

FEDERAL BUREAU OF INVESTIGATION

Precedence: ROUTINE

Date: 3/30/2005

To: IT Operations
Director's Office

Attn: Ms. Gail Scavongelli
Attn: Mr. Zalmi Azmi, CIO
Mr. [redacted] (OCA)
Mr. [redacted] (OPA)
Mr. [redacted] (OCTO)

CJIS
Training
Administrative Services
Counterintelligence
Criminal Investigative
Laboratory
Inspection
Security
Finance
Counterterrorism

Attn: Mr. [redacted]
Attn: Mr. [redacted]
Attn: Mr. [redacted]
Attn: Ms. [redacted]
Attn: Ms. [redacted]
Attn: Mr. [redacted]
Attn: Mr. [redacted]
Attn: Mr. Richard Haley
Attn: Mr. [redacted]
Mr. [redacted]
Mr. [redacted]
Attn: Mr. [redacted]
Attn: Mr. David Hardy
Ms. Marie Allen
Attn: Mr. [redacted]
Attn: Mr. [redacted]

b6
b7C

Cyber
Records Management

Investigative Technology
Office of Intelligence

From: General Counsel
Administrative Law Unit/ Room 7338
Contact: [redacted] (202) 324-[redacted]

b6
b7C
b2

Approved By: Kelley Patrick W
[redacted]

ALL INFORMATION CONTAINED
HEREIN IS UNCLASSIFIED
DATE 07-11-2007 BY 65179 DMH/BJA/CAL

Drafted By: [redacted] lbh

Case ID #: 66F-HQ-C1321794

1058805

Title: ADMINISTRATIVE LAW UNIT CORRESPONDENCE
PRIVACY IMPACT ASSESSMENT (PIA) UPDATE
INVESTIGATIVE DATA WAREHOUSE (IDW)

Synopsis: [redacted]
[redacted]

b2
b7E

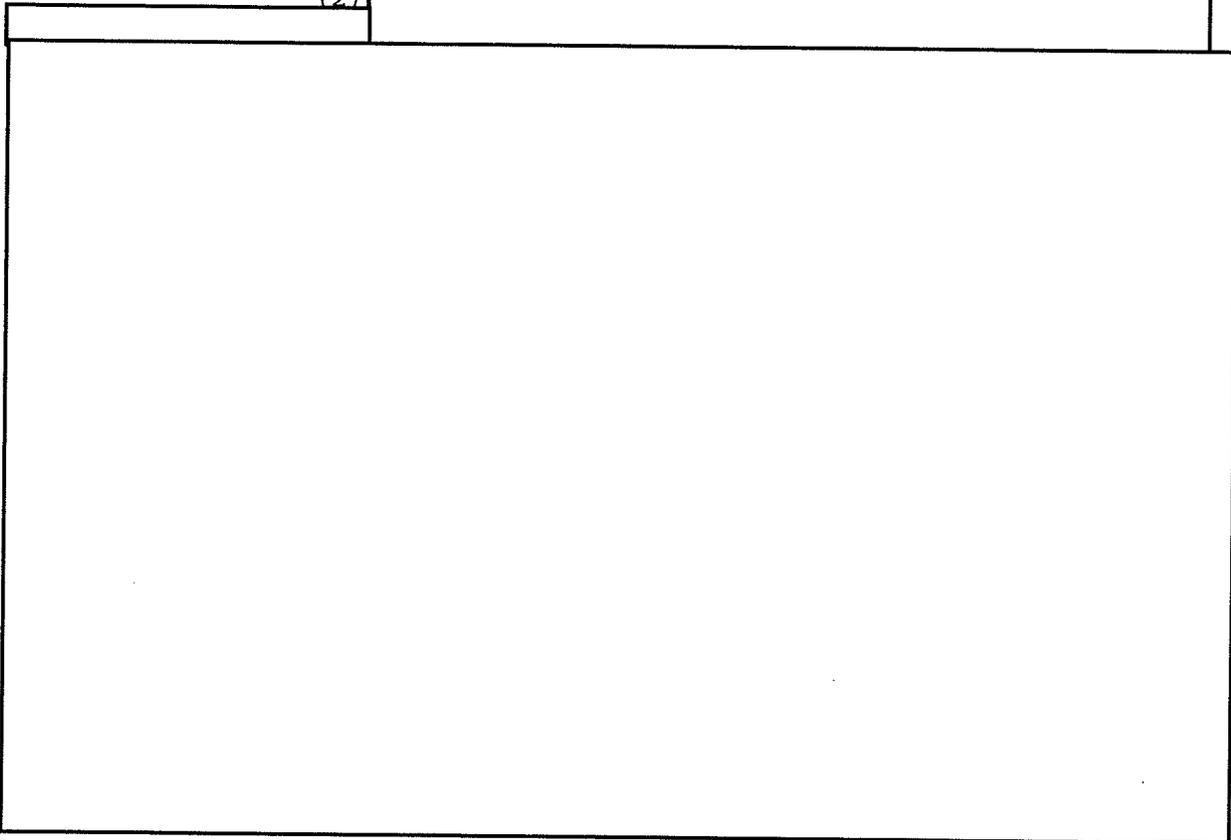
Reference: 66F-HQ-1321794 Serial 166
66F-HQ-1321794 Serial 197

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005

Administrative: This documents contains footnotes. To read the footnotes, download and print the document in WordPerfect. This is a privileged FBI communication; do not circulate outside the FBI without the permission of the Office of the General Counsel.

Enclosure(s): (1) Privacy Impact Assessment, [redacted]

(2) [redacted]



b2
b7E

¹Please see the referenced EC, 66F-HQ-1321794 Serial 166, for more information.

²Please see the referenced EC, 66F-HQ-1321794 Serial 197, for more information.

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005

LEAD(s):

Set Lead 1: (Action)

OCIO, OCTO

AT WASHINGTON, DC

Please take action consistent with this EC.

Set Lead 2: (Info)

ALL RECEIVING OFFICES

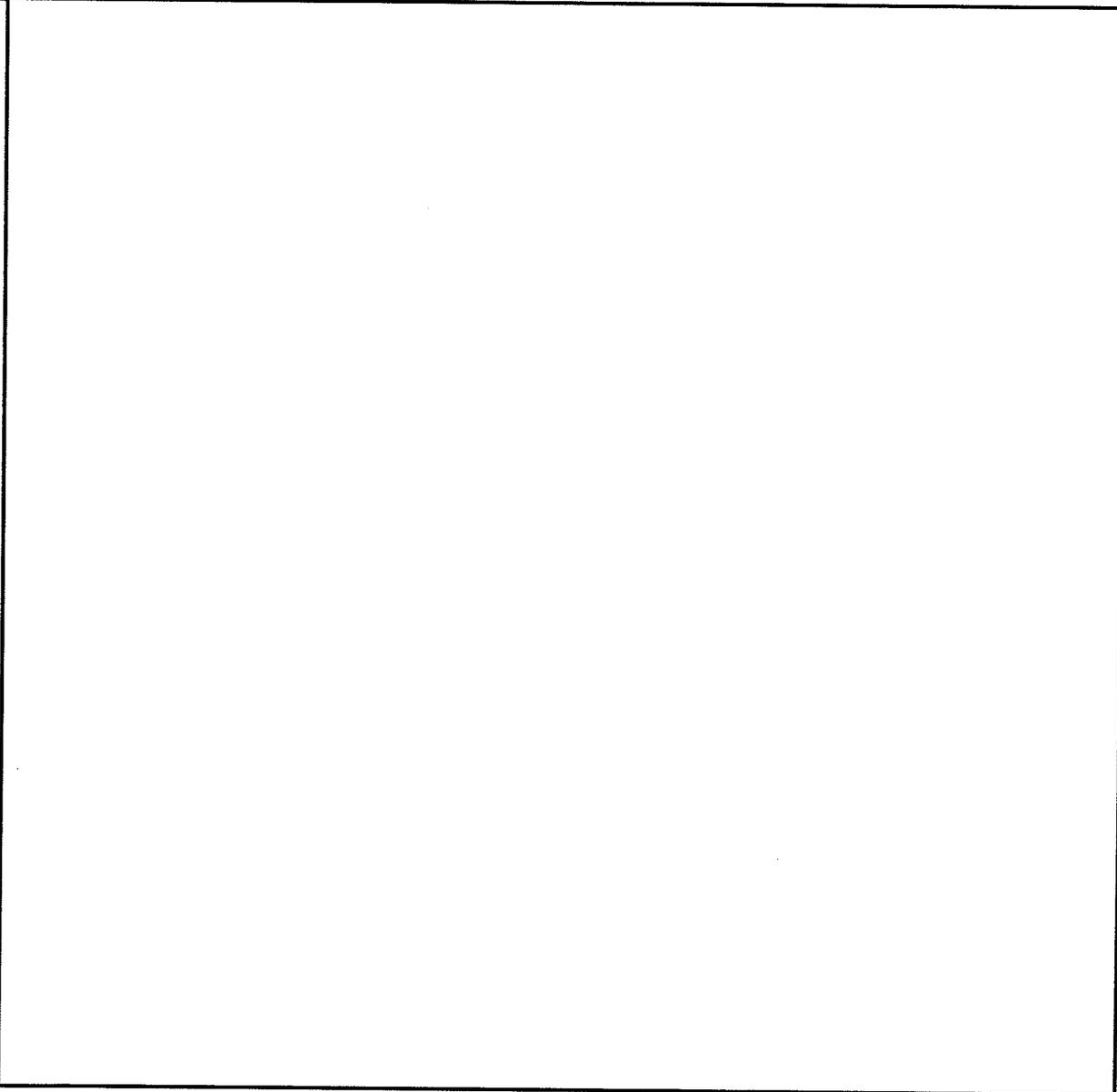
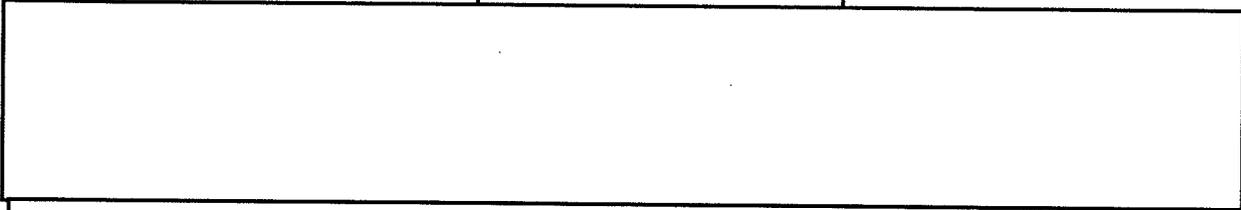
For information only.

cc: 1 - Ms. Caproni
1 - Mr. Kelley
1 - Ms. 
1 - Ms. 
1 - Mr. 
1 - Mr. 
1 - Mr. 
1 - Ms. 
1 - Ms. 
1 - Ms. 
1 - ALU Tickler
1 - ALU Library (InfP/PIA/2005)

b6
b7c

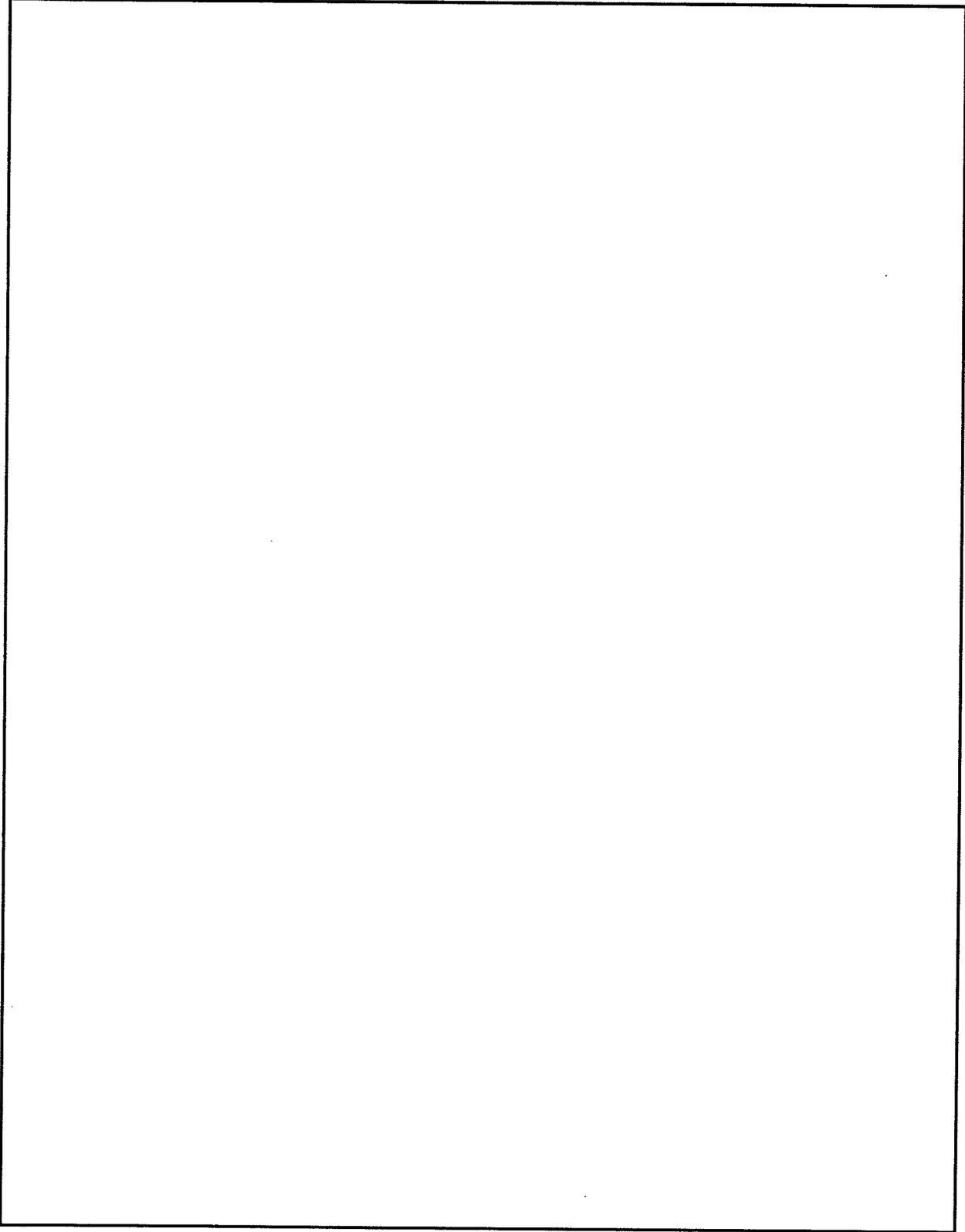
To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005

Enclosure 1
Privacy Impact Assessment



b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005



b2
b6
b7C
b7E

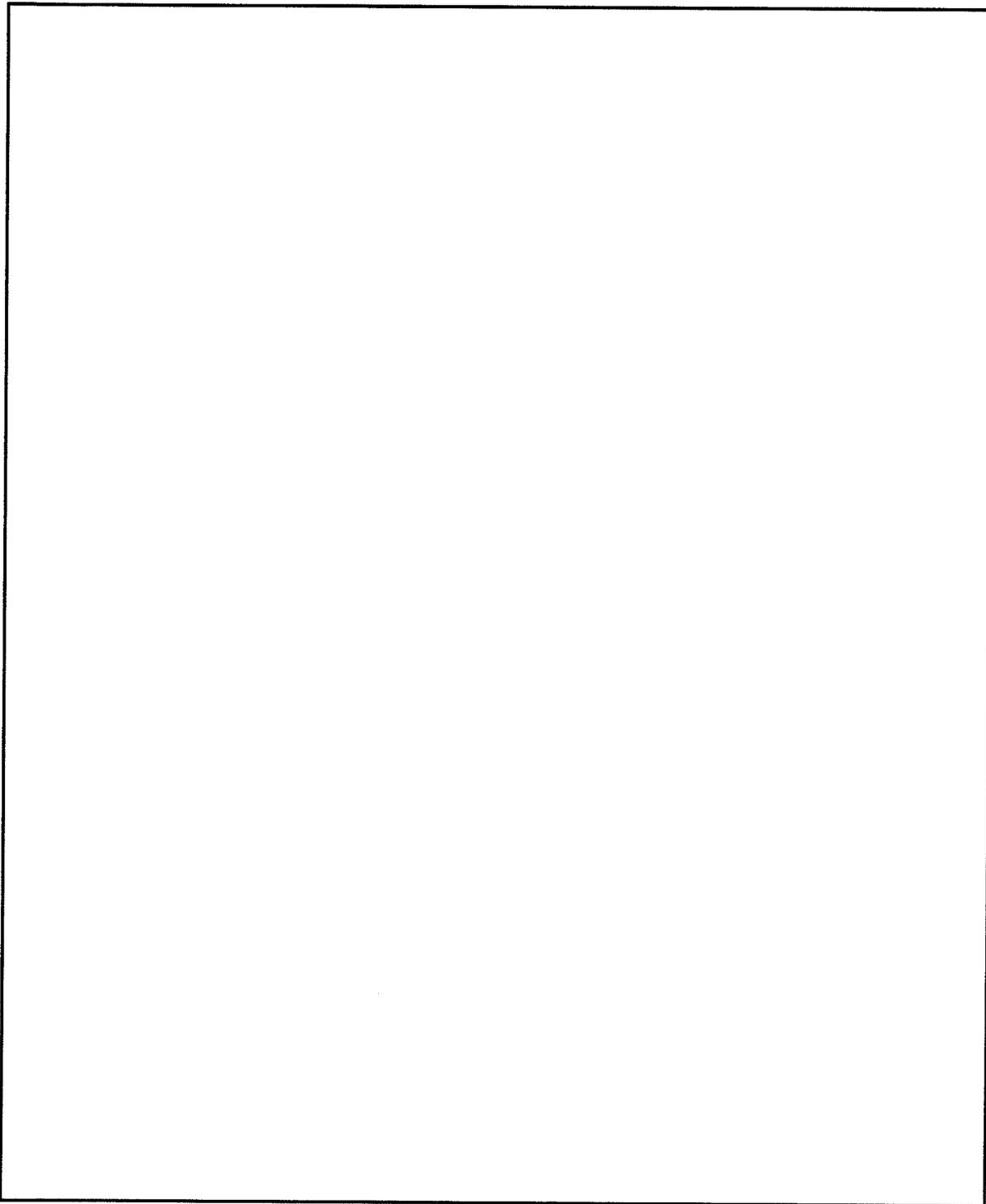
To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005

b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005

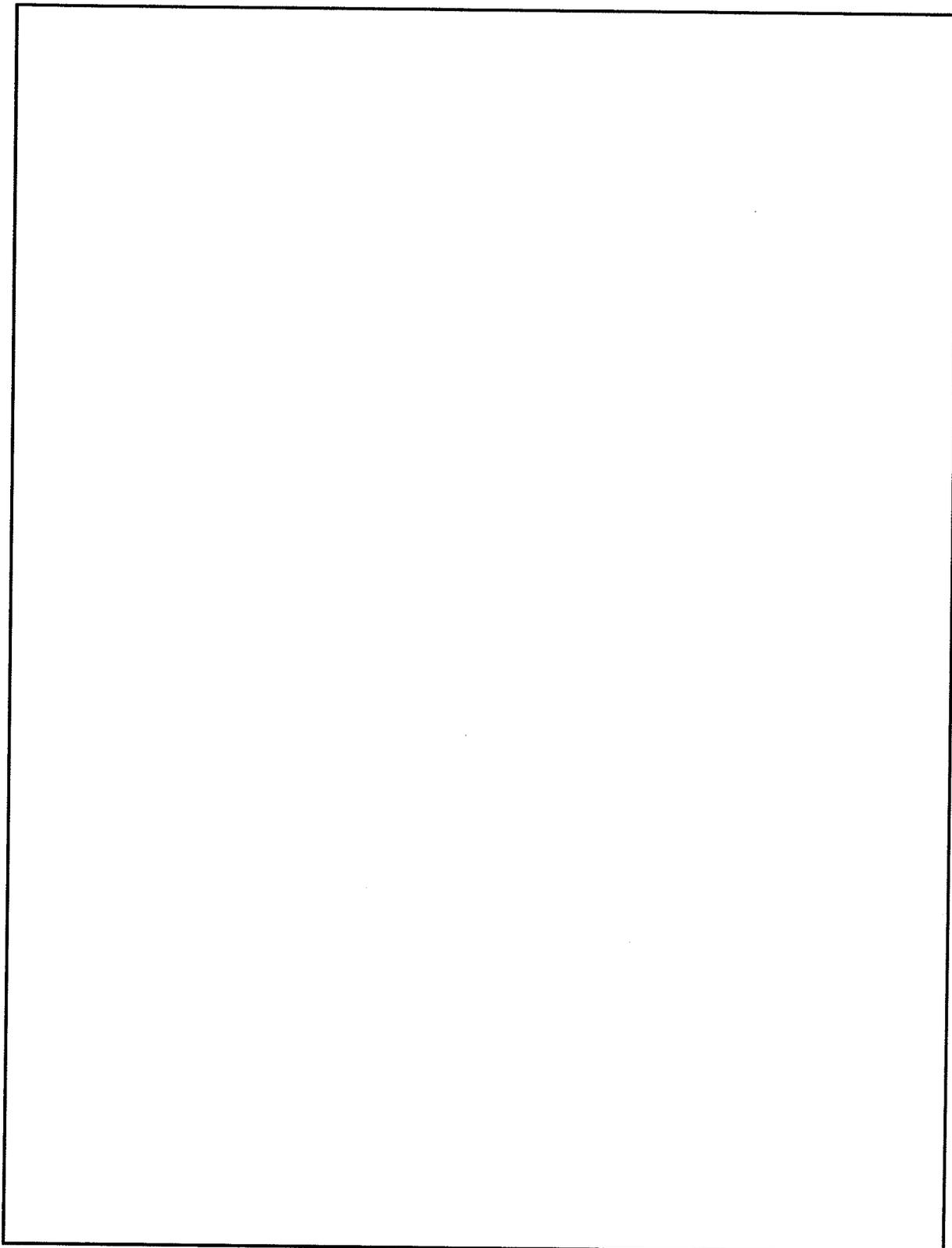
b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005



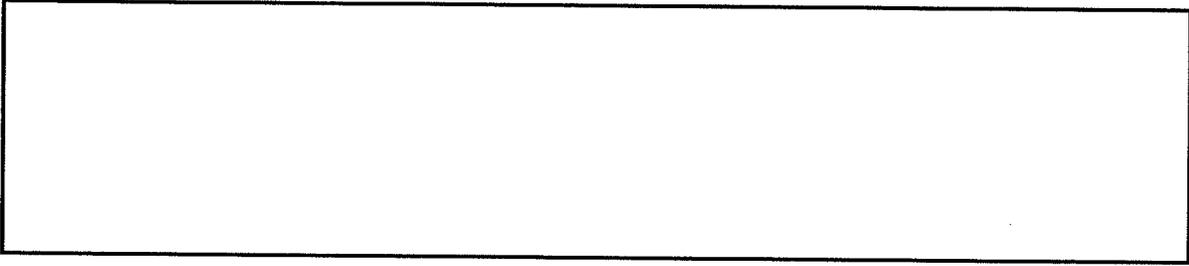
b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005

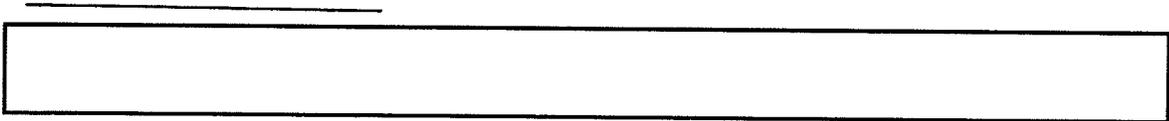


b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005



b2
b7E



b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005

Enclosure 2
Request for additional data sets

Original Message-----

From: [redacted] (CTD) (FBI)
Sent: Thursday, February 17, 2005 10:54 AM
To: [redacted] (OGC) (FBI) [redacted] (OI) (OGA)
Subject: RE: [redacted]

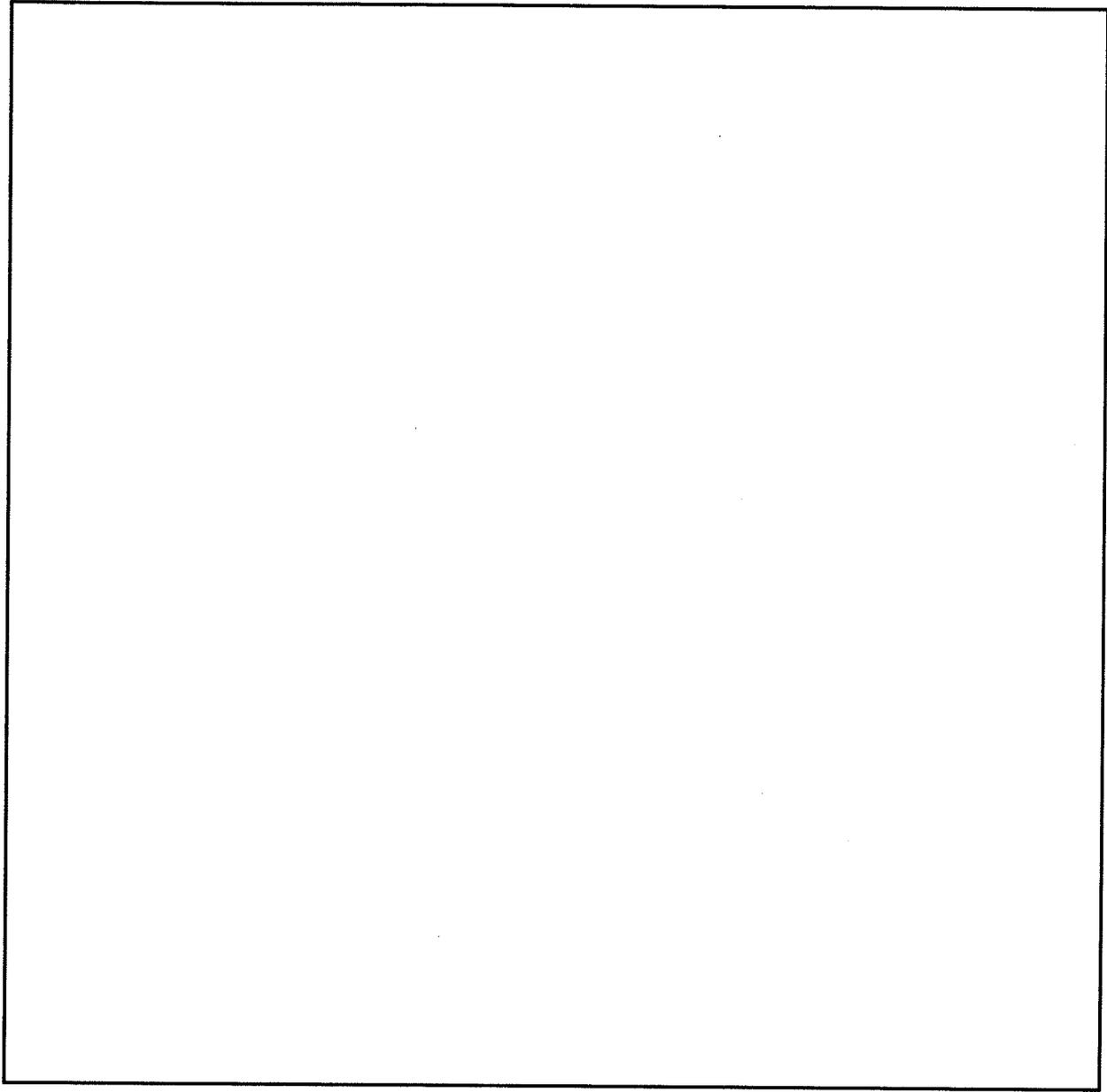
b2
b6
b7C
b7E

SENSITIVE BUT UNCLASSIFIED
NON-RECORD

SENSITIVE BUT UNCLASSIFIED
NON-RECORD

b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005



b2
b6
b7C
b7E

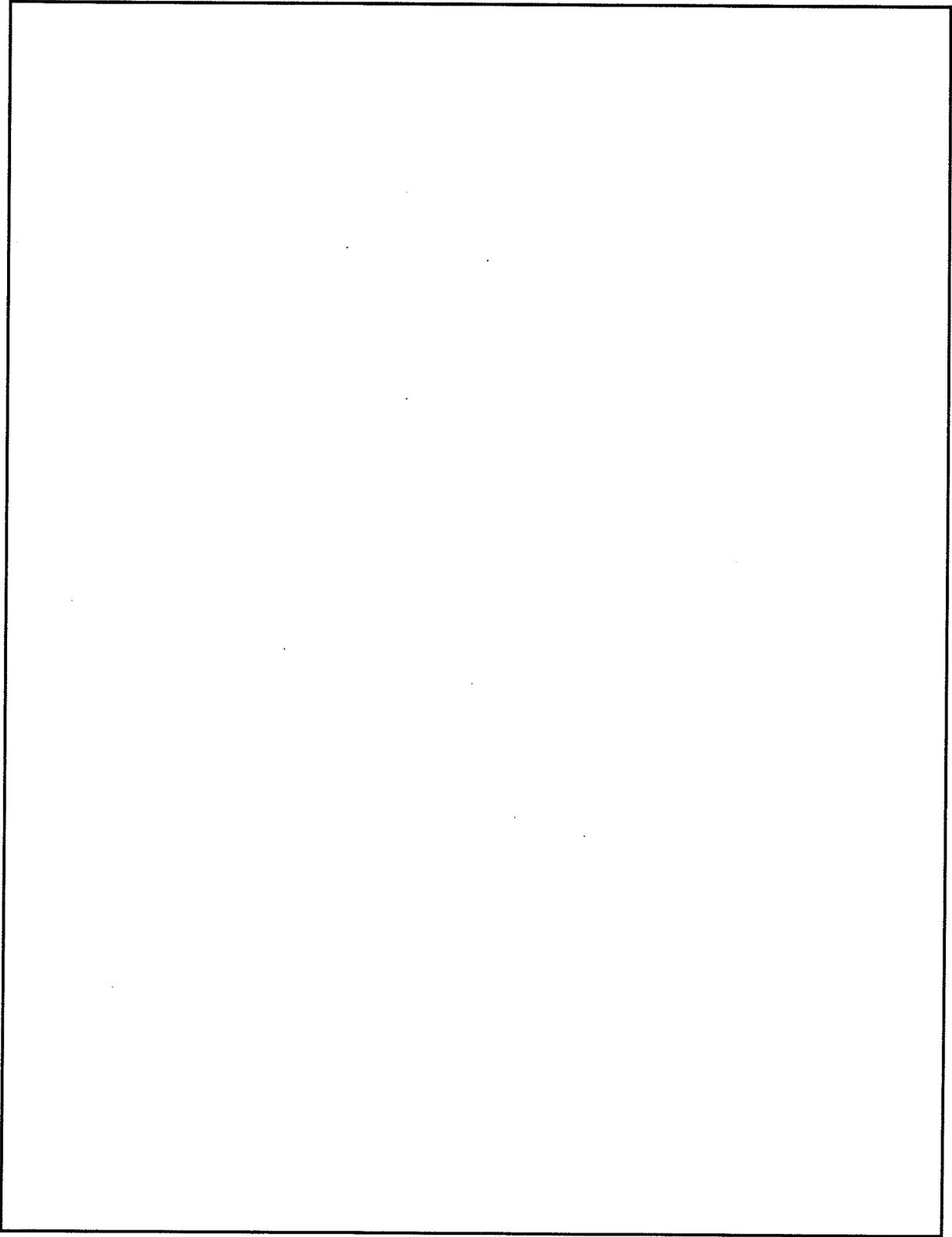
-----Original Message-----

b2
b6
b7C
b7E

From: [redacted] (OGC) (FBI)
Sent: Wednesday, February 16, 2005 1:04 PM
To: [redacted] (CTD) (FBI); [redacted] (OI) (OGA)
Subject: [redacted]

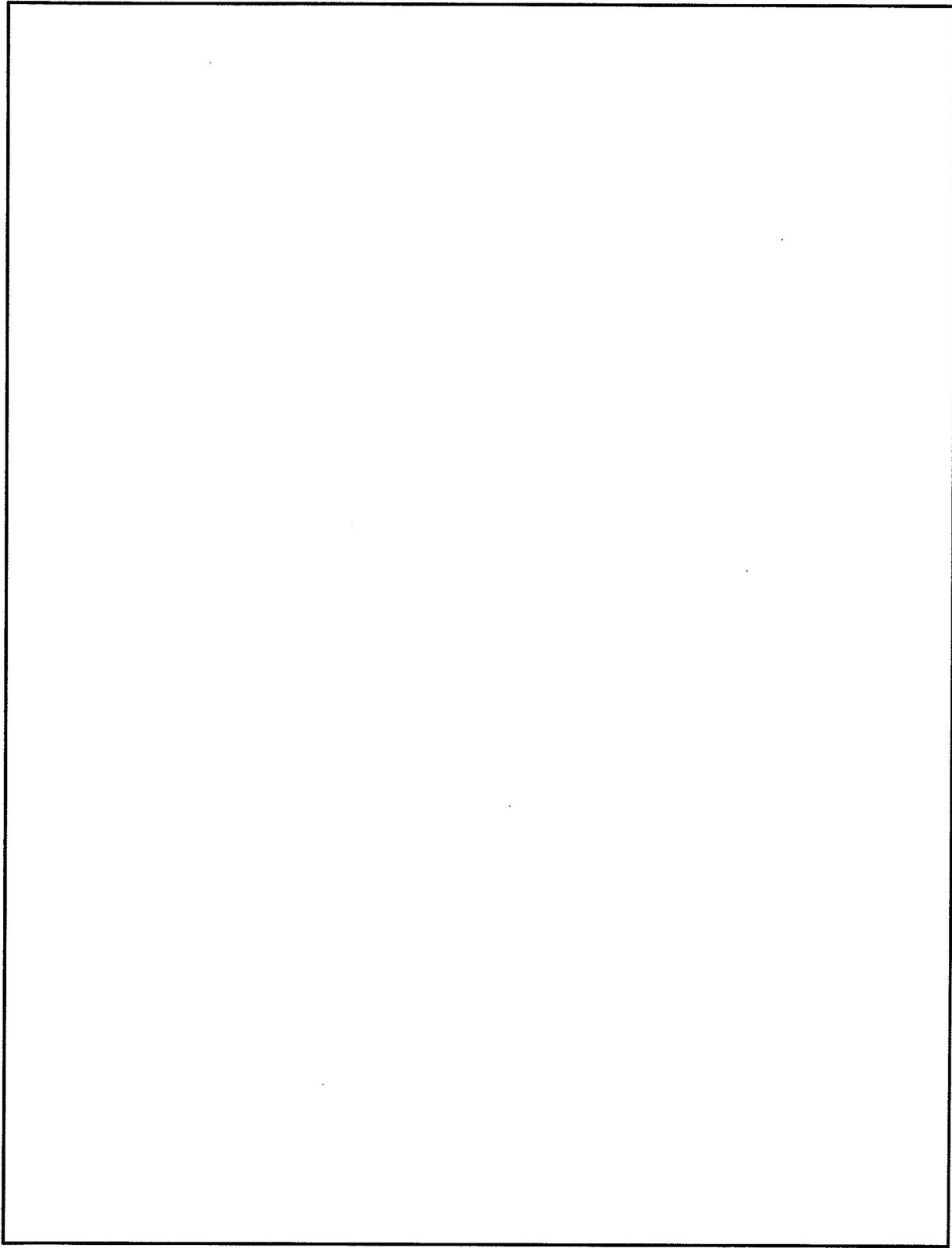
UNCLASSIFIED
NON-RECORD

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005



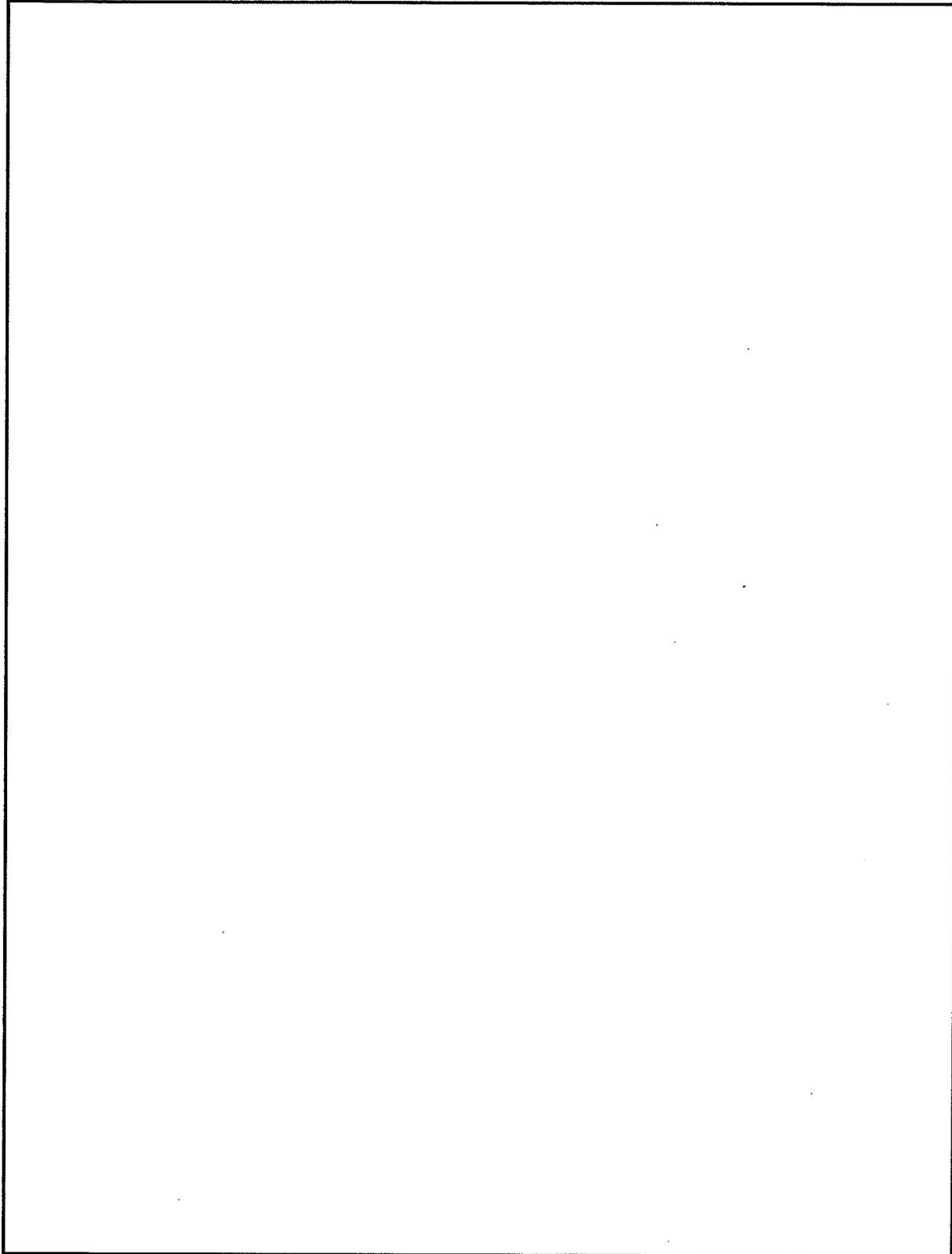
b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005



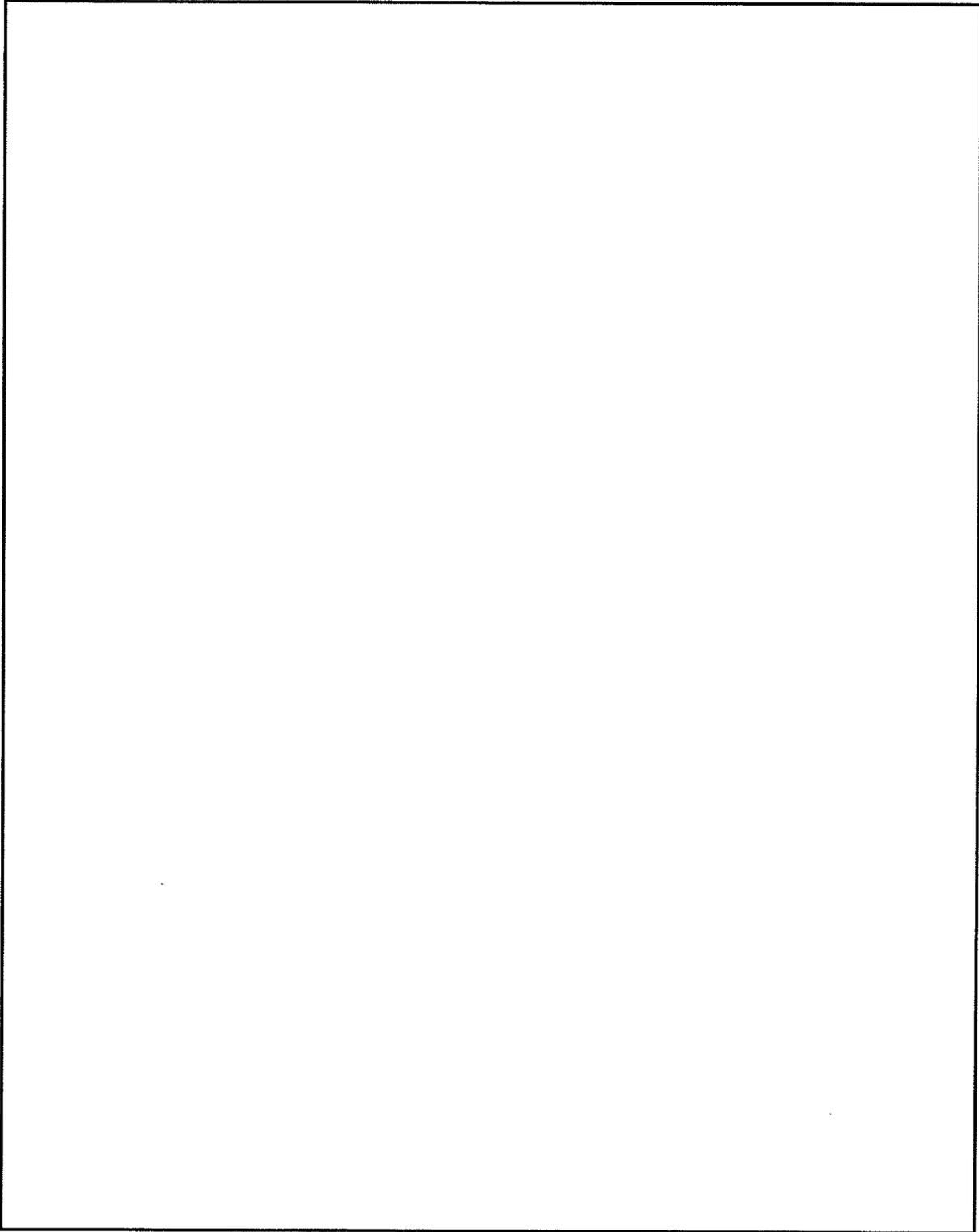
b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005



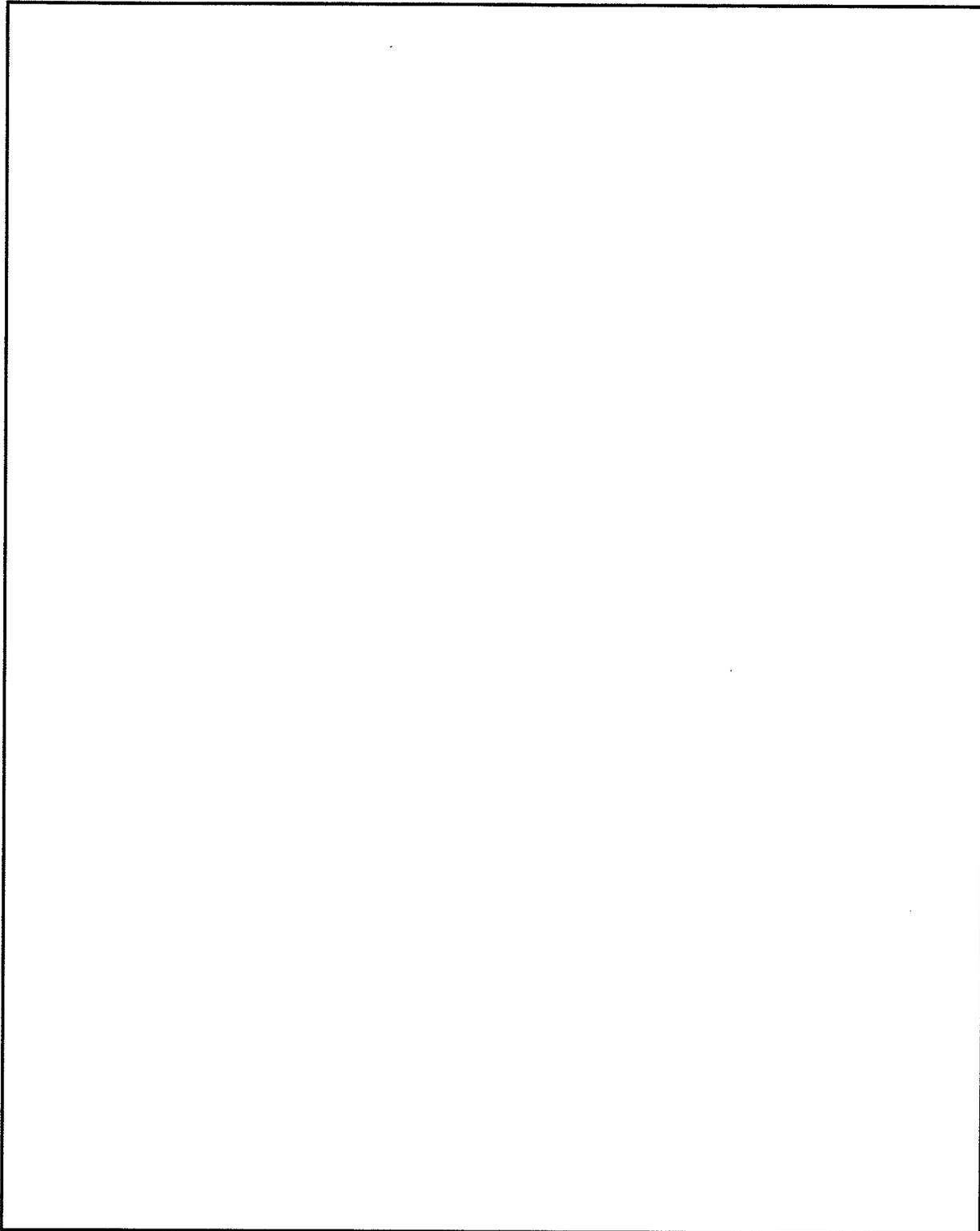
b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005



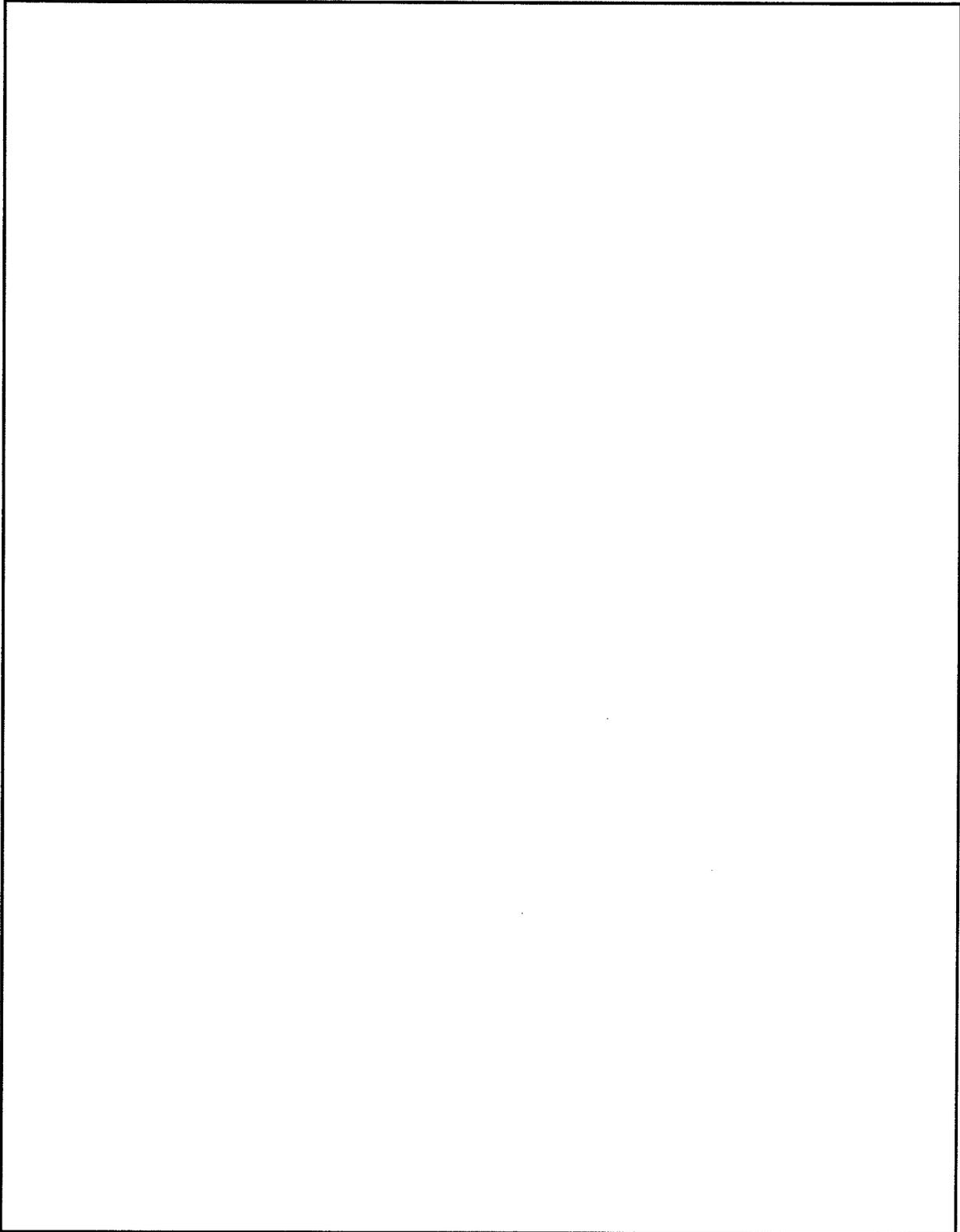
b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005



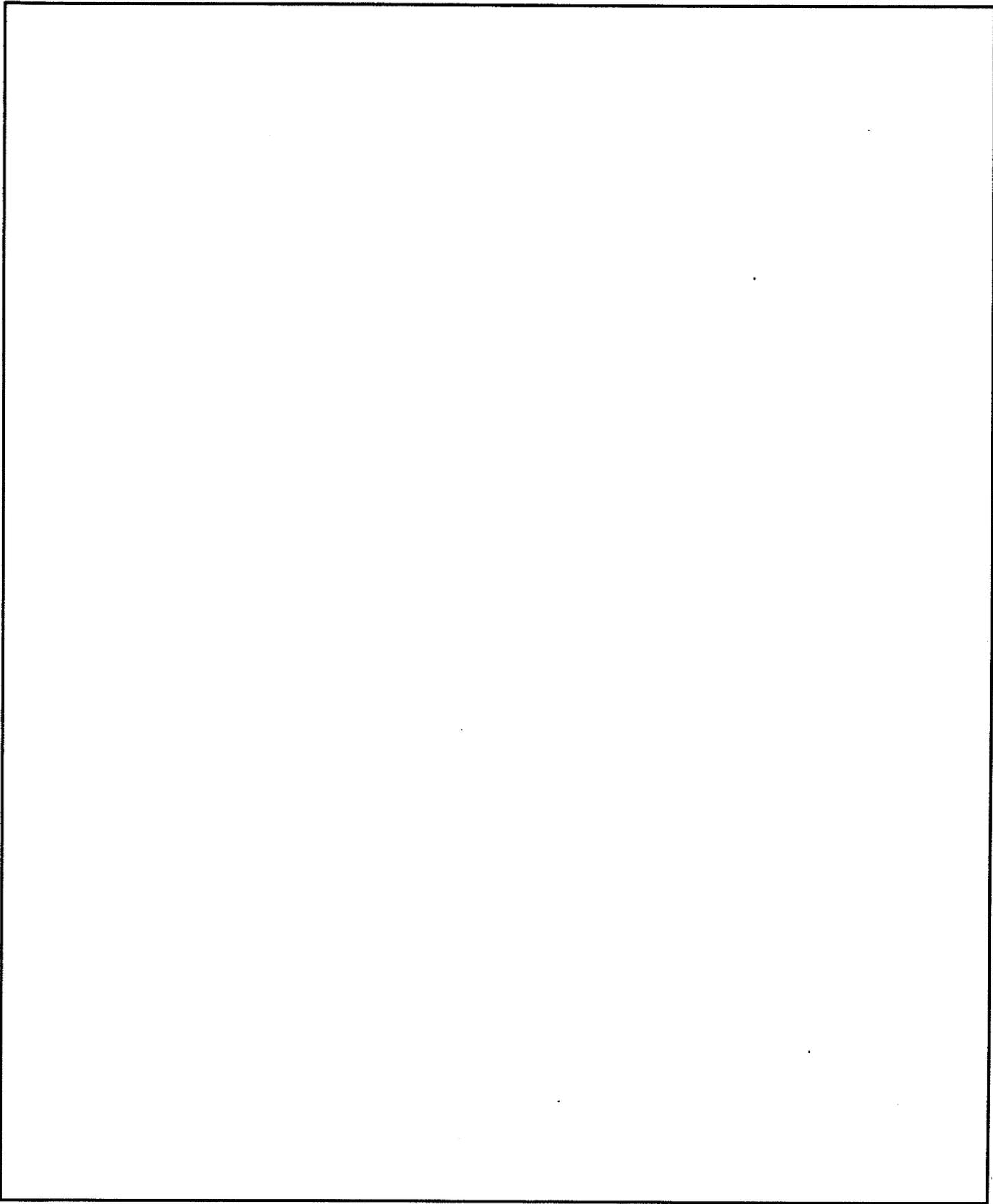
b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005

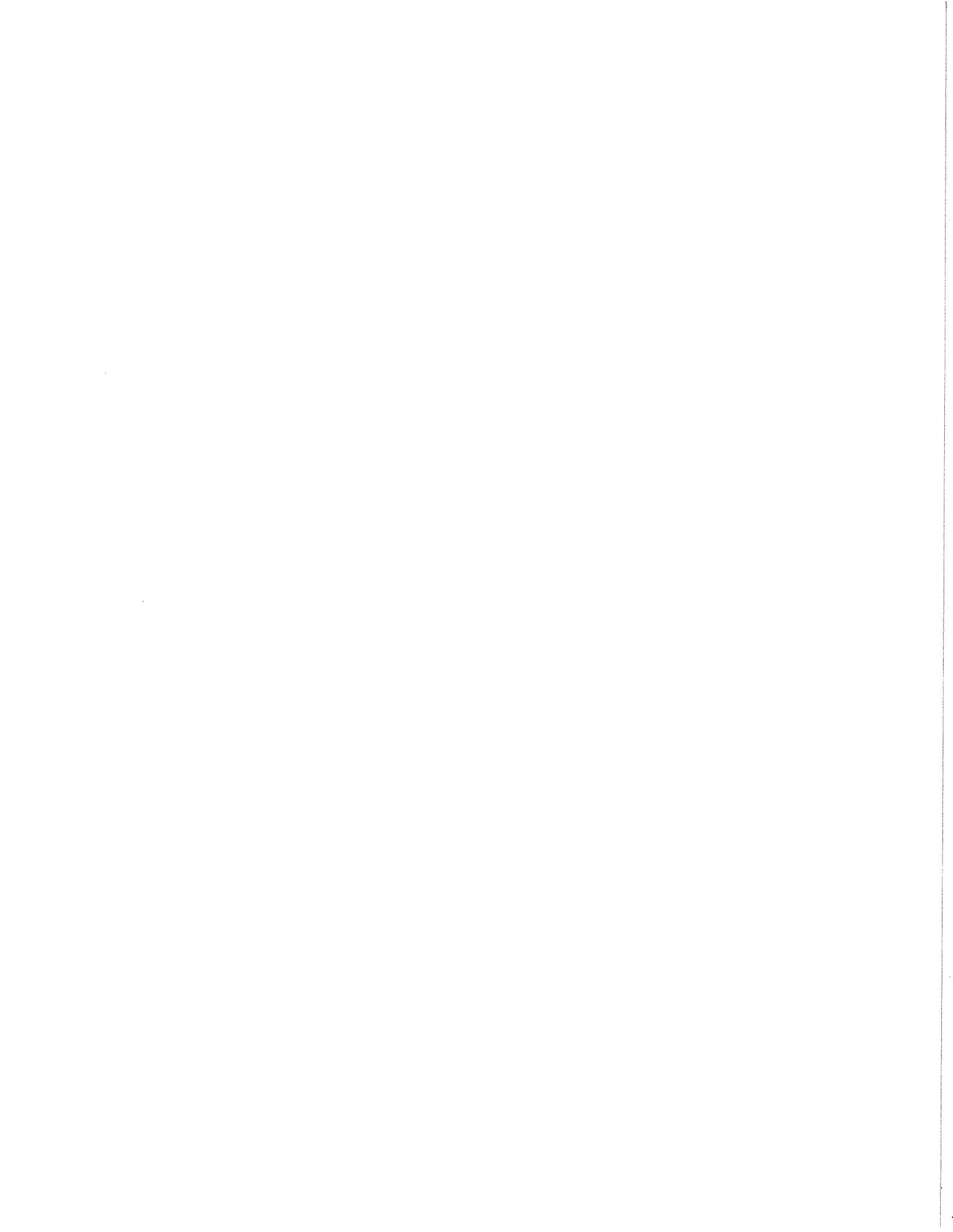


b2
b7E

To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005



b2
b7E



To: Counterterrorism From: General Counsel
Re: 66F-HQ-C1321794, 3/30/2005

b2
b7E

◆◆

~~SECRET~~

~~FOR OFFICIAL USE ONLY~~

DATE: 07-11-2007
CLASSIFIED BY 65179 DMH/BJA/CAL
REASON: 1.4 (C D)
DECLASSIFY ON: 07-11-2032

1/30/2007

IDW Data Contention and Audit Inventory for 2006 1058805

SUBJECT CASE ID	Serial	Collection	Requestor (via EC or email)	Date	ID Number	Action/Notes
281M [redacted]		ACS	[redacted]	21-Dec-06 2006-21DEC-01		4 files deleted under this case on 12/21/06
194A [redacted] 272B- [redacted]		ACS	[redacted]	20-Nov-06 2006-20NOV-01		No cases found under the 194 ID. Reclassed from 272B to a 194.
315M [redacted]	79	ACS	[redacted]	14-Nov-06 2006-14NOV-01		Removed serial 14NOV06. Removed meta data 16NOV06
332C [redacted]		ACS	[redacted]	1-Nov-06 2006-01NOV-01	332C [redacted]	removed 9 files.
65T [redacted]	2	ACS	[redacted]	31-Oct-06 2006-31OCT-01		Removed 65T's (33)
n/a		SAR	[redacted]	31-Oct-06 2006-31OCT-01A		Audited Revised SAR Losses per request.
66F [redacted]		ACS	[redacted]	23-Oct-06 2006-23OCT-01A		Provided Audits for the 2 issues.
315C [redacted]		ALL	[redacted]	29-Sep-06 2006-29SEP-01		Removed 19 doc id's per requested search terms.
101 [redacted]		SAMNET	[redacted]	27-Sep-06 2006-27SEP-01A		Blocked case id in [redacted] Index

b2
b6
b7C
b7A
b7E

~~SECRET~~

~~FOR OFFICIAL USE ONLY~~

ALL INFORMATION CONTAINED
HEREIN IS UNCLASSIFIED EXCEPT
WHERE SHOWN OTHERWISE

IDW Data Contention and Audit Inventory for 2006

3150 [redacted]	1 and 4	ACS	[redacted]	11-Sep-06 2006-11SEP-01A	Provided Audit for these case id's on 28Sept06
3150 [redacted] [redacted] [redacted] [redacted]	38 and 39	ACS	[redacted]	11-Sep-06 2006-11SEP-01A	Provided Audit for these case id's on 28Sept06
[redacted]		ACS	[redacted]	21-Aug-06 2006-21AUG-01	Removed from Quarantine
66F [redacted]		ALL	[redacted]	10-Aug-06 2006-10AUG-01A	No Results for requested audit.
315C [redacted]	ALL	ACS	[redacted]	21-Jul-06:2006-21JUL-01	Removed all docs from collection on 21JUL06.
n/a		ALL	[redacted]	15-Jul-06 2006-15JUL-01A	Provided Audit for this issue.
315H [redacted]	ALL	ALL	[redacted]	21-Jun-06:2006-21JUN-01A	Provided audit for specific documents.
100 [redacted] [redacted] 311A [redacted] [redacted]	24	ACS	[redacted]	9-Jun-06 2006-09JUN-01	Removed this document on 12JUN06
n/a		ALL	[redacted]	5-May-06:2006-05MAY-01A	Provided audit for all users and all data deletions for IDW from 31DEC06- 05MAY06 on 09MAY06.

b2
b6
b7C
b7A

IDW Data Contention and Audit Inventory for 2006

n/a

ACS

[Redacted]

17-Apr-06 2006-17APR-01A Provided audit for issue.

[Redacted]

2006-27APR-01 & 27-Apr-06 01A Removed 7 documents on 28APR06 and provided requested audit on 01MAY06.

[Redacted]

2006-27APR-01 & 27-Apr-06 01A Removed 7 documents on 28APR06 and provided requested audit on 01MAY06.

66F

[Redacted]

[Redacted]

6-Mar-06 2006-06MAR-01A Provided audit for issue.

b2
b6
b7C
b1
b7A

(S)

1058805

The data expungement and corrective actions processes that are utilized by IDW are identified in the Investigative Data Warehouse–Secret Version 1 (IDW-S V1) *Data Administration Manual (DAM)*, Version 0.6, 23 DEC 2005, Section 4, as excerpted below.

For files that are unauthorized due to classification issues, the following process applies.

4. IDW-S Data Security Administration

As noted earlier, the IDW-S system is authorized to hold and process national security data classified up to and including Secret. The IDW-S system is not authorized to process any Top Secret data nor any Sensitive Compartmented Information (SCI). To ensure that IDW-S contains only data for which it is authorized, all data received by IDW-S is subjected to an automated process of [redacted]

[redacted]

b2
b7E

[redacted]

b2
b7E

[redacted] The procedure for deleting individual files from IDW-S is provided below.

[redacted]

b2
b7E

[redacted] The procedure for secure deletion of individual files [redacted] is also provided below.

These process are also outlined in the Federal Bureau of Investigation (FBI) Investigative Data Warehouse (IDW) *System Security Plan*, Version 2.0, dated May 31, 2006, Section 3.1.3.

For files that are unauthorized due to categorization or content issues, the following process applies.

4.1 Deleting Individual Files from IDW-S

In spite of the many precautions taken, it can occur that data for which IDW-S is not authorized is ingested into IDW-S. When such data is discovered on IDW-S it is necessary to delete this data and to update the Document Tracking Database with the appropriate "DEL" status for the file. For this purpose [redacted]

[redacted] was created. There are three usages for [redacted]

b2
b7E

- Usage 1: [redacted]
- Usage 2: [redacted]
- Usage 3: [redacted]

where

- [redacted] is the option to create a "delete file" full filename(s) and filepath(s) of the files to be deleted.
- [redacted] is a text file containing the IDW Document ID's [redacted] of the files to be deleted.
- [redacted] is the option to delete all files with the given IDW Document ID's from the filesystem and to update the Tracking Database with the appropriate "DEL" status for the files.
- [redacted] is the name of the "delete file" containing the full filename(s) and filepath(s) of the files to be deleted. The [redacted] is created in the same filepath as the [redacted]. The format of [redacted] is [redacted]
- [redacted] is an option to update the Tracking Database with "DEL" status for the files but not to perform a delete action on the files. This option is provided for the case where the files have been previously (e.g., manually) deleted off the filesystem.

b2
b7E

b2
b7E

b2
b7E

Note that these three usages enable two modalities with respect to deleting files off of IDW-S:

- Mode 1: Usage 1 followed by Usage 2 deletes files with the IDW Document ID's specified in [redacted] from the filesystem updates the Tracking Database with the appropriate "DEL" status for the files.
- Mode 2: Usage 1 followed by Usage 3 updates the Tracking Database with "DEL" status for the files specified in [redacted]. This mode is used to reconcile the Tracking Database when the files have been previously (e.g., manually) deleted off the filesystem.

b2
b7E

When executed [redacted] reads the IDW Document ID values in [redacted] and for each IDW Document ID the program:

b2
b7E

- Retrieves the filename and filepath from the Tracking Database.
- Generates a batch ID and updates the [redacted] field of the [redacted] table in the Tracking Database with this batch ID.
- Inserts a new DEL event into the [redacted] table in the Tracking Database.
- Enters the notation "Security Delete" into the [redacted] field of the [redacted] table in the Tracking Database.

b2
b7E

b2
b7E

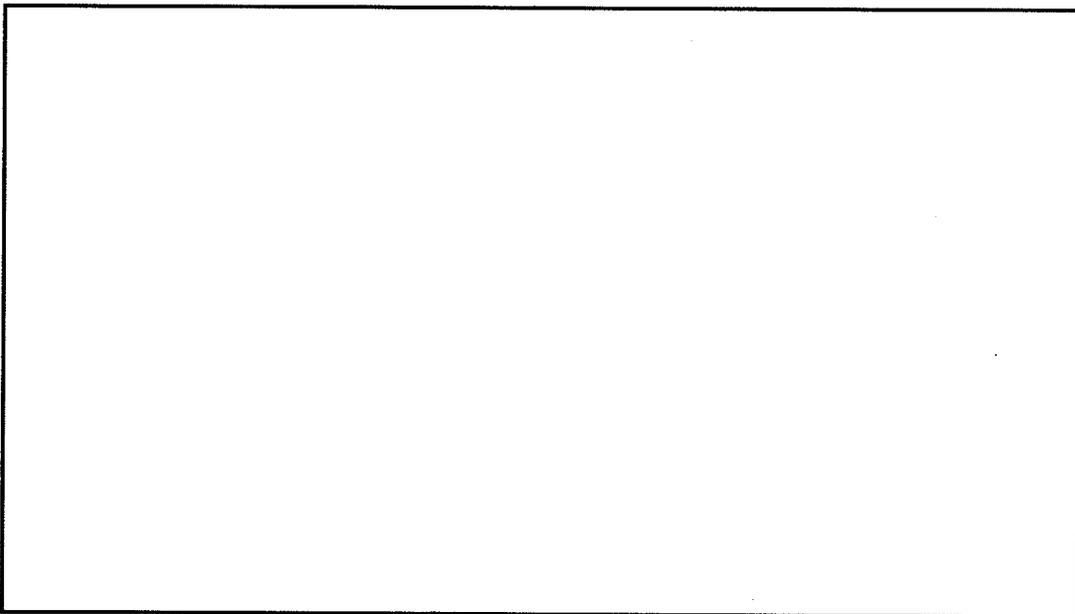
A log file that captures the file deletions and database update actions of [redacted] is created in the location [redacted]

b2
b7E

Auditing:

Specific auditing procedures and requirements are identified in the Federal Bureau of Investigation (FBI) Investigative Data Warehouse (IDW) *System Security Plan*, Version 2.0, dated May 31, 2006, Section 7.6.

IDW-S employs a combination of operating system, network, and application level auditing to record authorized activities and to detect and audit unauthorized system behaviors. All systems perform routine auditing of system and application level security events. Other commercial applications are used by IDW to enhance auditing and monitoring capabilities. Furthermore, specific application auditing provides final correlation of user-to-object access.



b2
b7E

Audit reports can be customized and provided upon request.



Congressional Affairs Office Congressional Contacts

Date Entered: 05/21/2004 Briefing Hearing Other

2004-736 Event Date: 5/12/2004

Subject: National Research Council Report

CAO Contact Person: _____

DOJ Notification: _____ DOJ Date/Time: _____

FBI Participants: CIO Zal Zami

Other Participants: _____

Committees /Subcommittees: HPSCI

Members/Staff: staff: Bob Myhill, Patrick Kelly, Mike Fogarty

Details of Briefing:

Zal advised that the NRC report is outdated and that the NRC would be producing a new, updated report to reflect the changes which the FBI has made to its information technology. He said that the NRC reps did not allow the FBI to respond to the findings before releasing the report. Zal discussed what IDW does (currently 9 data sources - analysis across these data sources) versus VCF (data flow and data generation). In response to Bob's question about who is responsible for enterprise architecture coordination within the IC, Zal said Alan Wade (overall) coupled with 5 working groups.

Follow Up Action:

b6
b7C

108th



Congressional Affairs Office Congressional Contacts

Date Entered: 01/14/2005 Briefing Hearing Other FOC

2005-1 Event Date: 1/13/2005

b6
b7C

Subject: VCF Status Briefing for Senate Select Committee on Intelligence (staff only)

CAO Contact Person: SSA [redacted]

DOJ Notification: None DOJ Date/Time: [redacted]

b6
b7C

FBI Participants: CIO Zalmay Azmi (Briefer), AD Eleni Kalisch, SSA [redacted] (OCIO)

Other Participants:

Committees /Subcommittees: Senate Select Committee on Intelligence

b6
b7C

Members/Staff: [redacted]

Details of Briefing:

[redacted]

[redacted] This is compared to IDW which is a warehouse containing 47 databases (including ACS) which also can be searched for data (including paper files). [redacted]

OTHER
O/S

[redacted]

Follow Up Action:

None

109 *th*



Congressional Affairs Office Congressional Contacts

Date Entered: 02/02/2005 Briefing Hearing Other FOC

2005-21 Event Date: 2/1/2005

OTHER O/S

Subject: IDW [redacted]

CAO Contact Person: [redacted]

b6
b7C

DOJ Notification: [redacted] DOJ Date/Time: [redacted]

FBI Participants: Zal Azmi [redacted] (ACS demo) [redacted]

Other Participants:

Committees: [redacted]

b6

/Subcommittees: House Appropriations

b7C

Members/Staff: [redacted]

Details of Briefing:

The staff were provided a demo and briefing on IDW and ACS. [redacted] conducted the IDW presentation/demo. He provided details on the sources of information contained in IDW, # of users (currently 6,000), plans for expansion, # of databases (47), privacy issues, mou(s) regarding information sharing with other federal agencies, states and local entities. [redacted] asked if DEA phone application information was contained in IDW. Answer: no due to security issues. A general discussion was held regarding the possibility of creating new IDWs for other crime problems/initiatives. [redacted]

b6
b7C
OTHER O/S

Follow Up Action:

[redacted]

OTHER O/S



Congressional Affairs Office Congressional Contacts

Date Entered: 05/19/2005 Briefing Hearing Other FOC

2005-178 Event Date: 5/20/2005

Subject: [Redacted]

CAO Contact Person: SSA [Redacted]

DOJ Notification: [Redacted] DOJ Date/Time: 1:00:00 PM

FBI Participants: SC Mike Morehart (TFOS) [Redacted] (TFOS, observer)

Other Participants: [Redacted]

Committees /Subcommittees: House Committee on Financial Services, Subcommittee on Oversight and Investigation

Members/Staff: [Redacted]

Details of Briefing:

[Redacted]

b6
b7C
OTHER O/S
b2
b7E

Follow Up Action:

[Redacted]

b6
b7C
OTHER O/S

b2
b6
b7C
b7E



Congressional Affairs Office Congressional Contacts

Date Entered: Briefing Hearing Other FOC

OTHER O/S

Event Date:

Subject:

b6
b7c

CAO Contact Person: SSA

DOJ Notification: DOJ Date/Time:

FBI Participants: SC Mike Morehart, TFOS, UC and SSA

Other Participants:

Committees /Subcommittees:

Members/Staff:

Details of Briefing:

and provided overview about IDW. Discussed information ingested by IDW and how said information is utilized. Discussed how all info is vetted through Privacy Impact and OGC. Then provided real time examples of data mining. There was discussion about the need to expand the system and how it currently hosts 41 million datasets. Discussion on awaiting financing to increase the system to ingest 71 million more data sets.

OTHER O/S

Follow Up Action:



Congressional Affairs Office Congressional Contacts

Date Entered: 08/01/2006 Briefing Hearing Other FOC

2006-721 Event Date: 5/22/2006

Subject: IDW

CAO Contact Person:

DOJ Notification: DOJ Date/Time:

b6
b7C
OTHER O/S

FBI Participants:

Other Participants:

CRS

Committees /Subcommittees: at the direction of House Approps SSJC

Members/Staff: not present

Details of Briefing:

IDW background and demonstration; users and availability; weaknesses and improvements needed; data composition; cooperation with outside agencies and DNI; intelligence products; Beta version; batch queries; training; financial resources.

OTHER O/S

[Redacted area]

Follow Up Action:

[Redacted area]

OTHER O/S



Congressional Affairs Office Congressional Contacts

Date Entered: 09/13/2006 Briefing Hearing Other FOC

2006-805 Event Date: 9/12/2006

Subject: [Redacted]
 CAO Contact Person: SSA [Redacted]
 DOJ Notification: [Redacted] DOJ Date/Time: [Redacted]
 FBI Participants: None
 Other Participants: [Redacted]
 Committees /Subcommittees: Senate Banking, Housing and Urban Affairs
 Members/Staff: Shelby, Hagel, Martinez, Allard

b6
 b7C
 OTHER O/S

Details of Briefing:

[Redacted]
 [Redacted] also made reference to a presentation he received from the FBI concerning IDW and how the FBI was able to link information received [Redacted] to subjects of ongoing criminal and terrorist investigations.

b2
 b6
 b7C
 b7E

Follow Up Action:

**Responses of the Federal Bureau of Investigation
Based Upon the August 19, 2004 Hearing Before the
Senate Committee on the Judiciary
Regarding "The 9/11 Commission and Recommendations
for the Future of Federal Law Enforcement and Border Security"**

Questions Posed by Senator Hatch

1. The 9/11 Commission has recommended that the position of deputy National Intelligence Director ("NID") for homeland intelligence be filled by either the FBI's executive assistant director for intelligence or the under secretary of homeland security for information analysis and homeland protection. Do you think this recommendation - by failing to specify precisely which official should hold the position - may create an unnecessary conflict between the FBI and the Department of Homeland Security ("DHS")? More generally, do you believe the FBI Office of Intelligence and the DHS Directorate for Information Analysis and Infrastructure perform similar functions, such that the heads of those entities would be interchangeable in the role of a deputy NID?

Response:

The FBI believes the Director of National Intelligence (DNI) should have one principal deputy. We believe the spirit of the 9/11 Commission recommendations can be better achieved through an intelligence coordinating council made up of NSC/HSC principals.

2. You have served in leadership positions within two different components of the Intelligence Community, the National Security Agency and the FBI. Moreover, you have had an opportunity to view the cooperation, or lack of cooperation, among intelligence agencies at the highest levels. If the 9/11 Commission's recommendations are adopted, you could end up serving as a deputy to the NID, as well as reporting to the FBI Director. Based on your experiences, do you think this type of "dual-hatting" can work? In your opinion, are there any conditions that might improve the likelihood of a successful merger of your potential NID and FBI roles?

Response:

We do not think a "dual-hatting" approach is the best answer. We are concerned about dual-hatting deputies who already have full time jobs, we may be replicating the situation underscored by the 9/11 Commission of intelligence community leaders having "too many jobs." In addition, maintaining the operational chain-of-command authority within the agencies that have the

to improve oversight of IT projects, to strengthen oversight of IT contracts, and to ensure that IT investments fully support the FBI's current and future missions.

c. What is the current projection for the final, total cost of the project?

Response:

It is too early to estimate the total cost of the program.

6. John Brennan, the Director of TTIC, testified on August 23, 2004, about the need to build an integrated information technology architecture, accessible to all members of the intelligence community. Do you agree? How would VCF or the Integrated Data Warehouse fit into this new architecture?

Response:

We agree with the need to build a government-wide integrated information architecture as outlined in the President's Executive Order entitled Strengthening the Sharing of Terrorism Information to Protect Americans. In the FBI's work processes, VCF, or its successor software, will be ingest tools (like the Automated Case Support system is now) for the Investigative Data Warehouse (IDW). VCF or its equivalent will be the first point of ingest for investigative and intelligence information and for records collected by Agents and others. IDW then allows the data to be accessed, analyzed, and used in the production of intelligence. IDW minimizes the compartmentalization of intelligence and/or terrorism-related data developed by the FBI and would fit within this new architecture. It would also allow the interchange between agencies, with the proper security and access controls necessary to protect methods and sources.

7. I understand that, after many millions of dollars spent, FBI agents now have the capability of e-mailing each other over a secure network. But I also understand that many field agents are still unable to send secure e-mails to other federal government agencies, or to state and local law enforcement and other entities outside the FBI. Is that true? If so, why does the FBI lack this basic capability, and what if anything is being done about it?

Response:

The FBI is faced with a unique challenge every day. Unlike other law enforcement agencies, we are responsible for communicating with the IC, other federal agencies, and our state and local partners in regional jurisdictions as it relates to our intelligence, counterterrorism prevention and criminal investigative responsibilities. This levies an enormous challenge on our IT resources and staff

The Inspection Division then obtained a copy of the Zyindex database from the OKBOMB investigation, which contained 167,000 documents, and obtained a comparison of the 15,200 documents from the "I" drive tapes, the 167,000 OKBOMB documents, and the documents in the FBI's Automated Case Support system. This comparison identified 891 questionable documents.

A CD-ROM containing the 891 questionable documents was forwarded to the Oklahoma City Division. Based on their knowledge of the documentation provided pursuant to the OKBOMB discovery process, the Oklahoma City Division was asked to determine whether any of these documents that should have been made available for discovery had, in fact, not been provided to the OKBOMB defense team.

The Oklahoma City Division advised that, of the 891 questionable documents, only four had not previously been reviewed by members of the OKBOMB Task Force. Two of the documents were first drafts of FD-302s that were later changed so they could be uploaded to the FBI's Automated Case Support system; one document was an FD-71 complaint form that mentioned OKBOMB and was generated by the Denver Division; and the fourth document was unidentifiable.

c. Were the existence and potential problems caused by the "I-drive" reviewed by the 9-11 Commission?

Response:

While the 9/11 Commission Report does not address the FBI's "I" drives, the 9/11 Commission did review the FBI's data automation and technology processes, finding its information systems "woefully inadequate" during this period (page 77 of the Commission's report).

d. Can analysts access data and documents on the "I-drive" through the Integrated Data Warehouse? If not, why not, and do you plan for this to change.

Response:

The purpose of the Integrated Data Warehouse (IDW) is to facilitate the analysis of data that has been collected and documented by FBI employees. While the IDW will utilize the FBI's network architecture to facilitate the analysis and sharing of data in FBI systems, it will not "see" or pull in data from the "I" drive. This is appropriate because the purpose of the "I" drive is to facilitate the mobility of the FBI's workforce by allowing employees to access their work-in-progress from any computer connected to the FBI network, and documents that have not been reviewed or approved by supervisors may contain inaccurate or incomplete

information. If this information were made available to all analysts, they would risk the possibility of reaching incorrect conclusions based upon unverified data. Once a document is approved, it is uploaded into the FBI's Automated Case Support system, from which information is retrievable and searchable by all employees. Except as described in question 11c, below, these documents could then be accessed by analysts through the IDW.

e. Will the "I-drive" still exist once VCF is implemented? Please explain.

Response:

The "I" drive is a networked computer drive that allows computer users to retrieve items that they are working on from any computer connected to the network. This type of network architecture facilitates the mobile nature of the FBI's workforce, while providing the appropriate security for information and intelligence gathered by the FBI. These network drives are not designed as repositories of information; they are designed to facilitate work that is in progress.

Because VCF, or its successor software, will permit documents to be drafted, reviewed, verified, and approved by supervisors within the workflow process defined by that software, the current use of the "I" drive will no longer be required after that software is deployed. Even then, however, networked drives that allow FBI employees to access their work in progress from any networked computer will still be a necessary part of the FBI's Enterprise Architecture. Consequently, while these shared drives may be called "I" drives or may use some other naming convention, shared drives will continue to have utility in the FBI, though for different purposes than the "I" drive is currently used.

11. During your testimony, you said that "case files" were included in the Integrated Data Warehouse (IDW). It is my understanding that FBI case files include documents such as FD-302's (interview memoranda), electronic communications, documents obtained by the FBI in the course of an investigation (and filed in "1A" envelopes with the case file), transcripts of wiretap recordings, as well as other materials.

a. Please confirm that these items are included in a typical FBI "case file" and explain what, if any, other types of documents or materials are kept in a "case file."

Response:

The above listed items are kept in a case file. In addition to electronic communications (ECs), FD-302s (Form for information that may become testimony), and transcripts, other types of data stored in a case file include

Facsimiles, FD-542s (Investigative Accomplishment Reports), Inserts, Teletypes, Letter Head Memorandums (LHM), Memorandums, and other miscellaneous documents.

b. Are all of these items accessible through the IDW?

Response:

Except for those items described below in item (c), all of these items are accessible through IDW.

c. What if any documents or materials kept or maintained in an FBI "case file" are *not* accessible in IDW, and why? Please be specific.

Response:

Most, but not all, electronic documents or materials kept in an FBI case file are accessible through IDW. A small number of case file documents that identify specific types of data too sensitive for all IDW users are not accessible through IDW. For example, information that reveals the identities of informants, information on public corruption investigations, and some administrative "case files" such as FBI employee disciplinary actions would not be accessible.

Prior to September 11, 2001, information in case files was primarily restricted to agents directly involved with the respective cases. Following September 11, 2001, Director Mueller established an "open data" policy, which permitted FBI analysts to access all data in FBI systems, with the exception of the most sensitive files identified by the EAD for Counterterrorism/Counterintelligence. This policy change allowed counterterrorism analysts to make more effective use of the FBI's collected data.

In accordance with the "open data" policy, the IDW system allows users to access all data in the system, although "need-to-know" principles still apply. The restrictions described above are intended to protect the FBI's most sensitive data from threats such as that posed by Robert Hanssen. To further protect against this type of threat, IDW audits all user activity.

As is further described in part (d) below, the FBI is aggressively developing a more advanced security system that would allow all documents to be included in

the data warehouse, with strict protections applied to the most sensitive documents.

In order to ensure that FBI policies create the most effective counterterrorism environment possible, Director Mueller established an Information Systems Policy Board that is charged with reviewing existing policies, modifying policies when necessary, and establishing new policies as needed to respond to a changing environment.

d. For any documents or materials not accessible through IDW, please detail how the FBI currently searches for data in such documents or materials, and how or whether the search is conducted differently today than it was prior to September 11, 2001. For documents not currently accessible in IDW, when will the FBI will be able to access such materials electronically?

Response:

The documents not available through IDW are currently accessed through their original sources' systems, as they were prior to September 11, 2001. However, the access rules applied to these systems have changed in response to the events of September 11 to provide greater access and enhanced auditing features. This provides a greater ability to locate and disseminate data than the FBI had prior to September 11, 2001.

The FBI is actively working on a project based on the IDW system that will add a more robust security layer, which includes the detailed discretionary access controls required for the FBI's most sensitive files. The FBI anticipates completion of the testing and evaluation of the new technology in the summer of 2005. If additional funding is secured, the FBI will initiate the process of loading the excluded documents described in part (c) above into the system with appropriate protections. Access will then be expanded to the full user base of IDW.

e. Is it true that IDW access to materials in an FBI "case file" is limited to only that information that has been typed by an agent or support personnel into an FD-302 or other report?

Response:

This is not true. There is a great deal of information in IDW other than that which has been typed by an agent or support personnel into an FD-302 or other report. With only the exceptions described in part (c) above, users have access to all electronic data that is stored in ACS, as well as other paper records which have

been automatically scanned and converted into computer text. These scanned documents include Bureau-generated documents related to terrorism, as well as other terrorism-related documents such as those seized in Afghanistan and Pakistan. Also large quantities of data from other agencies, including DIA, NSA, CIA, DOS, and FinCEN have been ingested into IDW.

f. Are all investigative materials obtained by the FBI by subpoena, by NSL or by other means always reviewed contemporaneously and summarized in report form, such that they are accessible through the IDW? If not, why not?

Response:

All investigative materials obtained by the FBI by subpoena, NSL, or by other means (such as that provided by 18 U.S.C. §2703) are reviewed contemporaneously. Not all investigative materials reviewed are deemed pertinent to a case. Those materials that are reviewed and deemed pertinent to a case are either summarized, in which the case summary is loaded into ACS, or the entire document is scanned, if necessary, and uploaded in its entirety into IntelPlus.

Many of the largest IntelPlus file rooms have been imported into IDW, so these documents would be accessible through the IDW in both text form and the original scanned images. Summaries loaded into ACS would be accessible through the IDW, except as noted in answer 11(c).

The only investigative materials that would not be available through the IDW are those that were not deemed pertinent to a case, those that were added to an IntelPlus file room that has not yet been incorporated into IDW, or those that are too sensitive to load into IDW, as described in answer 11(c).

g. What is the time frame for the dataset "case file" material that is currently accessible by IDW? In other words, are FD-302s that were written in 1995, 1990, or even prior to 1985 accessible?

Response:

The time frames for the datasets vary. Except as noted in part (c) above, all data stored in ACS, including FD-302s, are available in IDW. Since ACS was created in 1995, IDW contains ACS data from 1995 to present. IDW also contains millions of scanned paper documents, including those seized from suspected terrorists. Although the FBI knows the dates these documents were added into IDW, the date of origin of many of these documents is unknown.

As additional data sources continue to be added into IDW, most contain records dated prior to the date of ingest. All of this "day back" information will be included in IDW. The specific date ranges of the data will vary by source, and may include data prior to 1985. For example, IDW includes all CIA Intelligence Information Reports (IIR) at the Secret or lower classification levels issued from 1978 to present. Conversely, most data sources provide updates of new data created after the initial date of ingest. These "day forward" updates will continue to be added into IDW and appended to the appropriate data libraries.

h. You gave a "specific example" in order "to show this set of data that included a lot of different things, including case files, but not all case files, but terrorism information." Can you explain what you meant by this statement including the phrase "but not all case files, but terrorism information"?

Response:

The statement was intended to emphasize that the set of data includes terrorism information. The statement could be more clearly conveyed using two sentences: "The IDW included a lot of different types of data, including case files. IDW may not currently include all case file data (as discussed in question 11.c. above), but it does include terrorism information."

12. In early 2003, Director Mueller described the IDW as a future goal of the FBI that would encompass "31 different databases" and would be used to help the FBI conduct "data mining."

a. Please identify and provide a brief explanation of each database currently included in, or currently planned to be included in, the IDW. Approximately when was each database made accessible through IDW?

Response:

The following data sources are currently available through IDW. Other data sources that are planned to be added, pending approval by the Policy Board and the Office of General Counsel's (OGC) review of the Privacy Impact Assessment, are listed below in the response to (b).

Currently Included (Added Prior to January, 2004):

- Automated Case System (ACS), Electronic Case File (ECF)
- Secure Automated Messaging Network (SAMNet) – copies of all messaging traffic sent either from the FBI to other government agencies, or sent from other government agencies to the FBI through the Automated Digital Information Network (AutoDIN).

- Joint Intelligence Committee Inquiry (JICI) Documents – scanned copies of all FBI documents related to extremist Islamic terrorism between 1993 and 2002.
- Open Source News – various foreign news sources that have been translated into English, as well as a few large U.S. publications, such as the Washington Post.
- Violent Gang and Terrorist Organization File (VGTOF) – lists of individuals and organizations associated with violent gangs and terrorism, provided by the FBI National Crime Information Center (NCIC)

Currently Included (Added Between January 2004 and Present):

- 11 Financial Crimes Enforcement Network (FinCEN) Databases – data related to terrorist financing
- 2 Terrorist Financing Operations Section Databases - biographical and financial reports on terrorism-related individuals
- 11 Scanned document libraries – millions of scanned documents related to FBI's major terrorism-related cases
- CIA Intelligence Information Reports (IIR) and Technical Disseminations (TD) – copy of all IIRs and TDs at the SECRET security classification or below that were sent to the FBI from 1978 to present
- Foreign Financial List – copies of information concerning terrorism-related persons, addresses, and other biographical data submitted to U.S. financial institutions from foreign financial institutions
- Selectee List – copies of a Transportation Security Administration (TSA) list of individuals that warrant additional security attention prior to boarding a commercial airliner
- Terrorist Watch List (TWL) – the FBI Terrorist Watch and Warning Unit (TWWU) list of names, aliases, and biographical information regarding individuals submitted to the Terrorist Screening Center (TSC) for inclusion into VGTOF and TIPOFF watch lists
- No Fly List – copy of a TSA list of individuals barred from boarding a commercial airplane
- Universal Name Index (UNI) Mains – copy of index records for all main subjects on FBI investigations, except as mentioned in part (c) of question 11 above.
- Universal Name Index (UNI) Refs – copy of index records for all individuals referenced in FBI investigations, except as mentioned in part (c) of question 11 above.
- Department of State Lost and Stolen Passports - copy of records pertaining to lost and stolen passports
- Department of State Diplomatic Security Service – copy of past and current passport fraud investigations from the DOS DDS RAMS database

Planned Data Sources:

- (See part b below)

b. You stated in your testimony that the FBI "through a policy board" is looking specifically at IDW and trying to add to the data sets that are in there. How does the policy board operate and what other databases are being considered for inclusion in the IDW?

Response:

The Director created an Information Sharing Policy Group, co-chaired by the Executive Assistant Director - Intelligence and the Executive Assistant Director - Administration. This group reviews all requests for new data, as well as the dissemination controls imposed upon data sets. Before a data set can be approved by the policy board, or dissemination controls can be changed, the FBI's OGC must review and approve a Privacy Impact Assessment for the requested change.

Other primary data sources being considered include the FBI's Telephone Application, DHS data sources such as US-VISIT and SEVIS, Department of State data sources such as the Consular Consolidated Database (CCD), and Treasury Enforcement Communication System (TECS). Some of these sources will include very large amounts of data and funding has not yet been identified to complete their integration.

c. Does the FBI use IDW for "data mining?" If so, please describe the process, and indicate its effectiveness and reliability.

Response:

In its original statement, the FBI used the term "data mining" to be synonymous with "advanced analysis." The FBI does not conduct "data mining" in accordance with the GAO definition, which means mining through large volumes of data with the intention of automatically predicting future activities.

IDW allows for advanced analysis of large amounts of data, such as extracting all individuals from Suspicious Activity Reports and comparing the information against all individuals extracted from FBI terrorism investigations to look for overlap. All results are passed to FBI analysts for evaluation and further analysis. The FBI does not automatically generate predictions from IDW. Rather, it uses IDW to assist in identifying the most relevant elements of information that will allow trained analysts to make informed evaluations and predictions. This

approach saves analysts valuable time in gathering information from various sources, and has proven highly reliable.

d. Can other government agencies (federal, state or local) access IDW and if so, how?

Response:

Other government agencies can access IDW through their representatives to FBI Joint Terrorism Task Force (JTTF) members. JTTF members, including many federal, state, and local agencies, have been issued IDW accounts, and can access the system through any FBI computer connected to the FBI Intranet. These individuals must have completed background checks and been granted Top Secret clearances before they are granted access to FBI computers.

13. Do all FBI agents have access to the IDW on their desktops? If not, who has direct access to IDW? If agents do not have direct access, why not, and when can we expect them to have such access? Do you agree that it is important for the field agents to have access to all data at their fingertips in order to be able to react quickly in matters involving national security?

Response:

IDW is accessible from any FBI desktop; however, not all FBI agents have accounts. The Office of Intelligence Oversight Unit is responsible for evaluating user needs and prioritizing the creation of user accounts. Policy established by the Oversight Unit places priority on Field Intelligence Group members, and members of the Joint Terrorism Task Forces, in addition to the headquarters counterterrorism analysts that made up the initial user base. Since January 2004, IDW has issued more than 5,000 user accounts in accordance with the established policy.

The FBI agrees that it is important for field agents to have access to the data sets provided by IDW. The FBI intends to continue adding accounts and increasing the capability of the system accordingly; however, current funding does not support the provision of service to all FBI agents and analysts.

14. You also stated that the FBI can now do a "multi-word search" of data that is included in IDW. When was this capability made available through IDW? It is my understanding that these "multi-word searches" are still a long way from the type of multi-word searches that have become commonplace using the Internet or other search engines such as Lexis/Nexis or Westlaw. Thus, while the FBI can use multiple search terms like "flight school" and "lessons" to obtain some documents, it is my understanding that the FBI still

cannot find words within a certain defined parameter of one another. There may also be significant limitations when variations of spelling are used. Please explain in detail the types of searches of IDW that are currently available to FBI agents and any types of searches that are not currently available that you plan to add. Please include a timeline for any currently planned improvements to the search capability of your computer technology.

Response:

IDW included multi-word search ability when it was activated January of 2004. It provides greater search capability than that available through the Internet. Users can search for terms within a defined parameter of one another. For example, the search: 'flight school' NEAR/10 'lessons' would return all documents where the phrase "flight school" occurred within 10 words of the word "lessons." Users can also specify whether they want exact searches, or if they want the search tool to include other synonyms and spelling variants for words and names. Users can also combine all of these text search abilities with structured queries, such as limiting data by date ranges or FBI case classifications, within a single search.

IDW is also capable of extracting concepts such as names, phone numbers, and company names from unstructured text documents. This ability allows an IDW user the ability to perform concepts-related searches, rather than a list of documents. Users can then select concepts from the list, and browse through a series of related concepts that were extracted from the same document set. For example, a user could query information on a terrorist organization and retrieve a list of names extracted from documents about the terrorist organization. The user can then select a name from the list, and view a list of phone numbers extracted from the subset of documents that mention the selected name. At any point, the user can select a concept and view all related source documents for further analysis. This is a very powerful analytical method that is fundamentally different than standard search engines available through the Internet.

These capabilities are currently functional and available to all users. We are working on enhancing our ability to conduct multiple, large "batch queries." The example of advanced analysis provided in question 12(c), where the complete set of Suspicious Activity Reports is compared to the complete set of FBI terrorism files to identify individuals in common between them, is one type of "batch query."

15. The third phase of Trilogy – the Virtual Case File System, or VCF – was meant to replace the Automatic Case Support System (ACS). I took from your testimony that IDW is now adequately accessing ACS to ensure that all FBI information is capable of and is actually being mined for intelligence analysis and as an investigative tool. Many millions of

dollars have been spent in preparing for VCF and millions more will be spent to see that it is implemented.

a. Why is VCF still necessary if IDW and ACS are doing the job?

Response:

IDW addresses a subset of FBI investigative data while VCF, or its successor software, will provide access to all data resident in ACS. VCF and its successor software will provide enhanced workflow and case management functionality including the ability to search through various records, while that access is transparent to the user.

b. How (if at all) will VCF differ from IDW/ACS? In other words, will VCF be faster, easier, or more accessible to more agents and analysts? Will it have more sophisticated searching capabilities?

Response:

VCF, or its successor software, will far exceed the current ACS capabilities. It will essentially migrate the FBI from a "green screen" to a web interface, leaping several generations of technology. This capability will provide a faster and more user friendly interface for the agents and analysts. The greatly improved search capabilities will significantly improve their overall effectiveness and efficiency. VCF, or its successor software, also will contain a considerably larger repository of records than the IDW.

c. How is the continued delay of VCF's implementation adversely affecting the FBI's abilities?

Response:

The current paper-oriented workflow requires added time for data to be entered into the system of record, thereby delaying access to others. In addition, the lack of a search capability across records limits the FBI's ability to perform its intelligence and investigative functions. Despite the FBI's delay in implementing VCF, the FBI has achieved savings through the use of IDW.

d. The OIG noted in its September 2003 report that "unlike the currently used ACS system, agents will not be able to circumvent the use of the VCF." What do you understand that statement to mean and how does the ability of agents to circumvent ACS affect the IDW search engines?

Response:

Currently, the lack of controls with ACS prevents some users from submitting data in order to protect sources. VCF and its successor software will provide access controls that will require users to submit required data fields without later revealing critical source information to IDW users.

e. **The same September 2003 OIG report stated that with the release of VCF, agents will be provided with "content management capability" to "help agents access information from the FBI's data warehouse, regardless of where in the system the information was entered, [and] provide a single query for all of the FBI's systems that are connected to the Integrated Data Warehouse." Since VCF is still delayed, do the agents have this "content management capability" at this time and if not, when can we expect this capability to be in place?**

Response:

Agents do not currently have content management capability.

16. The OIG once described VCF as a "web-based 'point and click' case management system" through which "agents are expected to have multi-media capability that will allow them to scan documents, photos, and other electronic media into the case file." Am I correct that the FBI does not have that ability at present and that, therefore, scanned documents, photos and other electronic media are not accessible through the IDW at this time?

Response:

The FBI currently has the ability to make scanned documents and other electronic media available through the IDW.

VCF, or its successor software, will simplify the process of scanning documents and photos, and adding other electronic media into the case files, but it is still possible with current systems. Agents can use scanners provided by Trilogy, as well as the more robust services provided by the Document Conversion Laboratory (DOCLab) and Document Exploitation group (DocEx) to convert data into electronic form. Millions of these scanned documents have already been loaded into IDW and are available to users. In addition to scanned document libraries, the Violent Gang and Terrorist Organization File (VGTOF) library already has photographs imbedded with the electronic records and are accessible through IDW.

17. Earlier this year, with Senators Hatch, Grassley and Durbin, I asked the Government Accountability Office (GAO) to review the approximately \$600 million in costs attributed to the Trilogy system, which is still not in place. Can you assure me the FBI is fully cooperating with the GAO's audit, and doing so on a timely basis? Please explain what you are doing internally to ensure that the GAO is getting the materials it needs.

Response:

The FBI has and will continue to cooperate fully with the GAO auditors by providing timely, accurate, and complete information. Materials and information in response to GAO's requests have been provided. As an interim step to ensure the GAO is receiving the requested material in a timely fashion, in lieu of waiting until all material in response to a single request is available, the FBI will provide the information incrementally.

18. The September 2003 OIG report on Trilogy also commented upon the problems at the FBI regarding entry of foreign names into the FBI's existing databases (ACS) and explained that VCF would facilitate indexing on various web-based documents by providing data fields in searchable databases.

a. Does this mean, for example, that a VCF search of materials about Moammar "Gadhafi" will yield reports that spell the Libyan leader's name as Qaddafi, Qatafi, Quahthafi, Ghadafi, Kadafi or Kaddafi?

Response:

The VCF design included a wildcard search ability, but in its initial release would not have searched across name variants. In later releases, VCF was planning to incorporate Language Analysis Services (LAS), which has a robust name expansion utility to provide this service.

IDW has partially integrated LAS, and has already used it to support critical investigations, such as the 2003 holiday threat. This allowed IDW to expand a name into alternate spelling variants for comprehensive searching and analysis. This capability continues to be available to support special cases, and IDW plans to complete the integration and expose the name expansion capability to end users in a future release. Current funding, however, does not include this integration. At present, IDW allows users to manually create name expansion lists that would allow IDW to search across all identified variants. If LAS were fully integrated, users would have the option of manually creating a list, or using the automatic expansion provided by LAS.

b. Regarding IDW's capabilities as you described them in your testimony, are fundamental spelling issues still causing problems in search engines? Please explain how, if at all, VCF will rectify this situation.

Response:

IDW includes the ability to search across spelling variants for common words, synonyms and meaning variants for words, as well as common misspellings of words. If a user misspells a common word, IDW will run the search as specified, but will prompt the user to ask if they intended to run the search with the correct spelling. In addition, users can create a list of name variants they wish to use and IDW will search across all identified name variants. As mentioned in the question 18(a), it is anticipated that VCF (or its successor software) and IDW will incorporate the capabilities provided by LAS that would provide automatic expansion of name variants.

19. On April 8, 2004, the Subcommittee on Terrorism, Technology and Homeland Security of the Senate Judiciary Committee held a hearing on "Keeping America's Mass Transportation System Safe: Are the Laws Adequate?" At that time, I posed a written question to the Amtrak representatives about whether or not rail police have direct access to law enforcement records systems while performing pedestrian and vehicle investigations. A copy of Amtrak's response is attached as Exhibit A to these Written Questions. Please provide your position on the legislative proposal suggested by Amtrak in which rail police that are certified and commissioned law enforcement officers would be provided equal footing with state and local law enforcement for purposes of access to criminal history data.

Response:

28 U.S.C. § 534(4)(d)(1) authorizes the Attorney General to exchange records and information with railroad police departments which perform the administration of criminal justice, have arrest powers pursuant to a state statute, allocate a substantial part of their budget to the administration of criminal justice (defined in 28 C.F.R. Part 20, Subpart A), and meet the training requirements established by law or ordinance for law enforcement officers.

Under this authority, upon request, the FBI assigns Originating Agency Identifiers (ORIs) to railroad police departments meeting the criteria of 28 CFR Part 20. A National Crime Information Center (NCIC) ORI is a nine-character alpha-numeric identifier assigned to authorized agencies, permitting access to the NCIC Interstate Identification Index (III). Amtrak has been assigned eight ORIs that permit access to NCIC/III for criminal justice purposes.

FINAL

FOR OFFICIAL USE ONLY

FEDERAL BUREAU OF INVESTIGATION
PROJECT MANAGEMENT OFFICE

ALL INFORMATION CONTAINED
HEREIN IS UNCLASSIFIED
DATE 08-01-2007 BY 65179 DMH/EJA/CAL
1058805



SYSTEM

CONCEPT OF OPERATIONS (CONOPS)

INVESTIGATIVE DATA WAREHOUSE (IDW)

FOR OFFICIAL USE ONLY

26 March 2004
Version 3

FINAL

FOR OFFICIAL USE ONLY

EXECUTIVE SUMMARY

The Investigative Data Warehouse (IDW) provides FBI users with the capability to view, query, search, retrieve, correlate, integrate, synthesize, share, and protect information from multiple data sources in support of intelligence and investigative activities. As a single point of entry for accessing both FBI data and non-FBI data, IDW provides FBI users with information needed to successfully accomplish the FBI's counter-crime, counter-intelligence, and counter-terrorism missions.

This Concept of Operations (CONOPS) documents IDW as an evolving family of systems that will provide near- and long-term operational and developmental capabilities to the FBI. When fully deployed, IDW will include four (4) systems:

- The IDW Secret-level operational system (IDW-S) consists of those builds which have undergone appropriate security and operational testing and have been approved by the responsible FBI authorities for operational use. IDW-S V1.0 received Interim Authority to Operate (IATO) on January 23, 2004 and began operations for approved users over FBI Net on January 25, 2004. IDW-S V2 is currently being developed.
- The IDW Integration system (IDW-I) is a Secret-level representation of IDW-S that serves as an environment in which maintenance fixes and proposed new capabilities can be realistically tested before being released into IDW-S.
- IDW-TS/SCI is a version of IDW-S that, when built, will be approved for data that is classified as Top Secret and/or Sensitive Compartmented Information (TS/SCI). It should be noted that because the IDW Program has given high priority to IDW-S, the IDW-TS/SCI system is currently in the definition stage.
- The IDW Development system (IDW-D) is an Unclassified prototyping environment used to facilitate experimentation with proposed new IDW technologies.

This initial IDW CONOPS is focused on IDW-S and IDW-I, the two IDW systems currently developed. As noted above, IDW-S operates under an IATO, whereas IDW-I will operate under an Interim Authority to Test (IATT). This is appropriate to the role of IDW-S as an operational system with a general user base and to the intended role of IDW-I as a test environment. This CONOPS identifies all major IDW system processes and internal and external interfaces. It provides an overview of the IDW-S conceptual design and a high-level description of IDW system requirements. This IDW CONOPS is intended to complement other IDW Program documentation, in particular the IDW Program Management Plan, the IDW-S System Security Plan, and the Target IDW/Virtual Case File (VCF) Business Architecture.

FINAL

FOR OFFICIAL USE ONLY

SECTION 3

DESCRIPTION OF THE IDW PROJECT

Pertinent details regarding the IDW project are:

Project Name: Investigative Data Warehouse (IDW)

Account Identification Code: 15-0200-0-1-999,

Project Initiation Date: March 2003,

Project Planned Completion Date: December 2006 (As the Master Data Warehouse¹).

As a single point of entry for accessing investigative data sources, IDW provides FBI users with the capability to readily acquire, store, share, use, disseminate, and protect the information needed to successfully accomplish their assignments and the FBI's overlapping missions in intelligence, counter-terrorism, and criminal investigations.

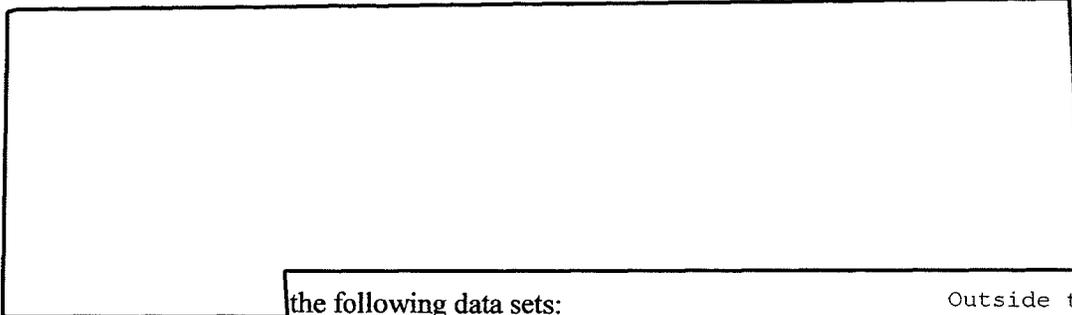
The IDW system environment consists of a collection of UNIX and NT servers that provide secure access to a family of very large-scale storage devices. The servers provide application, web servers, relational database servers, and security filtering servers. User desktop units that have access to FBI Net can access the IDW web application. This provides browser-based access to the central databases and their access control units. The environment is configured so that the FBI analytic and investigative users can access any of the data sources and analytic capabilities of the system for which they are authorized. The entire configuration is scalable to enable expansion as more data sources and capabilities are added.

The FBI currently owns or has access to over 30 information technology systems and well over 100 enterprise level applications that support investigative functions. At the user level, the number of databases containing case-centric intelligence is estimated to be in the thousands, a number that has increased largely due to the lack of an enterprise-wide application for data analysis. The IDW project initiative will ultimately integrate many of the underlying system data sources into a single Investigative Data Warehouse that will support data mining and target searching of both FBI data and data from external sources. The project will also support the selective sharing of data with other Federal agencies as part of the Department of Homeland Security's (DHS) Horizontal Information Sharing Initiative and the Joint Terrorism Task Forces (JTTFs). The data warehouse capability will permit the abundance of this investigative data to be shared on an FBI-wide basis, providing a complete data picture to analysts and agents.

¹ The Master Data Warehouse is the next generation investigative warehouse which expands the analytical tool capabilities and includes administrative data sets so that the FBI can adequately evaluate return on investment in applying resources to investigative programs.

FINAL

FOR OFFICIAL USE ONLY



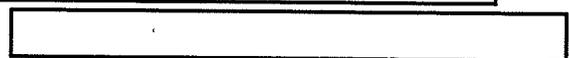
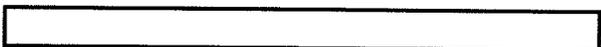
the following data sets:

Outside the Scope

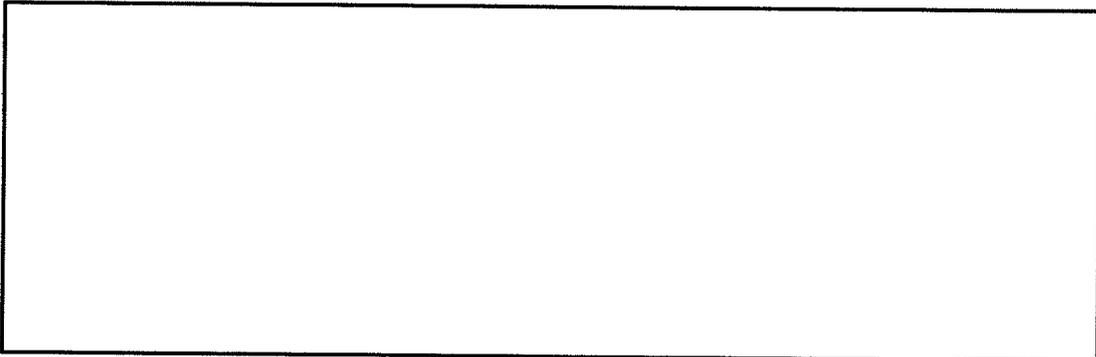
- Approved case files from the FBI's Automated Case Support (ACS) case management system,
- Electronic versions of the Joint Intelligence Committee Inquiry (JICI) archived documents,
- Secure Automated Messaging Network (SAMNet) message traffic;
- IntelPlus File Rooms (IDW V1.0 does not currently update this information),
- Violent Gang and Terrorist Organization File (VGTOF) data from the Criminal Justice Information Systems (CJIS) Division (IDW V1.0 does not currently update this information).



following additional databases and/or data sources:

-  - Data provided from FINCEN system
- 
- 
- Virtual Case File (VCF)
- 

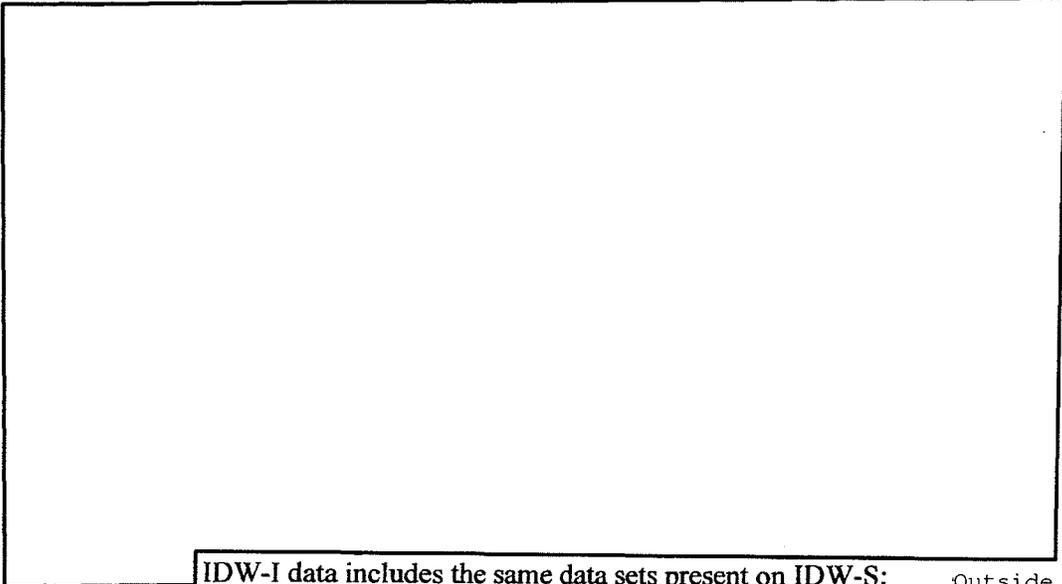
b2
b7E



Outside the Scope

FINAL

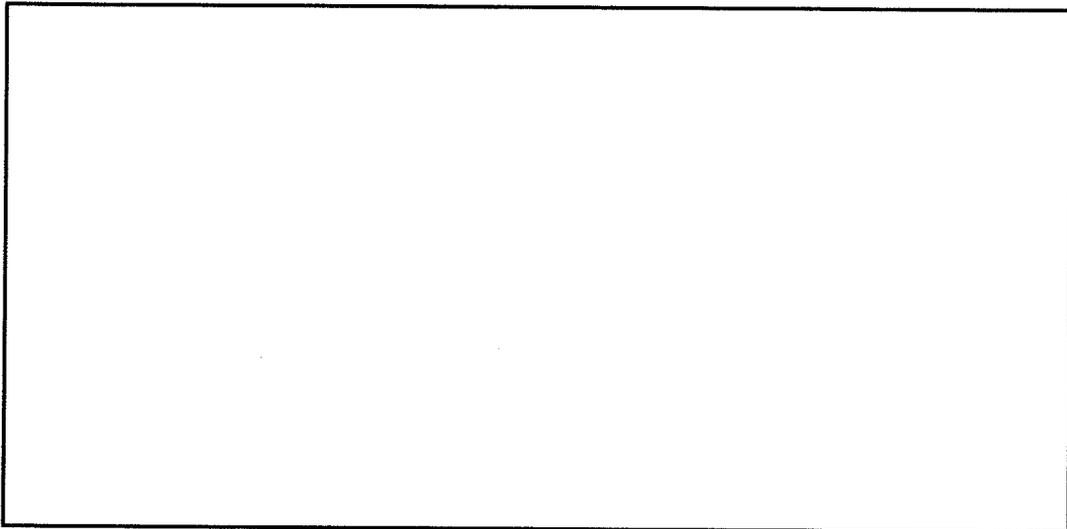
FOR OFFICIAL USE ONLY



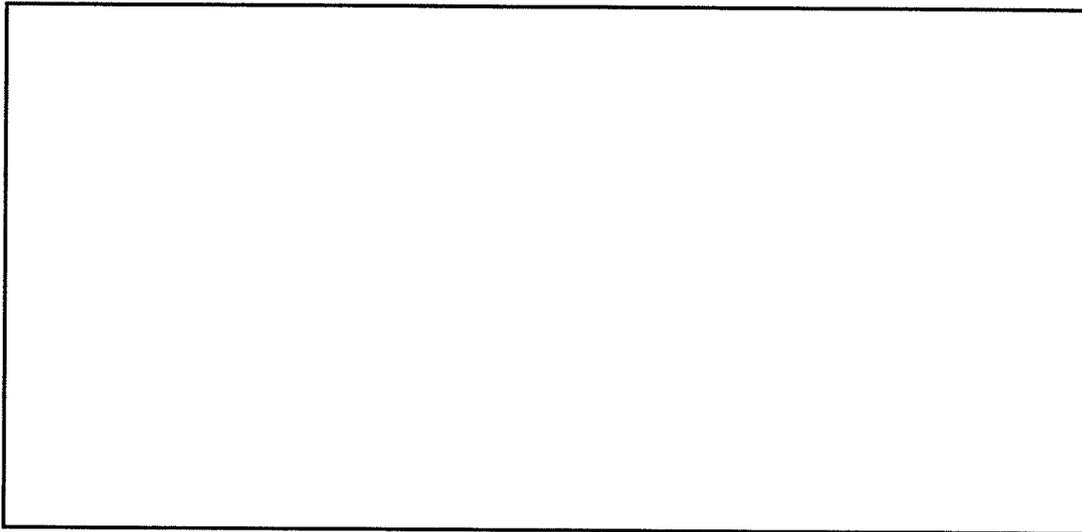
IDW-I data includes the same data sets present on IDW-S:

Outside the Scope

- Approved case files from the FBI's ACS case management system,
- Electronic versions of the JICI defined archived documents,
- SAMNet message traffic;
- IntelPlus File Rooms,
- VGTOF data from the CJIS Division,
- Translingual Information Detection, Extraction and Summarization (TIDES)
Program Open Source News Data



Outside the Scope



Outside the Scope

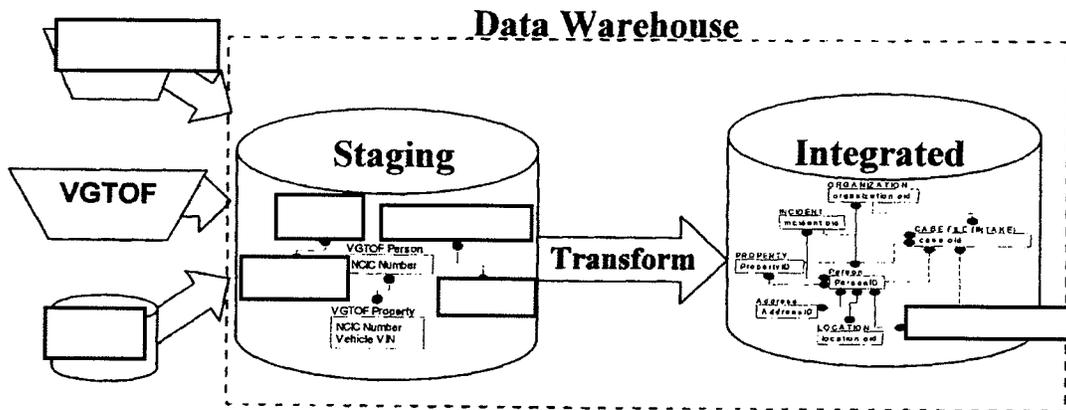
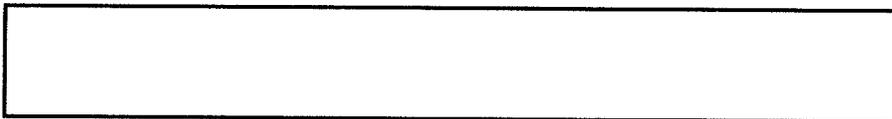


Figure 5-2 Data Ingest Data Transformation Process

The ingested data is transformed from a source structure to a staging structure to provide more efficient searches and database organization (see Figure 5-2). During the transformation process, the data obtained from multiple sources is integrated so that relationships between the several data elements of the original data source can be established. The new structure provides the basis for analysis by the BI tools. In addition to the transformation, the Data Ingest analyzes the data quality and will maintain this metric for manual use by the analysts and automated use by the analysis tools.



Outside the Scope

Outside the Scope

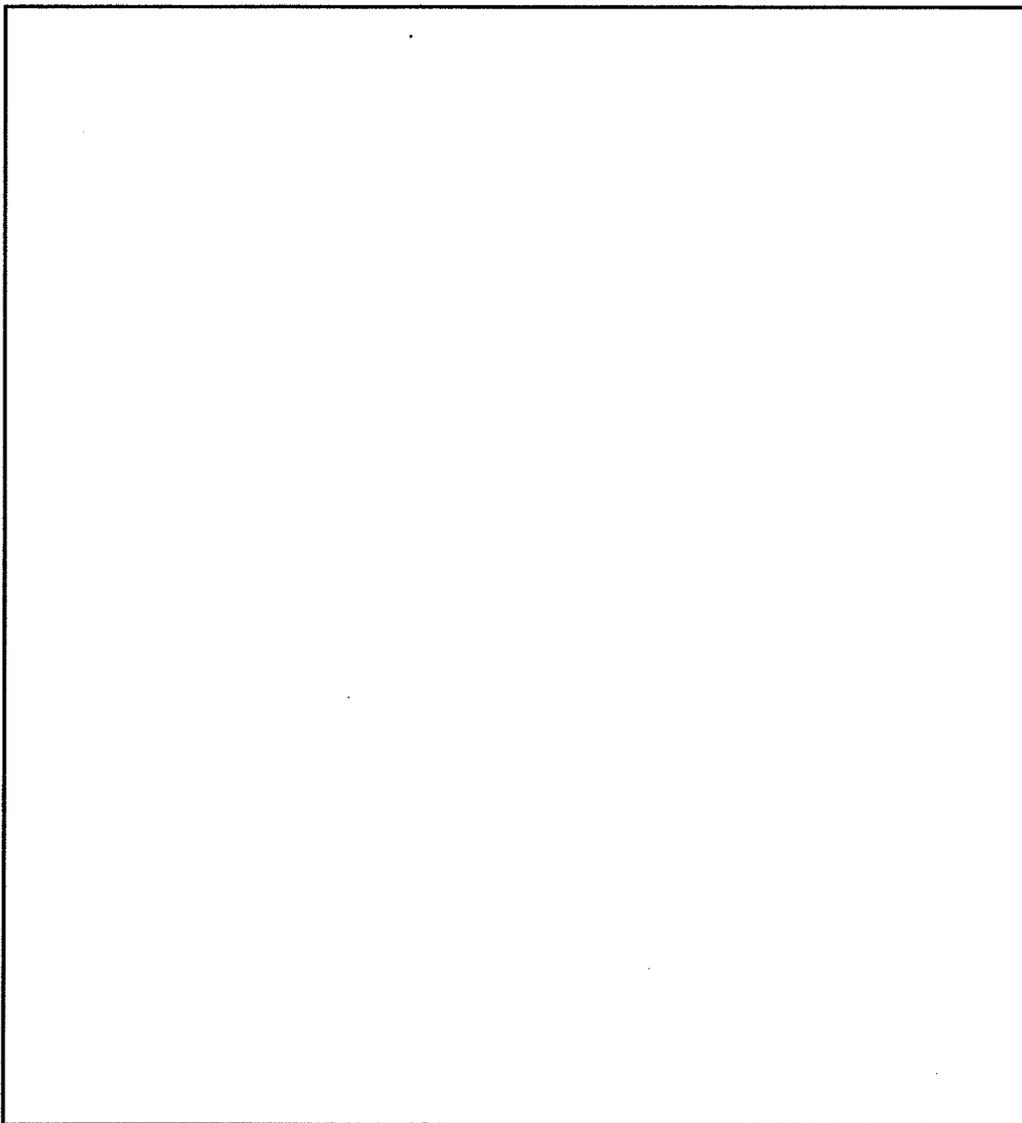


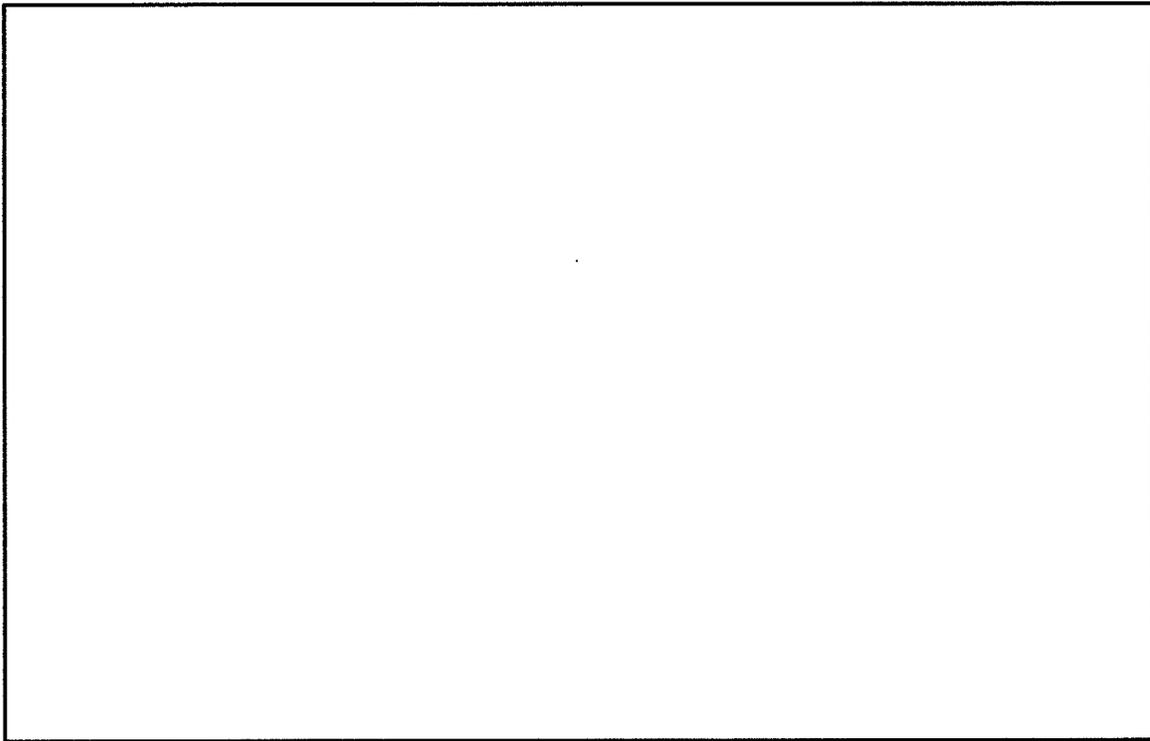
The SDMA component of the IDW-S system will process and manage all structured data by providing the following major processes:

- Store and manage structured data from all external and legacy systems (IDW V2 will initially include VCF, VGTOF, [redacted] UNI, and [redacted] and other future data sources. This data will be provided to SDMA by Data Ingest.
- Store and manage unstructured data as part of the IDW Data Store.

b2
b7E

Outside the Scope



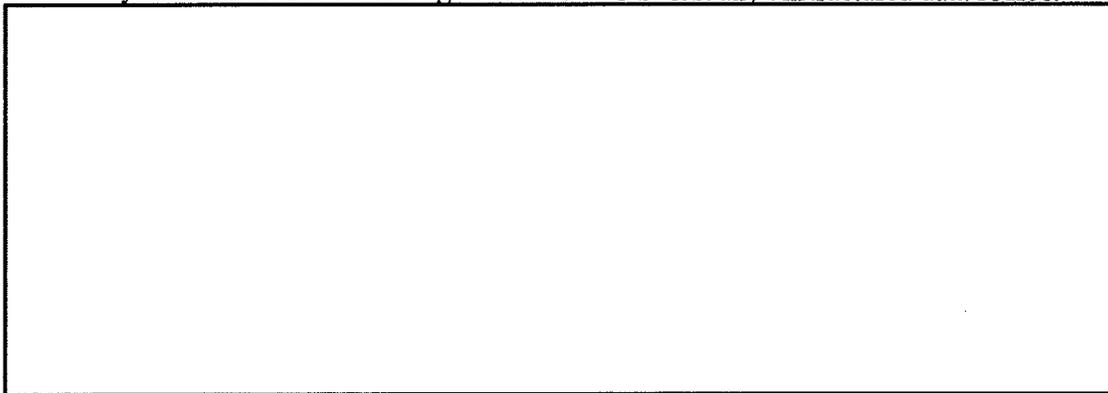


Outside the Scope

6.3 Manage and Analyze Unstructured Data

The unstructured data management and analysis subsystem (UDMA) provides functionality for storing, indexing, searching, and extracting information from unstructured information. This unstructured information is documented and associated with the document's metadata. The initial data sources for IDW UDMA are: ACS/VCF, INTEL+/JICI, SAMNET, and data, however, except for some key metadata that are captured for these sources there is nothing unique about the unstructured data handled by this subsystem. UDMA will be designed to work with most any unstructured data source.

b2
b7E

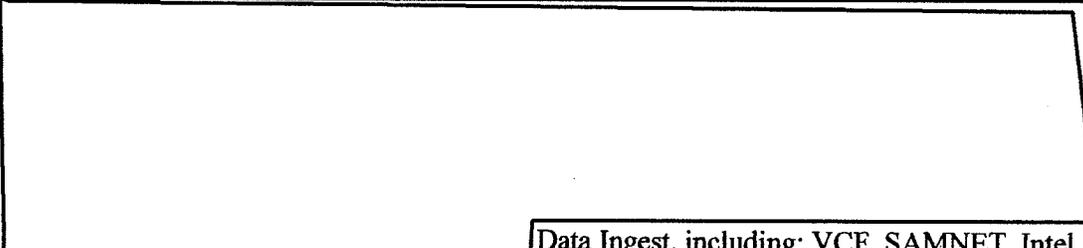
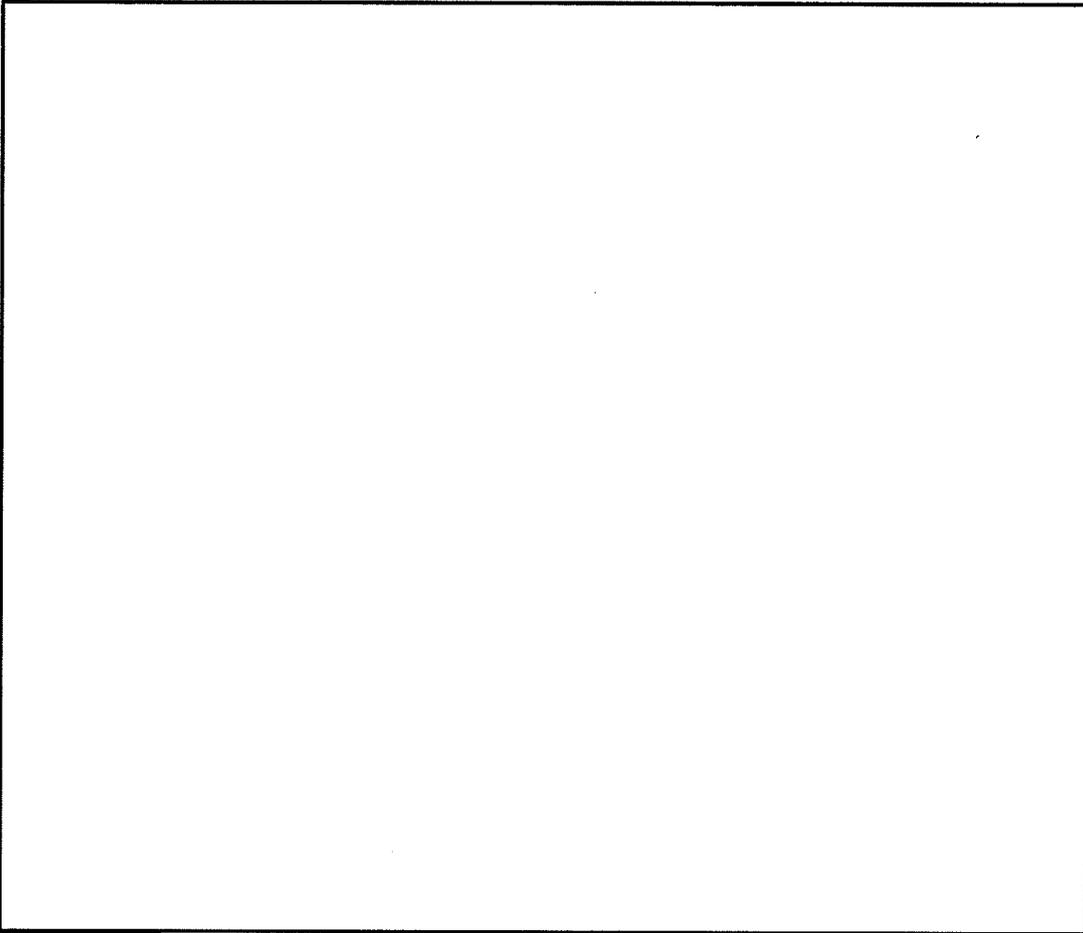


Outside the Scope

FINAL

FOR OFFICIAL USE ONLY

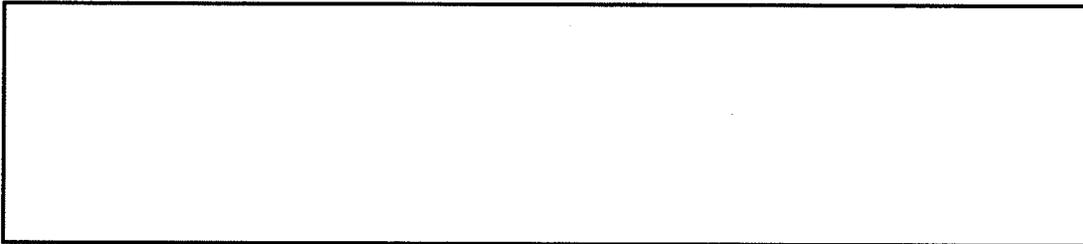
Outside the Scope



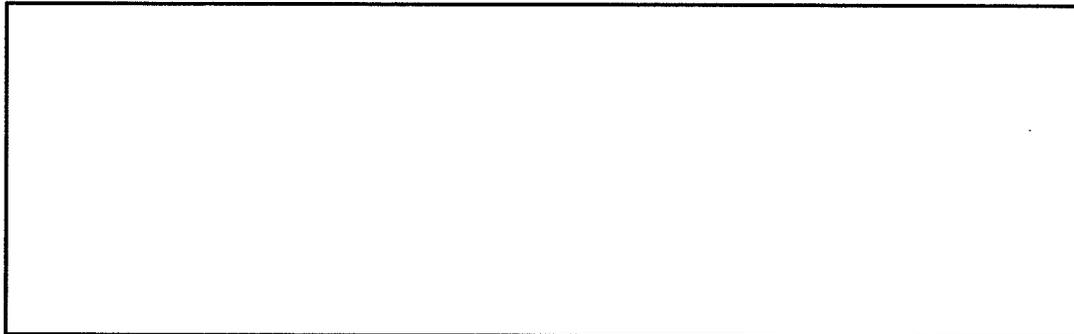
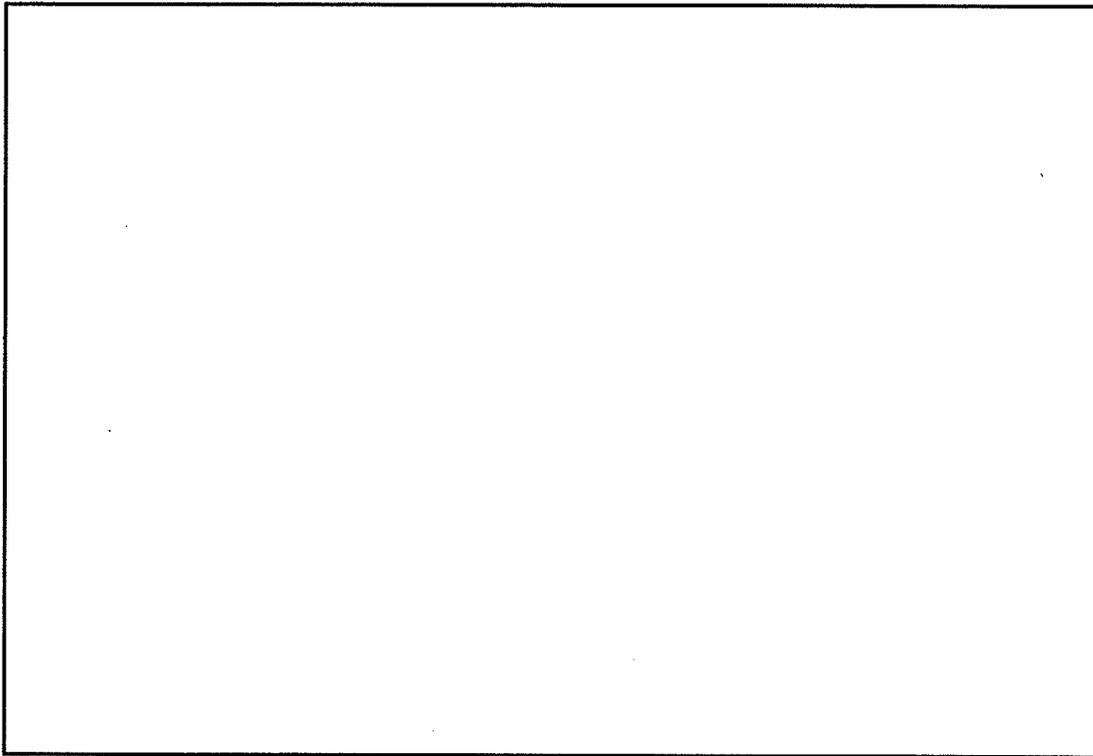
b2
b7E

Plus, JICI, and 

Data Ingest, including: VCF, SAMNET, Intel



Outside the Scope



6.10 Data Sources

The IDW V2 system will provide one web-based interface to the user thereby allowing access to any of nine data sources with an access control system applying to all of the data sources in order to conduct global searching and analysis. These sources include four that are structured ([redacted] VGTOF [redacted] and [redacted] and five that are unstructured (VCF, Intel+, JICI, SAMNet [redacted] Versions of IDW beyond V2 will support the processing from additional data sources.

b2
b7E

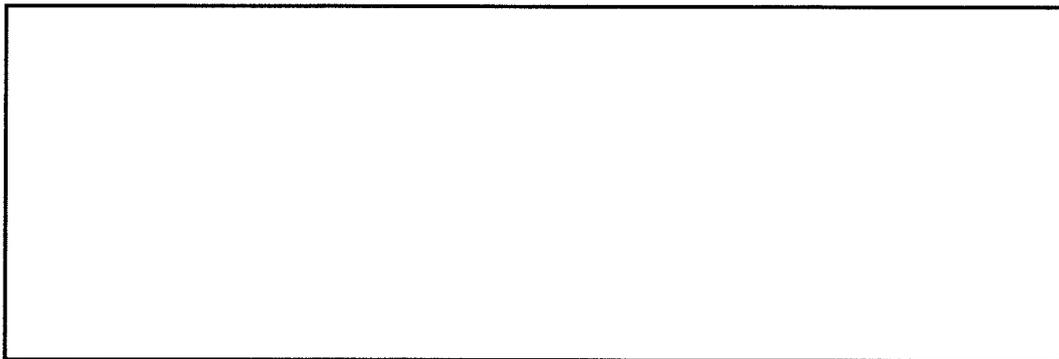
- [Redacted]
- [Redacted]
- [Redacted]
- [Redacted]

The users have requested that IDW system support searches of the following databases. Some of these are currently not funded:

- Legacy Systems
 - SAMNet-S
 - JICI
 - ACS or VCF (since VCF is planned to replace ACS)
 - VGTOF
 - [Redacted]
 - [Redacted]
 - [Redacted]
 - IntelPlus
 - FINCEN - [Redacted]

b2
b7E

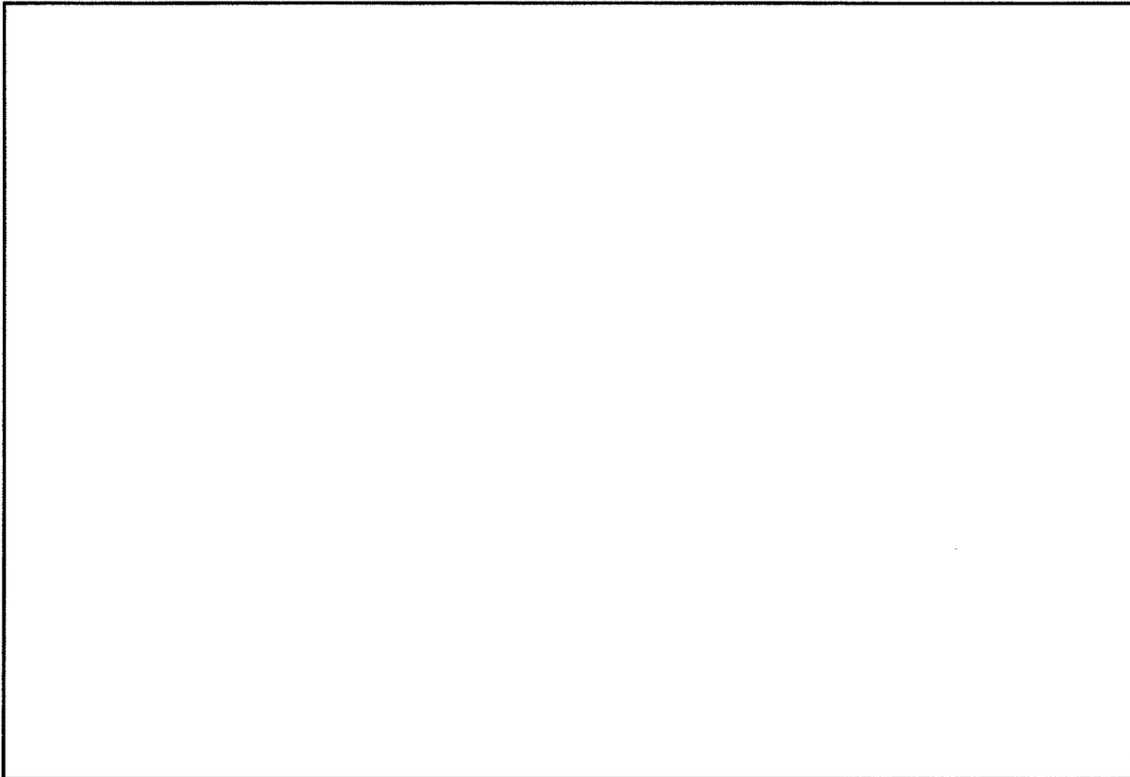
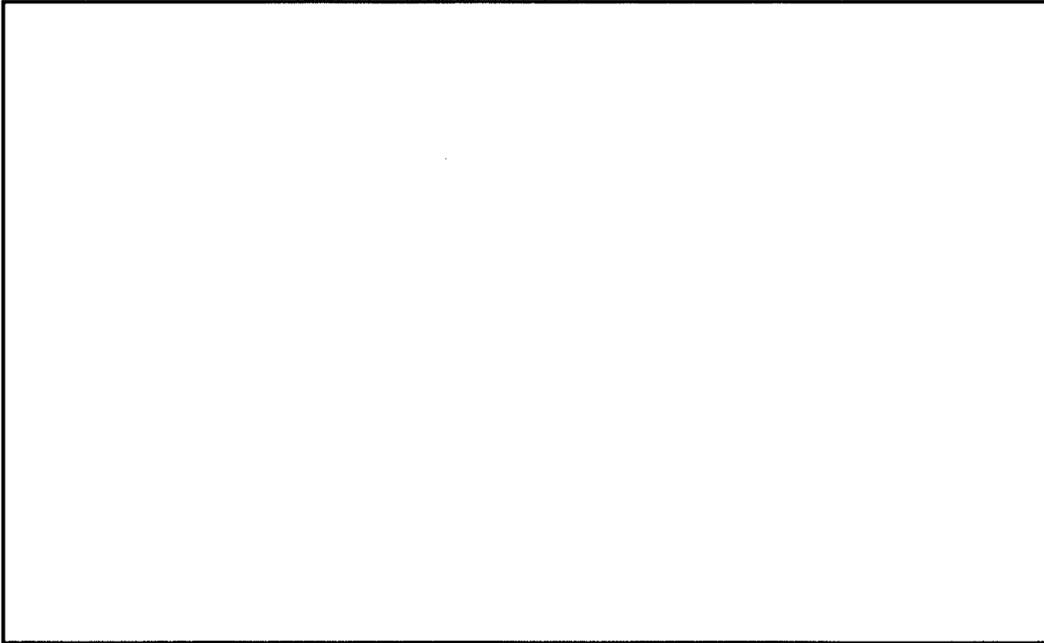
Outside the Scope



FINAL

FOR OFFICIAL USE ONLY

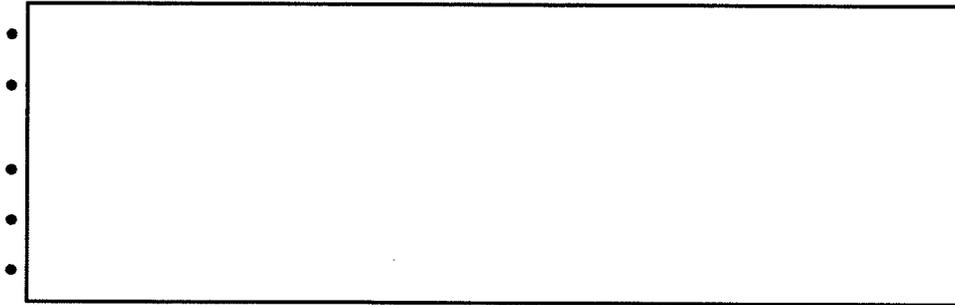
Outside the Scope



b2
b7E
b5

FINAL

FOR OFFICIAL USE ONLY



b2
b7E
b5

FINAL

FOR OFFICIAL USE ONLY

APPENDIX B
Document References

- Department of Justice (DOJ) System Development Life Cycle (SDLC) Guidance Document
- SCOPE Concept of Operations, September 2002.
- Department of Justice, Federal Bureau of Investigation, FY2004 Exhibit 300 Capital Asset Plan, Secure Counter-terrorism Operational Prototype Environment (SCOPE) Investigative Data Warehouse (IDW), July, 2002 version.
- *"FBI Data Warehousing, Data Mining & Collaboration: An Enterprise View of Data"* a public briefing 5/30/03. Mr. Kenneth Ritchhart, Section Chief, Data Engineering & Integration Program Management Office.
- System Requirements Document, Investigative Data Warehouse, version 9, dated March 24, 2004
- Investigative Data Warehouse Business CONOPS, version 1.0, dated February 2, 2004
- System Security Plan (SSP) for the Investigative data Warehouse – Secret (IDW-S), Version 0.9, dated January 8, 2004
- System Engineering Management Plan (SEMP), Final Draft, version 1.3, dated March 17, 2004
- Unstructured Data Management and Analysis Subsystem Design Document, draft document dated February 17, 2004
- Structured Data Management and Analysis Subsystem Design Document, draft document, version 1.2, dated February 25, 2004
- Data Ingest Subsystem High-Level Design Specification, Version 0.4, dated February 4, 2004

APPENDIX D
Data Source Descriptions

D.1 VGTOF: Violent Gang and Terrorist Organizations File

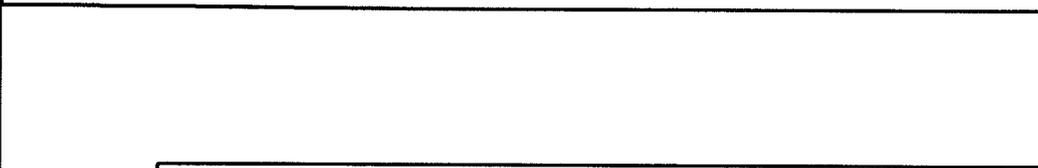
Since September 11, 2001, Director Mueller has directed field offices of the FBI to place the subjects of open terrorism related investigations into the FBI's Terrorism Watch List which is part of the Violent Gangs and Terrorist Organization File (VGTOF) maintained by the National Criminal Information Center (NCIC). The Terrorism Watch List is currently the Counterterrorism Division's integrated listing of lone terrorists, or terrorist groups, of investigative interest to the FBI.

The subjects of counterterrorism investigations are being added to the file daily and are accessed by other Federal, State and local law enforcement agencies whenever these agencies access the system for the purpose of running criminal history checks on individuals of interest to their own investigations (i.e., during routine traffic stops). When accessed by an officer, an application used with the database is capable of automatically notifying the officer that the name is of interest to the FBI and should be treated with caution. The system can provide further instructions such as requesting the officer to notify the FBI of the reason for the inquiry.

The purpose of the data base is to share pertinent biographical information with other Federal, State and local law enforcement agencies for officer safety and mutual investigative interest.

The Terrorism Watch List (VGTOF) is in the process of being consolidated into a single data base managed by the Terrorist Threat Integration Center (TTIC) and the recently announced Terrorist Screening Center (TSC).

D.2



The application is also used as an analytical tool and a source for intelligence information.

b2
b7E

FOR OFFICIAL USE ONLY

D.3 [redacted]

[redacted] is an investigative tool that also serves as the central repository for [redacted] data obtained throughout the course of an FBI investigation, to include [redacted]. The [redacted] source data is in part obtained from FBI operations.

b2
b7E

D.4 [redacted]

This is actually a dual program [redacted] serving as one repository.

[redacted]

b2
b7E

provides the ability [redacted] within various criteria and provides statistical reports. [redacted] is the repository for [redacted] provided to the FBI by FINCEN.

D.5 VCF: Virtual Case File

This will be the central repository for FBI investigations. It is currently under development and is expected to become operational by the end of 2004. The VCF system will be a structured central database supporting investigative activities enterprise-wide. The database will contain relevant data to all cases opened for investigation including court files and related law-enforcement information from state and local field office sources. Due to security and access control challenges and programmatic constraints, current plans are for the IDW system to use a subset of the entire VCF content. This subset of VCF content will be consistent with the case classification restrictions, and other (Federal Grand Jury, Federal Taxpayer Information, Bank Secrecy Act Information) restrictions which are currently on ACS documents being copied into the IDW.

D.6. SAMNET: Secure Automated Message Network

This system is used to transmit and receive messages from the Intelligence Community and other agencies. SAMNET is also used by Legat Offices, and Field and Headquarters Divisions to exchange messages up to the TS/SCI level. The system is being modernized to include a migration to the Defense Message System (DMS), and replacement of an existing manual method of printing and delivering paper, with electronic delivery to the

appropriate desktop, based on functional profile. SAMNET information can be from DoD and other Intelligence Community sources.

D.7 JICI: Joint Intelligence Community Inquiry

This a static collection of anti-terrorist files collected from FBI field office files offices following 9/11. The collection represents a historical record of field office files and is not currently updated on a regular basis. The files are searched for target words usually in conjunction with other associated databases.

D.8 IntelPlus: Intelligence Plus

This is an application which allows the users to view "Table of Contents" lists from large collections of records in various formats. The user is able to display the document whether it is in text form or one of several graphic formats and then print, copy or store the information. The application allows researchers in tracking associated documents by assisting the user on going to related topics and provides a convenient search capability. The Intel Plus application is currently organized around six separate counterterrorism collections:

-
-
-
-
-
-
-

b2
b7E

D.9 [Redacted]

[Redacted]

The service is available to government organizations.

Table 2 Data Ingest Initiated Transaction

Item #	Ingest Initiated Transaction	2004	2005	2006	2007	2008	Record Size (average)	Reference HLF
1	Ingest data from external source							REQT30
2	Ingest data from ACS/MCF							REQT30.1
3	Ingest data from VGTOF							REQT30.2
4	Ingest data from SAMNET-S							REQT30.3
5	Ingest data from [redacted]							REQT30.4
6	Ingest data from [redacted]							REQT30.5
7	Ingest data from [redacted]							REQT30.6
8	Ingest data from IntelPlus							REQT30.7
9	Ingest data from [redacted] from the field							REQT30.8
10	Ingest data from [redacted]							REQT30.9
11	Ingest data from JICI							REQT30.11
12	Ingest data from [redacted]							REQT30.12
13	Ingest data from [redacted]							REQT30.13
14	Ingest data from [redacted]							REQT30.14
15	Ingest data from [redacted]							REQT30.15
16	Ingest data from [redacted]							REQT30.16

b2
b7E

Table 2 Ingest Initiated Transaction

Item #	Ingest Initiated Transaction	2004	2005	2006	2007	2008	Record Size (average)	Reference HLFR
17	Ingest data from [redacted]							REQT30.17
18	Ingest data from [redacted]							REQT30.18
19	Ingest data from [redacted]							REQT30.19
20	Ingest data from [redacted]							REQT30.20
21	Ingest data from [redacted]							REQT30.21
22	Ingest data from [redacted]							REQT30.22
23	Ingest data from [redacted]							REQT30.23
24	Ingest data from [redacted]							REQT30.24

b2
b7E

Table 2 Ingest Initiated Transaction

Item #	Ingest Initiated Transaction	2004	2005	2006	2007	2008	Record Size (average)	Reference HLFR
25	Ingest data from FinCen - [redacted] Information.							REQT30.25
26	Ingest selectively from FinCen Data							REQT65.2
27	Ingest selectively [redacted]							REQT65.3
28	Ingest selectively [redacted]							REQT65.4
29	Ingest selectively [redacted]							REQT65.5
30	Ingest selectively from [redacted]							REQT67
31	Ingest selectively from [redacted]							REQT67.1
32	Ingest selectively from [redacted]							REQT67.2
33	Ingest selectively from [redacted]							REQT67.3
34	Ingest selectively from [redacted]							REQT67.4

b2
b7E

Table 2 Ingest Initiated Transaction

Item #	Ingest Initiated Transaction	2004	2005	2006	2007	2008	Record Size (average)	Reference HLFR
35	Ingest selectively from [redacted]							REQT67.5
36	Ingest selectively from [redacted]							REQT67.6
37	Ingest selectively from [redacted]							REQT67.7
38	Ingest selectively from [redacted]							REQT67.8
39	Ingest selectively on an ad hoc basis, data from open sources.							REQT69
40	Ingest selectively from the [redacted]							REQT69.1
41	Ingest selectively from [redacted]							REQT69.2
42	Ingest selectively from [redacted] data.							REQT69.3
43	Ingest selectively from [redacted]							REQT69.4
44	Ingest selectively from [redacted]							REQT69.5

b2
b7E

Table 2 Ingest Initiated Transaction

Item #	Ingest Initiated Transaction	2004	2005	2006	2007	2008	Record Size (average)	Reference HLF
45	Ingest selectively from [redacted]							REQT69.6
46	Ingest selectively from [redacted]							REQT69.7
47	Ingest selectively from [redacted]							REQT69.8
48	Ingest selectively from [redacted]							REQT69.9
49	Ingest selectively from data about [redacted]							REQT69.10

b2
b7E