications Laboratory

Georgia Tech Research Institute

Atlanta, GA

Technical Report ITTL/CSITD 07-04

June 2007

Not for General Distribution

Distribution authorized for NARA, GTRI and ARL only.

The Army Research Laboratory (ARL) and the National Archives and Records Administration (NARA) sponsor this research under Army Research Office Cooperative Agreement W911NF-06-2-0050. The findings in this paper should not be construed as an official ARL or NARA position unless so indicated by other authorized documentation.

Abstract

The Presidential Electronic Records Pilot System (PERPOS) is a research prototype for investigating advanced technologies supporting archival decisions in processing Presidential e-records. PERPOS supports accession and storage of collections of Presidential e-records. It also supports systematic processing (arrangement, preservation, review and description) and FOIA processing (FOIA search, review, creation of FOIA collections and finding aids) of Presidential e-records. Previously, systematic processing activities were pilot tested by archivists at the Bush Presidential Library. This report records observations made by two archivists who pilot tested FOIA Processing activities as supported by PERPOS. It also reports the dialog between the archivists and the researchers who developed the prototype system. The results of the pilot test are:

- The identification of a number of bugs, which have been fixed;
- A better understanding by two archivists of the functionality of the prototype;
- The identification of additional features that would enhance productivity of archivists in FOIA processing of e-records but that require additional research.
- The inclusion of some of these features in the prototype archival processing system.

This evolving research prototype provides an environment for evaluation of the performance of methods being developed for automatic extraction of information required for withdrawn records, automatic extension of folder titles, automatic description of the contents of records, folders and record series, and support of archivists in making review decisions.

Table of Contents

1. INTRODUCTION	1
2. ARCHIVIST REVIEW OF FOIA PROCESSING TECHNICAL REPORT	2
3. PILOT TEST OF FOIA PROCESSING	9
4. SUMMARY	18
REFERENCES	21

1. Introduction

1.1 Background

The Presidential Electronic Records Pilot System (PERPOS) is a research project sponsored by the ERA Program of the National Archives and Records Administration and lead by the Georgia Tech Research Institute. PERPOS is also the name of an archival repository and processing prototype that supports exploration of issues in storing, arranging, preserving, reviewing and describing electronic records.

PERPOS is being developed using a method known as evolutionary prototyping. An initial prototype is constructed to learn more about the problems of processing Presidential e-records. Once the prototype has been used in processing actual e-records and the requisite knowledge gained, the prototype is adapted to satisfy the now better-understood needs. This process of prototype reuse, learning and re-adaptation repeats until the prototype system satisfies all needs and has thus evolved into a system. Research is ongoing to provide advanced technologies to support archivists in processing electronic records. These new technologies include: automatically recognizing document types, extracting metadata, automatically describing the contents of records, folders and records series, and supporting archivists in making review decisions [Underwood et al 2007].

Five years after the end of the Bush Presidential Administration, archivists at the Bush Presidential Library began to respond to Freedom of Information Act (FOIA) requests, and that continues today. This requires the capability to index and search the repository for Presidential e-records relevant to the request. It also requires the capability to support review of the relevant records including redaction of copies of the records. In addition, finding aids must be created for FOIA collections, and the collections must be made available to the requestor. The PERPOS prototype was extended to include some of these capabilities, and was installed at the Bush Presidential Library so that it could be pilot tested by archivists [Underwood et al 2006].

1.2 Purpose

The purpose of this report is to describe the first use of PERPOS by archivists at the Bush Presidential Library in support of FOIA processing of Presidential e-records. The report also describes what the archivists and the researchers learned and how the prototype is being readapted in light of this new knowledge.

1.3 Scope

In 2006, researchers submitted a technical report describing the prototyping of support of FOIA Processing. The next section summarizes the suggestions of Bush Library archivists on this technical report. The third section describes the pilot test and reports the

archivists' questions and recommendations and the researchers' responses. The fourth section summarizes the results of the pilot test.

2. Archivist Review of FOIA Processing Technical Report

Two archivists at the Bush Presidential Library were assigned to the pilot testing of FOIA Processing using PERPOS. They were asked to review the technical report describing the FOIA Processing functions of PERPOS [Laib and Underwood 2006]. This section summarizes their comments and the responses by the researchers. This dialog is important because it captures the rationale behind changes made to the FOIA Processing features and recommended features that have not yet been incorporated.

2.1 Carter and Clement: Although the terminology "OAID" (the unique number for our textual boxes) was originally suggested, we're now wondering if this term could be altered for the electronic media to container, hard drive, or disk number. This would eliminate any confusion on the part of staff members or researchers, and emphasizes that this material is different media.

Underwood: Good idea. The following footnote has been inserted in the technical report.

"The abbreviation OAID stands for "Over-Sized Attachment ID." This is a 5-digit number assigned by the White House Office of Records Management (WHORM) to the Federal Record Center (FRC) boxes of records transferred to the National Archives. The Presidential Library assigns additional OAID numbers to other collections. Previously, archivists suggested that 5-digit OAID numbers be assigned to the containers of electronic records accessioned by the Library. They now suggest another name for containers of digital materials, for example, Digital Container ID (DCID), to distinguish them from FRC boxes."

While we are doing this, what would you think of automatically generating the next Digital Container ID during Accession, rather than the archivist having to find the last DCID and enter the next sequential DCID? I found that I sometimes entered the wrong container ID. Furthermore, one would not have to reserve a block of OAID numbers for DCIDs, because these would be IDs for a different type of container.

Carter and Clement: Can't we use the container's original number assigned by the White House rather than generating a new number?

Underwood: You will find that there are many different schemes being used by the White House for labeling the media containing digital records and that sometimes these schemes have gaps or duplicate IDs. During Accession, PERPOS captures the identifier from the original medium, and provides the library a unique identifier for controlling its digital containers of e-records.

2.2 Carter and Clement: (Section 5.1) There may be cases where the archivist who created the FOIA case is not the same one who processes or completes the processing of

the FOIA. Is there a mechanism for allowing this, such as an overall system administrator?

Underwood: Currently, the archivist who created the FOIA case or the user with the user name "Administrator" can check the container back in and anyone can change the archivist name associated with the FOIA Case. A footnote was inserted.

"There may be cases where the archivist who created the FOIA case is not the same one who processes or completes the processing of the FOIA. The user with the User name "Administrator, can also check a container back in. Anyone can change the Archivist name associated with the FOIA Case."

Clement and Carter reply: Does this change the name of the first processor (and all of their work) to the second processor's name? We have found, with the textual collections, that it is useful to know who has worked on which FOIA, and many times it is several people.

Underwood reply: Currently, this changes the name of the first processor to the second processor's name. We are adding a feature that only allows the archivist's name associated with a FOIA case to be changed if all the containers checked out for that case have been checked back in. A feature can be added to keep a list of prior archivists who have worked on the case. Currently, archivist's names (user names) are associated only with withdrawn records. Is it desirable to know who reviewed each record? The archivist's user name could be associated with each section of the manifest that refers to a record. If that is done, one should probably save information as to who re-reviewed a record.

Carter and Clement: Whether archivists' names are associated with the records they have reviewed is probably an administrative decision.

2.3 *Carter and Clement:* (Section 5.1, p. 15) Is there a mechanism for making the read-only copy the active copy if the active copy is inadvertently deleted or corrupted?

Underwood: Yes. Undo checkout (Sect. 5.6), and re-checkout the container (Sect. 5.1).

2.4 *Carter and Clement:* (Section 5.2) In the past, has there been any consideration to making the non-relevant (non-bold) records inactive or not shown at all?

Underwood: No. But it's a good idea. This is one of those many situations in which we could provide an Option under the View or Tools drop-down menu: (1) Show in boldface those file names of records relevant to request and needing review or (2) Show only those files relevant to the request and needing review. Our prototyping philosophy has been to select a reasonable solution and put off optional views for the future. Adding this feature would let you try both options.

Laib: One reason for showing all records and boldfacing only those found in the FOIA search is that some not boldfaced might be relevant to the FOIA case. In some cases not all relevant records are found in the FOIA search. There may be some records that have not been indexed due to their format (images) that turn out to be relevant. That is why the ability to add documents to FOIA Cases was included. A second reason for showing all records is that most of the files in a series or folder might be relevant and thus boldfaced, and there are just a few additional records. The archivist might review them to complete the review of the series or folder. This would be a case of records that were not relevant, but systematically processed.

2.5 *Carter and Clement:* (Section 5.3) It might be useful to add a review action for situations when the archivist needs further guidance before making a review decision; something like "Pending" or "To Be Determined" or "Needs Decision" that will highlight that the document is outstanding.

Underwood: Your comment is added as a footnote.

Laib: During review of a FOIA case, the fact that a document that has not been reviewed is bolded but the icon beside it is white instead of blue, red, green, or yellow should make it stand out as needing further action. Maybe bold directories should only get a grayed out check mark during FOIA review if some of the bold or relevant documents have not had a review action performed on them. That way the directories would stand out as having a document that still needed a decision.

2.6 *Carter and Clement:* (Section 5.3) Will (a)(3) and/or (b)(3) closures allow the selection or entry of the specific statute as required by FOIA?

Underwood: "Edit Reasons Withdrawn" discussed in the Review Section of the Reference Manual for PERPOS describes how one can enter subcategories of FOIA exemption (b)(7) "Law Enforcement Investigations" so that they can be selected as a Reason Withdrawn during review of a record. Similarly, one could enter statutes that were reasons for withdrawal under (a)(3) and/or (b)(3). They would actually only have to be entered under (b)(3) since they are the same for (a)(3). A preferable solution may be to allow reasons withdrawn to be edited during review when the (a)(3) and/or (b)(3) had been entered. This involves highlighting the filename, selecting View>Properties and editing reasons withdrawn. Currently, it is not editable. The following footnotes have been added.

"FOIA exemption (b)(7), Law Enforcement Investigations, allows selection of subcategories of that exemption."

"Reasons for Withdrawal (a)(3) and/or (b)(3), Exempted by Statute, should enable selection or entry of the relevant statute(s). One option would be to have a list of statutes for which there are exemptions to release and when (a)(3) and/or (b)(3) was selected, the archivist was required to pick one or more statutes from that list. The list could be extended by the capability to Edit Reasons Withdrawn."

2.7 *Underwood*: Can you give us any examples of any donor restrictions on access, e.g., codes and criteria? They don't actually need to be Bush Library restrictions.

Carter and Clement: Although there are several reasons for materials to be closed under a donor's deed of gift [personal privacy, national security, specific request of donor, etc.], the Bush Library does not specify the reason a document or a portion of a document is closed under the deed of gift. The one exception is for donated materials that have classification markings similar to federal records ("C" for confidential, "S" for secret, and "TS" for top secret). Deed of gift restrictions are cited on Bush Library withdrawal sheets as "C" with the explanation: "Closed in accordance with restrictions contained in donor's deed of gift."

Other libraries may have different practices, for example the Ford Library's withdrawal sheets have three restriction codes for donated materials:

- A. Closed by Executive Order 12356 governing access to national security information. [Note that EO 12356 has now been replaced by EO 12958]
- B. Closed by statute or by the agency which originated the document.
- C. Closed in accordance with restrictions contained in the donor's deed of gift.

2.8 *Underwood*: Could you suggest some shorter descriptions of the subcategories of FOIA (b)(7) than those shown below? They are needed for the dialog box for selecting restrictions on disclosure.

(b)(7)(a) Law enforcement investigations: Could interfere with pending or prospective enforcement proceedings

(b)(7)(b) Law enforcement investigations: Would deprive a person of a right to a fair trial or an impartial adjudication

(b)(7)(c) Law enforcement investigations: Could constitute an unwarranted invasion of personal privacy

(b)(7)(d) Law enforcement investigations: Could disclose the identify of a confidential source

(b)(7)(e) Law enforcement investigations: Would disclose techniques, procedures, or guidelines for law enforcement investigations or prosecutions

(b)(7)(f) Law enforcement investigations: Could endanger the life or physical safety of any individual

Carter and Clement: Suggestions for shorter descriptions:

(b)(7)(a) Could interfere with law enforcement proceedings

- (b)(7)(b) Could affect right to fair trial or impartial adjudication
- (b)(7)(c) Unwarranted invasion of personal privacy
- (b)(7)(d) Identity of or information provided by a confidential source
- (b)(7)(e) Law enforcement or prosecution techniques, procedures, or guidelines
- (b)(7)(f) Could endanger the life or physical safety of individuals

2.9 *Underwood*: For inclusion in the dialog box for selecting exemptions from (restrictions on) disclosure under (a)(3) and/or (b)(3), we need a few examples of statutes that have restrictions on disclosure.

Carter and Clement: When the Bush Library has cited (b)(3) in withdrawn documents, these are the statutes that have been applied the most often, in order of prevalence:

50 USC 403g – CIA Act

26 USC 6103 – Internal Revenue Code

Rule 6(e)(6) Federal Rules of Criminal Procedure

42 USC 2161-66 Atomic Energy Act of 1954

22, USC 2751 Arms Export Control Act¹

Of course, it is always possible that additional statutes may be applied in the future.

2.10 *Carter and Clement*: (Section 5.3) The “Reasons Withdrawn” screen does not have the minimize function that the “Withdrawal Info Screen” has; it might be useful, as the archivist may need to look at the document in order to determine all of the reasons withdrawn.

Underwood: Footnote added:

"It might be useful for the Reasons Withdrawn window to have a minimize function, as the archivist may need to look at the record in order to determine all of the reasons withdrawn."

Laib: The Reasons Withdrawn dialog box does not have a minimize function. However, it does have a “View {filename}” menu item at the top that allows the archivist to see the record.

2.11 *Carter and Clement*: (Section 5.3, p. 21) Archivists need the ability to update and/or reverse a previous review decision. Examples include when the document has been declassified, opened on appeal, or at the end of the 12-year PRA restriction period (when only FOIA exemptions stand).

¹ <http://www.usda.gov/da/ocpm/Security%20Guide/S2unclas/Export.htm>

Underwood: Paragraph added.

"An archivist may update and/or reverse a previous decision. Examples include when the record has been declassified, opened on appeal, and at the end of the 12-year PRA restriction period (when only FOIA exemptions stand). Support for these capabilities is discussed in the PERPOS Reference Manual. [Underwood et al 2006, Sect 2.7.9]."

The capabilities to open for public access a previously withdrawn record and to redact a previously withdrawn record need to be implemented.

2.12 *Carter and Clement:* (Section 7.0) Do you want the most up-to-date FOIA Finding Aid form? We based it on the new professional standard [Describing Archives: A Content Standard (DACS)] adopted in 2004. The Finding Aid should specify that these are electronic records, not paper records.

Underwood: We would like to have a copy of the most up-to-date FOIA Finding Aid form. We have ordered a copy of DACS. The following footnote is inserted:

"The Finding Aid should specify that these are electronic records, not paper records."

Carter: You will find attached, as requested, up-to-date versions of our Library Reference Request Form, Library Reference Search Form, and the FOIA Finding Aid.

2.13 *Carter and Clement:* (Section 7.3) Consider adding the following to the Withdrawal Sheet: document date, creator name and office, document subject, document type.

Underwood: On page 19, a withdrawal information dialog box is shown in which this information is entered. It is saved in the manifest and displayed in the right windowpane corresponding to the filename of a record (see page 20). At the top of page 30 one can see that this withdrawal information is shown in the right windowpane and is associated with the filename of the withdrawal sheet. It should have shown up in the withdrawal sheet itself, as seen on the lower part of page 30. The following footnote has been added.

"The withdrawal information shown in the right windowpane (Document Type, Subject or Title, Chronological Date, Creator Name and Office) should have also been shown on the withdrawal sheet."

2.14 *Carter and Clement:* (Section 7.4) Concerning the issue of incidentally processed records: one option for completing a container or processing a record that has been inadvertently reviewed [i.e., after reviewing the document the archivist determines the record is not-relevant to the FOIA] is for the archivist to handle the document under systematic processing. At this point, the Bush Library does not incidentally process FOIA records.

Underwood: While reviewing records in a container (or directory), an archivist may find there are additional records not in the result set that if reviewed would complete the review of the container (or directory). If reviewed at that time, these additional records are said to be incidentally processed. It was stated in an earlier version of the Archival Processing Manual, that folders of incidentally processed records should be indicated in the finding aid. Since archivists do not incidentally process records during FOIA review, the FOIA Reference Container and the Finding Aid will not indicate such records.

Two related questions are:

- (1) What should be done with a record that is in the result set of the query, reviewed, and then it is concluded that it is not actually relevant to the request, and
- (2) What should be done when a record is discovered that is not in the results set, but is relevant to the FOIA query?

To address these questions, two additional actions have been added to the *Action* drop down menu in the *Review Activity*. These actions are *Remove from Case* and *Add to Case*. The first action, *Remove from Case*, is required because the results set of a text-based query is often imprecise. It may contain records that are not relevant to the FOIA request. For example, if an archivist's query is on the term "Space," thinking that the result set will include items pertaining to NASA or the Space Council, they will inevitably find some documents that do not pertain to those organizations at all. Rather, they may get documents referring to office space.

The second action, *Add to Case*, is required because a text-based query does not recall in the results set all the records relevant to a FOIA request. It is also possible that some records were not returned in the result set of a FOIA search because the records were not indexed. This situation can occur when files have not been extracted from an archive file, a file is password protected or the record is an image or audio file.

2.15 *Carter and Clement:* (Section 7.4) At the present time, for a FOIA request, the Bush Library assigns separate FOIA case numbers for Bush Presidential, Bush Vice Presidential, and Quayle Vice Presidential records, for example 2007-0020-F for Bush Presidential records, 2007-0021-F for Bush Vice Presidential records, etc. This is unique to the Bush Library, as the Reagan and Clinton Libraries do not have Vice Presidential records.

Underwood: This may imply that an option is needed for indexing and searching the Bush Presidential e-record collection separate from the Bush Vice-Presidential and Quayle Vice-Presidential e-record collections. This is easily accomplished because e-records, including e-mail, are associated with the appropriate collection at the time they are accessioned. Alternatively, a separate FOIA case number could be generated for electronic records.

2.16 *Carter and Clement*: (Section 8) The Bush Library now files FOIAs into the shadow file soon after they are completed & opened. Could you give us more information on “obsolete FOIAs?”

Underwood: We have changed the term "obsolete" to "inactive." Our basic idea is to include in a FOIA collection all copies of records that are relevant to the FOIA request, even if they were reviewed in another FOIA case and are included in the FOIA collection for that case. This facilitates access to the records relevant to a FOIA case. For a similar reason, copies of records that are relevant to a FOIA case, but were systematically reviewed could also be included in the FOIA collection. When there was no longer a need to keep the FOIA collection together, it could be deleted. The master copy of all records, closed, opened, or redacted is in archival holdings.

Carter and Clement Reply: FOIAs are always “active” – regardless if they’ve been archivally arranged. We recently received an administrative FOIA, asking for information on FOIAs requested during a certain period of time. Knowing this, we’ll need to have the ability to go back, even years later, and have access to this information.

Underwood Reply: There is no need to delete a FOIA e-record collection. A copy of it can be kept indefinitely.

2.17 *Carter and Clement*: (Section 9.1) Could you provide us more information on how FOIA/Systematic case numbers are generated? Presently, our case numbers are generated by the Bush Library’s Access database.

Underwood: The FOIA/Systematic case numbers are generated by the Bush Library's Access database. They are transcribed into the dialog boxes in *Add FOIA Case* and *Add Systematic Case*.

3. Pilot Test of FOIA Processing

Two authors of this report, archivists at the Bush Presidential Library, pilot tested the FOIA Processing functions of PERPOS the week of February 9, 2006, and then again on February 12, 14, 23, and 27.

The archivists searched ten filtered containers that were available on the local PERPOS machine. The following Boolean searches were conducted:

- 2007-0001-F “China AND MFN”
- 2007-0002-F “Gerald AND Ford”
- 2007-0003-F “John Tower”
- 2007-0004-F “(home AND \$school)”
- 2007-0005-F “state dinner”
- 2007-0006-F “(\$point NEAR light)”
- 2007-0007-F “NAFTA”

2007-0008-F “(state AND \$dinner)”
2007-0009-F “(Camp David) OR (Camp Marvin)”

All searches resulted in at least one responsive document.

Archivists then checked out the appropriate containers for the first FOIA case and opened the APT to review the highlighted (responsive) documents. Throughout these FOIA cases, archivists tested the functionality of the various restrictions, including Presidential Record Act (PRA) restrictions and FOIA exemptions. Archivists also applied the redaction and stamping tools to several records. When archivists completed the review of all responsive documents in a FOIA case, they checked in the containers using the ART, and proceeded to generate the reference copy and the FOIA Finding Aid.

The following are comments, questions, and suggestions from the archivist's tests. They are followed by the Georgia Tech researcher's replies.

3.1 *Carter and Clement*: Using the Archival Processing Tool (APT) the tester was able to checkout [using “Undo Checkout”] a container checked out to another tester; is this a capability that the system should have or should only the original user or administrator have this power? We realize now that you probably gave both of us administrator privileges and this may account for the tester’s ability to do this.

Underwood: I think what is meant is that the first user checks out a container associated with a FOIA case. A second user undoes the checkout. The second user can then checkout the container, from a different FOIA case.

Laib: The Archival Repository Tool does not check privileges. It checks for the current USERNAME, which is a system variable that gets set when a user logs on in Windows 2000.

The problem occurs because the current USERNAME is being compared to the user who owns the case instead of comparing the current USERNAME to the archivist that checked out the container. The problem occurs when two different archivists have the same container in their Systematic or FOIA Case. One archivist checks out the container under their case, then the other archivist tries to Undo Checkout of the same container but under his assigned case. The second archivist who had not checked out the container should not be able to undo checkout. This problem has been fixed and when one user has a container checked out, another user will not be able to undo checkout of that container.

3.2 *Carter and Clement*: In the Image Redactor, spreadsheets show up as letter-sized portrait layouts. Can the layout be changed so that the redactor shows up as a legal-sized landscape?

Underwood: This may require manual intervention by the review archivist for each spreadsheet or database table. In Quick View Plus Help, under View> Quick View Plus Options, the Print Tab, one can control the printed characteristics of spreadsheets,

databases, bitmaps and drawings. Furthermore in File > Page Setup you can choose legal size and portrait. You can also see the result in File > Print Preview. The question to Sandy is: Do we have enough control of Quick View Plus that the customizations selected for the current spreadsheet or database are the ones that will be used when the APT calls Quick View Plus to print to a TIFF image(s)?

Underwood: Research question—Is it possible to represent the knowledge that review archivists have to decide on the most effective, intuitive way to layout the printing of a particular spreadsheet or database table, and adjust the printing characteristics of a viewer to achieve those characteristics? If so, we would have a very useful productivity tool for review.

Underwood: Second research question—Spreadsheets and database in their native file format are not viewed on a computer display as pages, but as virtual spreadsheets and databases that can be viewed using the horizontal and vertical scroll bars. By scaling spreadsheets before printing to a TIFF image, can't intermediate size spreadsheets and databases be put into a single high-resolution image whose details can be viewed by zooming in?

Laib: We have no control over Quick View Plus other than telling it to print. It will follow whatever has been set up in the View->Options->Print Tab as far as printed characteristics of spreadsheets, databases and bitmaps and drawings. What gets set from File->Print Setup are printer options which are not tied to Quick View Plus but rather controlled by the printer object. I tried to change both the orientation to landscape and the paper size to legal in the properties of the printer object from inside the Archival Processing Tool, but it did not work. I tried to change the properties from the File->Print Setup inside of Quick View Plus. Again, the image that was created was letter size and portrait style. The problem appears to be with the Informatik Print Driver. Even opening a spreadsheet directly in Quick View Plus, when the Informatik driver is selected as the printer and Landscape is selected as the Orientation, the TIFF image is still created in Portrait orientation.

3.3 *Carter and Clement:* In the Image Redactor, for documents without margins, will there be room for the Redaction Stamps to be added to the documents without overlaying text, such as spreadsheets, for example?

Underwood: Quick View Plus by default creates 1-inch margins. In View > Quick View Plus Options, the Print Tab, these defaults can be changed. Do we have enough control of Quick View Plus that the customizations selected for the current spreadsheet or database are the ones that will be used when the APT calls Quick View Plus to print to a TIFF image(s)?

Laib: We have no control over Quick View Plus other than telling it to print. It will follow whatever has been set up in the View->Options->Print Tab as far as printed characteristics of spreadsheets, databases and bitmaps and drawings. Whatever margins

are set, will be used the next time the Archival Processing Tool creates a TIFF image from a document.

3.4 *Carter and Clement*: Within the same document, can the Image Redactor default to the same closures throughout the document? For example, separated columns of social security numbers?

Underwood: If I understand you correctly, if there are two or more sentences or paragraphs on the same page or different pages that have the same restrictions or exemptions, and that are separated by one or more sentences or paragraphs or pages that have no restrictions, is it possible to select the restrictions once and apply it to the separate redacted areas? Similarly, if there is a spreadsheet or database table with a column of SSNs, and this table crosses several pages, is it possible to select the restrictions once and apply it to the redacted areas across several pages? The answer is *Yes*.

Select the Block rectangle. Mark each sentence, paragraph, or column on a page (or multiple pages) in the document that has the same restriction(s). Select the stamp and enter the restrictions, select OK, and go through the pages marking each blocked rectangle with the restriction.

3.5 *Carter and Clement*: Using the Image Redactor, testers attempted to undo redactions. The only option is to delete the entire redacted version of the document. Archivists need the capability to change portions of redactions whenever additional information can be released. For example, a document may contain personal information on four job candidates; all four sections would be redacted for privacy. At the death of the first candidate, the relevant section could be unredacted and released to the public. With paper records, archivists would make the changes on the one page rather than having to re-redact the entire document.

Underwood: We have not implemented re-redacting redacted records [see section 2.7.9.3 of Help or in the Reference Manual "Re-redacting a Previously Redacted Record" (page 159)]. However, your suggestion is excellent. We hadn't thought of doing this until you suggested it. See Sandy Laib's comment below.

Laib: The system is working the way it was designed. The block and the stamp that is used to implement redaction, during image redaction, are in fact annotations made to the TIFF image. Currently, when the user saves the modifications made to the TIFF image the annotations are burned into the image. In the future it should be possible to put off burning the annotation into the image until the reference copy has been made. The reference copy would have the annotations burned in. The redacted image that is stored in the master container would have the annotations, but they would not be burned in. Therefore, the annotation could be removed or modified during re-redaction. Each time a reference container was created, any redacted image documents that were copied to a reference container would have their modified annotations burned into the reference copy.

3.6 *Carter and Clement*: A document marked “remove from case” in Archival Processing Tool (APT) still showed up in the ART list of documents relevant to the FOIA Case.

Laib: This was indeed a bug. At one point there was a loop made through manifest items to update FOIA Case documents. But, this loop was replaced by a set of queries. One of the queries is missing. The query that adds documents to the CaseDocument table and the query to add documents to the ContainerDocument table are there, but the query that removes documents from the CaseDocument table based on the manifest is missing. This problem has been fixed and when a document has been marked "remove from case," it will no longer show up in the list of documents relevant to a case.

3.7 *Carter and Clement*: In the Quick View Plus Viewer, we suggest adding a pop-up box with “Find” and “Find Next” for use in searching for relevant text within a document.

Underwood: In Quick View Plus, on the Edit drop-down menu, there are *Find*, *Find Previous*, and *Find Next* functions.

3.8 *Carter*: We also wondered if FOIA finding aids could be adjusted so that the container number would be associated with the staff member (if known) rather than at the document level. In the textual collections, there are multiple OAID numbers associated with an individual staff member’s records. However, we assume that only one hard drive is associated to a staff member. Is this correct? Is this an easy fix?

Underwood: There are sometimes multiple hard drives associated with a staff member's name, but not often, and usually no more than two. Hence, if there was only one container that contained relevant records, the container ID could be associated with the staff member's name. However, recall that the finding aids for paper records usually show the OAID number at the folder level, not at the series or the record level. PERPOS is showing the path (including directories) to each file (record).

President, Office of the
George H. W. Bush's Files
EO\121.doc [OAID 15004]
MEMO\013.wds [OAID 15004]

We need to give this more thought.

3.9 *Carter*: The closed document “WORK\HOLLY\JOURNAL” (10008) did not show up on a FOIA finding aid except for the following: “[OAID 10008].” When I checked the FOIA Reference Copy I noticed that the withdrawal sheet was attached at the container level NOT at the document level as were other withdrawal sheets. I don’t know if this is the reason for the problem with the FOIA finding aid.

Laib: I would like to see a copy of the finding aid and a copy of the Reference copies

manifest. A copy of the manifest of OAID 10008 copied while it is checked out under the FOIA Case to the owner of the FOIA Case would also be a big help. Seeing the manifest after it is properly checked out allows me to see how the documents should have been marked when the APT opened the container. Documents that belong to a FOIA Case get an extra property that is used by both the APT and ART when the container is checked back in.

1. A way to get a copy of the manifest of OAID 10008 is to open the container in the APT. Once the container has been opened select “Manifest” from the “View” main menu. This will open the manifest in Quick View Plus. Select “Save Copy...” from the “File” main menu of Quick View Plus. A dialog will appear that will allow you to choose where you want to save it and what file name to give it. If you select the down arrow to the right of the Save in: field you can go to the My Documents directory and save the file in a folder named Sandy. The default filename will be MANIFEST. I suggest you suffix that name with the OAID number of the container (10008 in this case). Then Exit Quick View Plus. I cannot emphasize how important it is that I see the manifest. The manifest is how information is passed between the ART and the APT.
2. A way to get a copy of the finding aid for the Reference copy is to select the FOIA case in ART. Then select Open Finding Aid from the File drop-down menu. The finding aid will be displayed in Internet explorer. Select Save as from the File drop-down menu of Internet Explorer and save the Finding aid in the same folder for Sandy. It can have the file name displayed in the address field of Internet Explorer, which will be the FOIA case number followed by htm. Exit Internet Explorer.
3. A way to get a copy of the Manifest of the Reference Container for the FOIA case is to select the FOIA Case in ART. Then select Open FOIA Ref Copy from the File drop-down menu. This will open the FOIA Reference Container inside the Archival Processing Tool. Then select Manifest from the View main menu. This will open the manifest in Quick View Plus. Select Save as and proceed as in item 1 above except the filename should be MANIFEST suffixed with the FOIA Case number.

Clement: Here is the manifest, finding aid, and reference copy for container 10008 under FOIA 2007-0003-F [the issue was the file WORK\HOLLY\JOURNAL] that Sandy requested.

Laib: Thank you for the copies of the manifests and finding aid. There was a problem in building the new path for closed documents in the Reference copy and finding aid. This is being fixed.

3.10 *Carter:* I conducted a FOIA search on NAFTA and only 1 document in 10005 was listed in the ART (“WORK\WP50\CHEAT.STA”). However, in the APT two documents were highlighted for 10005 and both were relevant: “WORK\WP50\CHEAT.STA” and

“WORK\WP50\CHEAT.” I marked both as open, however only “WORK\WP50\CHEAT.STA” shows up in the FOIA Reference Copy and FOIA finding aid. Could this have something to do with the documents having almost identical names?

Laib: This time a copy of only the manifest of OAID 10005 copied while it is checked out under the FOIA Case to the owner of the FOIA case might give me an idea of what was going on. Follow the instructions for item 1 in my response to the previous question.

Clement: Here is the manifest for container 10005 in Debbie's NAFTA FOIA that Sandy requested.

Laib: Thank you for the manifest for container 10005. There was problem in boldfacing relevant documents in a container. This problem has been fixed.

3.11 *Clement:* In the APT, is it possible to change the format of the drop-down menu when you open a container? It would be nice to be able to see the actions taken (the color-coded actions) on all of the documents first thing, rather than having to either click on the document to read the action, or switching to review for that information to become readily available

Underwood: An option could be added that determined the default activity when opening the APT. It currently defaults to Explore, but it could be set to default to any of the activities including Review, especially for those only performing FOIA Processing. Then one would see the color coded actions immediately. This capability will be implemented.

3.12 *Clement:* I ran a search on “Points of Light” (2007-0006-F) and found in container 10003 press releases for each Point of Light. The thing is, only the first press release was in bold, and there are over 1000 press releases in the directory, which means that we will have to go and add each of the press releases to the case during processing. Is there some way we can capture several documents and add them to the FOIA case, rather than having to do each one separately?

I re-ran the search as (\$point NEAR light), and this problem was solved, but it may be useful in the future to have the ability to highlight several documents and add them all at the same time.

Laib: Currently, the proper thing to do is to return to the FOIA case and re-perform the FOIA search in light of the new information you obtained. The new search will replace the previous search. However, doing this brings up the issue that if records have already been reviewed for that FOIA case one would like to see that they were included in the new results set. In this case, your query captured all the records in the original results set, so any records that you had reviewed would remain in this FOIA case. Generally speaking, if records had been reviewed, the new FOIA search should include the original query “OR”ed with the new query, for example, "Points of Light" OR (\$point NEAR light).

Underwood: Alternatively, when a new FOIA search is performed for a FOIA case in which a search has already been performed, there could be an automatic check to determine whether records have already been reviewed for that FOIA case. If so, they could automatically be included in the results set.

Laib: With regard to the suggestion that in review it might be useful to highlight several documents and add them to the FOIA case all at the same time, we could add the capability to add several documents to a case simultaneously, but the programming might be quite complicated. It might be possible to add to the review activity the capability to add an entire directory to a case, rather than to highlight several documents to add to the case all at the same time. However, that would also be quite complicated to accomplish.

Underwood: Another alternative would be to add a data element to items and folders called item description and folder (or file unit) description, respectively, and to provide the capability for archivists to describe items and the folder content. One of our current research tasks is to automatically generate such item and folder descriptions [Underwood et al 2007]. Then, the capability could be provided to index and search on item descriptions or folder descriptions and return relevant items or folders in the results set.

One might be inclined to believe that text-based search for documents relevant to a FOIA request would always be better than search of item or folder descriptions. However, experiments with text-based search of large document collections have shown that precision and recall measures of relevant records are usually less than .50. The capability to search item and folder descriptions as an alternative to text-based search would allow us to determine via an experiment the relative merits of text-based search and search of item and folder descriptions.

Carter and Clement: We think the last option sounds like a great idea!

3.13 *Clement:* When working in container 10005 for a FOIA on “State Dinner” (2007-0005-F), one document called “STATE” was in bold. When I clicked once on the document, the information on the right said that the document had 316 bytes of information, but when I tried to open the document, I got a blank screen. I then tried to remove the document from the case, but the menu options does not offer me that option, but instead offers to add the document to the FOIA case. We believe this may be connected to the same issue of this being a document without an extension when a document with an extension follows.

Underwood: The reason that the document is blank when viewed with Quick View Plus is probably because there is no content. The word STATE probably appears in the header of the file. To check this hypothesis, select the FOIA case in ART and check out the container. In the APT select the review activity and open the container. Click on the file name to view the File with name STATE. Now in Quick View Plus, select View As from the View pull-down menu. You will see two or three option. If there are three, the first will be checked and it is a viewer for the file format type of the file, e.g., WP 5.0 for DOS. The second and third options are Text and Hexadecimal. Select one of those two

and you will see either the text that is in the file or the hexadecimal display of the 317 bytes in the file with the displayable characters shown at the right. See if the word STATE doesn't appear there, but otherwise the file has no displayable characters. If this is the case, we may have learned that some files that we realize as formats of word processing documents or spreadsheets or databases that have no content may be indexed and found in a search because of words that appear in the header.

Laib: With regard to the second issue, the filename extension should not have any effect. I need to see the manifest for the container. Again it needs to be obtained after the container is checked out under the FOIA Case to the owner of the FOIA Case. How to obtain a copy of the manifest was described earlier.

Laib: Thank you for the manifest for container 10005. The reason that the file named STATE could not be removed from the case was that it was incorrectly boldfaced and not in the case so could not be removed. The fix to the problem reported in paragraph 3.10 also fixed this problem.

3.14 *Clement:* After processing all of the files in 2007-0005-F, I ran a FOIA Reference Copy which pulled up only 3 of the 6 documents associated with it, and only listed 3 of the 4 documents from container 10005. The two documents from 10008 did not display at all in the reference copy. I determined that the fourth document from 10005 was “WORK\WP50\MARLIN.MEM” which had been marked as PRM during review. Shouldn't this document have shown up, along with a withdrawal sheet? When I ran the finding aid for this FOIA, “MARLIN.MEM” is still missing, but this time the name of the other staff and office file that was part of the FOIA (container 10008) appears – but just the name of the office and staff member. The files “WORK\HOLLY\GERMANY.HS” and “WORK\HOLLY\STDIN327” still do not appear in the finding aid. I believe this may be attached to a similar problem Debbie Carter had with the same container.

Laib: I would like to see a copy of the 10008 container checked out under this FOIA Case, the finding aid for this FOIA case, and the manifest of the Reference Copy for this FOIA case. The instructions for creating these three files are the same as in my comments for the problem Debbie is having with the same container, except use filenames for this FOIA case. We will add withdrawal sheets for PRMs.

Clement: Here are the manifests for containers 10005 and 10008 under FOIA case #2007-0005-F that Sandy requested.

Laib: Thank you for the manifests and finding aid. We now understand that withdrawal sheets for PRMs should be included in the reference copy and will thus show up in the finding aid. This is being fixed.

3.15 *Carter & Clement:* A colleague at the Office of Presidential Libraries involved in ERA brought the following question to our attention: “Does PERPOS have the capability to search for a specific office or staff member rather than by keyword?” We need this

capability because in the past researchers have submitted FOIA requests for specific offices and staff members, for example: the Chief of Staff's Office.

Underwood: This question is related to my earlier comment that it might be desirable to have the capability to search on item and folder descriptions, as this includes the capability to search on document type, correspondents' names and topic. Currently, there is not a capability to include in the results set all the records associated with a staff member or office, but this capability could easily be added.

3.16 *Laib:* There may be a way to get information into the Oracle index for documents that cannot be indexed. A description field (item description) would have to be added to documents (filenames). Then, Oracle would have to return a list of files in each container that it could not index (e.g., images). Using the list of un-indexed files an archivist would have to view the files then fill in the description field. This would be done in the APT. Then Oracle could index both the description field for the documents that did not have textual content as well as the content of the textual documents. FOIA Search would then return more complete results.

Underwood: We might want to run an experiment to determine how many records are indexable and how many are not, and the file formats of those that are not indexable.

4. Summary

The following table summarizes the results of the archivists' review of the FOIA Processing technical report and their testing of PERPOS in support of FOIA processing of Presidential e-records.

	Paragraphs
Questions answered regarding the use of PERPOS for FOIA processing	2.3, 2.5, 2.10, 2.14, 2.16, 2.17 3.3, 3.4, 3.7
Software bugs detected and fixed	3.1, 3.6, 3.9, 3.10, 3.13, 3.14
Recommended new features implemented	2.2, 2.6, 2.7, 2.8, 2.9, 2.13, 3.14
Recommended new features that require research or are not yet implemented	2.1, 2.2, 2.4, 2.11, 2.12, 2.13, 2.15, 2.16 3.2, 3.5, 3.8, 3.11, 3.12, 3.15, 3.16

Many of the questions asked by archivists were prompted by their need to accomplish a task and the Reference manual not adequately explaining how it could be accomplished. This is in large part because this is the first time that FOIA Processing has been tested by archivists.

The half-dozen software bugs that were detected by the archivists occurred because this is the first test of the FOIA Processing support capability by archivists using actual Presidential records. The source of the errors was easily found and the software rapidly corrected.

Recommended features implemented in response to recommendations include:

- Inclusion in *Reasons for Withdrawal of*
 1. subcategories of FOIA exemption (b)(7) Law Enforcement Investigations (2.8)
 2. specific statutes for restriction (a)(3) and/or (b)(3) Exempted by Statute (2.6, 2.9)
 3. Donor restrictions on disclosure (2.7)
- The name of the archivist who is processing a FOIA can be changed only if all the containers associated with a FOIA case have been checked back in. (2.2)
- Inclusion on the Withdrawal Sheet the document date, creator name and office, document subject and document type. (2.13)
- Withdrawal information should be collected for Personal Record Misfiles and included in the FOIA Reference collection and FOIA finding aid. (3.13)

Recommended features that require more research or are not yet implemented include:

- Change the data element and terminology used for containers from OAID (Oversized Attachment ID to DCID (Digital Container ID). (2.1)
- Automatically generate the next Digital Container ID during Accession, rather than requiring the archivist to find the last DCID and enter the next sequential DCID. (2.1)
- In a container's manifest, associate with the filename of a reviewed item, the name of the archivist who reviewed the item. (2.2)
- In the APT under the review activity, add an Options menu item under Tools and provide the option to either boldface the items relevant to a FOIA case, or to show only those items relevant to a FOIA case. (2.4)
- In the Review Activity, implement the capability to open to public access a previously withdrawn record and to redact a previously withdrawn record. (2.11)
- Implement re-redacting a previously redacted record so that one does not have to redact the original, but can simply remove those redactions where restrictions no longer apply. Put off burning the annotation into the TIFF image until the reference copy has been made. The reference copy would have the annotations burned in. The redacted image that is stored in the master container would have the annotations, but they would not be burned in. Therefore, the annotation could be removed or modified during re-redaction. Each time a reference container was created, any redacted image documents that were copied to a reference container would have their modified annotations burned into the reference copy. (2.11) (3.5)

- Create a new version of Bush Library FOIA Finding Aid form. It is based on the Society of American Archivists professional standard [Describing Archives: A Content Standard (DACS)]. The Finding Aid should specify that these are electronic records, not paper records. (2.12)
- Currently, PERPOS only supports indexing and searching the contents of the entire repository. Provide the capability (option?) to Index and Search collections separately. (2.15)
- Include copies of records that are relevant to a FOIA case, but included in another FOIA collection in the FOIA case. This facilitates access by the requestor to those records. For a similar reason, also include in the FOIA collection copies of records that are relevant to the FOIA case, but were systematically reviewed. A question to be resolved is whether to keep the FOIA case marker in the FOIA reference collection and finding aid and whether to put asterisks by record series that have already be opened in systematic processing. (2.16)
- For spreadsheet and database files, experiment with print scaling in Quick View Plus to determine how to best create TIFF files for redaction. (3.2)
- Consider associating the Digital Container ID in a finding aid with directories or record series rather than individual records. (3.8)
- In the APT, in the Tools menu under Options, provide an option to determine the default activity when entering the APT. (3.11)
- Provide an alternative method of FOIA search that allows indexing and search on folder titles. (3.12)
- Provide the capability for archivist to describe the contents of folders and a method of FOIA search that allows indexing and search of folder descriptions. (3.12)
- Provide an alternative method of FOIA search that allows search on Office name or person's name (series title). (3.15)
- Add item description as a data element of record (file) and add the capability perform a FOIA search on item description. (3.16)

In addition to the features identified by archivists who used PERPOS for FOIA processing, the following features described in the Reference Manual and technical report [Laib and Underwood 2006] have not yet been included in the prototype.

- Currently, the FOIA Case Management activity and FOIA Search capability only support saving the result set of one query for a FOIA Case. For complex FOIA

requests or for follow-up searches, it will be necessary to have the capability for additional queries and result sets to be associated with a FOIA case. This can be accomplished this by extending the FOIA Case No. to have subcases, e.g., 2006-0020-F[1], 2006-0020-F[2], and creating FOIA search criteria and results sets for each subcase.

- Archivists need an estimate of the number of pages to be reviewed in a FOIA case as an estimate of workload. While that capability is included in PERPOS, it is not accurate enough. A better method needs to be implemented. An estimate of the number of pages, and the exact number of files and bytes in a FOIA case is also needed for the FOIA Finding Aid. While PERPOS currently accumulates the number of files and bytes in a container, that capability needs to be extended to FOIA collections.
- A Public Access System is needed for viewing FOIA collections and systematically processed opened record series.
- Print the Reference Search Form for inclusion in the FOIA case Folder.
- Integrate REDAX, a PDF redactor, into the Review Activity and capability to convert files in legacy format to PDF.

The next steps are to adapt the prototype to the now better-understood needs of FOIA processing of Presidential records and to provide a new version of the PERPOS prototype to the archivists for reuse and learning of additional FOIA processing requirements.

References

[Laib and Underwood 2006] S. Laib and W. Underwood. FOIA Processing in the Presidential Electronic Records Pilot System. Technical Report 06-05 CSITD/ITTL, GTRI, July 2006.

[Spencer et al 2005] L. Spencer, S. Oriabure and W. Underwood. Launching E-Records with a PERPOS: The Presidential Electronic Records Pilot System. NAGARA Annual Meeting 2005, Richmond, Virginia, July 20-23, 2005.

[Underwood et al 2006] W. Underwood, S. Laib and M. Hayslett-Keck. Reference Manual for PERPOS: An Electronic Records Repository and Archival Processing System, Version 3.1. PERPOS TR ITTL/CSITD 06-2, Information Technology and Telecommunications Laboratory, Georgia Tech Research Institute, September 2006.

[Underwood et al 2007] W. Underwood, S. Isbell, M. Underwood. Grammatical Induction and Recognition of the Documentary Form of Records. *Digital Curation Symposium* (DigCCurr2007), Chapel Hill, NC, April 19-20, 2007.
http://www.ils.unc.edu/digccurr2007/papers/underwood_paper_4-5.pdf