U.S. Department of Justice

Federal Bureau of Investigation

Washington, D.C. 20535

SEP 28 2007

MARCIA HOFMANN ESQ
ELECTRONIC FRONTIER FOUNDATION
454 SHOTWELL STREET
SAN FRANCISCO, CA 941110

Subject: INVESTIGATIVE DATA WAREHOUSE

FOIPA No. 1058805- 000

Dear Requester:

The enclosed documents were reviewed under the Freedom of Information/Privacy Acts (FOIPA), Title 5, United States Code, Section 552/552a. Deletions have been made to protect information which is exempt from disclosure, with the appropriate exemptions noted on the page next to the excision. In addition, a deleted page information sheet was inserted in the file to indicate where pages were withheld entirely. The exemptions used to withhold information are marked *below and explained on the enclosed Form OPCA-16a:*

Section 552		**Section 552a**
☒(b)(1)	☒(b)(7)(A)	☐(d)(5)
☒(b)(2)	☐(b)(7)(B)	☐(j)(2)
☐(b)(3)_____	☒(b)(7)(C)	☐(k)(1)
_____	☐(b)(7)(D)	☐(k)(2)
_____	☒(b)(7)(E)	☐(k)(3)
_____	☐(b)(7)(F)	☐(k)(4)
☐(b)(4)	☐(b)(8)	☐(k)(5)
☐(b)(5)	☐(b)(9)	☐(k)(6)
☒(b)(6)		☐(k)(7)

204 **page(s)** were reviewed and 169 **page(s)** are being released.

☐ Document(s) were located which originated with, or contained information concerning other Government agency(ies) [OGA]. This information has been:

- ☐ referred to the OGA for review and direct response to you.
- ☐ referred to the OGA for consultation. The FBI will correspond with you regarding this information when the consultation is finished.

☐ You have the right to appeal any denials in this release. Appeals should be directed in writing to the Director, Office of Information and Privacy, U.S. Department of Justice,1425 New York Ave., NW, Suite 11050, Washington, D.C. 20530-0001 within sixty days from the date of this letter. The envelope and the letter should be clearly marked "Freedom of Information Appeal" or "Information Appeal." Please cite the FOIPA number assigned to your request so that it may be easily identified.

☐ The enclosed material is from the main investigative file(s) in which the subject(s) of your request was the focus of the investigation. Our search located additional references, in files relating to other individuals, or matters, which may or may not be about your subject(s). Our experience has shown, when ident, references usually contain information similar to the information processed in the main file(s). Because of our significant backlog, we have given priority to processing only the main investigative file(s).

If you want the references, you must submit a separate request for them in writing, and they will be reviewed at a later date, as time and resources permit.

☒ See additional information which follows.

Sincerely yours,

David M. Hardy
Section Chief
Record/Information
Dissemination Section
Records Management Division

Enclosure(s)

The enclosed documents represent the first of a series of interim releases that will be made with regard to your request pertaining to the Investigative Data Warehouse.

For your information, some of the enclosed pages contain information about other subject matters. We consider this information "outside the scope" of your request, and have redacted the information and marked it "o/s".

EXPLANATION OF EXEMPTIONS

SUBSECTIONS OF TITLE 5, UNITED STATES CODE, SECTION 552

(b)(1) (A) specifically authorized under criteria established by an Executive order to be kept secret in the interest of national defense or foreign policy and (B) are in fact properly classified to such Executive order;

(b)(2) related solely to the internal personnel rules and practices of an agency;

(b)(3) specifically exempted from disclosure by statute (other than section 552b of this title), provided that such statute(A) requires that the matters be withheld from the public in such a manner as to leave no discretion on issue, or (B) establishes particular criteria for withholding or refers to particular types of matters to be withheld;

(b)(4) trade secrets and commercial or financial information obtained from a person and privileged or confidential;

(b)(5) inter-agency or intra-agency memorandums or letters which would not be available by law to a party other than an agency in litigation with the agency;

(b)(6) personnel and medical files and similar files the disclosure of which would constitute a clearly unwarranted invasion of personal privacy;

(b)(7) records or information compiled for law enforcement purposes, but only to the extent that the production of such law enforcement records or information (A) could be reasonably be expected to interfere with enforcement proceedings, (B) would deprive a person of a right to a fair trial or an impartial adjudication, (C) could be reasonably expected to constitute an unwarranted invasion of personal privacy, (D) could reasonably be expected to disclose the identity of confidential source, including a State, local, or foreign agency or authority or any private institution which furnished information on a confidential basis, and, in the case of record or information compiled by a criminal law enforcement authority in the course of a criminal investigation, or by an agency conducting a lawful national security intelligence investigation, information furnished by a confidential source, (E) would disclose techniques and procedures for law enforcement investigations or prosecutions, or would disclose guidelines for law enforcement investigations or prosecutions if such disclosure could reasonably be expected to risk circumvention of the law, or (F) could reasonably be expected to endanger the life or physical safety of any individual;

(b)(8) contained in or related to examination, operating, or condition reports prepared by, on behalf of, or for the use of an agency responsible for the regulation or supervision of financial institutions; or

(b)(9) geological and geophysical information and data, including maps, concerning wells.

SUBSECTIONS OF TITLE 5, UNITED STATES CODE, SECTION 552a

(d)(5) information compiled in reasonable anticipation of a civil action proceeding;

(j)(2) material reporting investigative efforts pertaining to the enforcement of criminal law including efforts to prevent, control, or reduce crime or apprehend criminals;

(k)(1) information which is currently and properly classified pursuant to an Executive order in the interest of the national defense or foreign policy, for example, information involving intelligence sources or methods;

(k)(2) investigatory material compiled for law enforcement purposes, other than criminal, which did not result in loss of a right, benefit or privilege under Federal programs, or which would identify a source who furnished information pursuant to a promise that his/her identity would be held in confidence;

(k)(3) material maintained in connection with providing protective services to the President of the United States or any other individual pursuant to the authority of Title 18, United States Code, Section 3056;

(k)(4) required by statute to be maintained and used solely as statistical records;

(k)(5) investigatory material compiled solely for the purpose of determining suitability, eligibility, or qualifications for Federal civilian employment or for access to classified information, the disclosure of which would reveal the identity of the person who furnished information pursuant to a promise that his/her identity would be held in confidence;

(k)(6) testing or examination material used to determine individual qualifications for appointment or promotion in Federal Government service the release of which would compromise the testing or examination process;

(k)(7) material used to determine potential for promotion in the armed services, the disclosure of which would reveal the identity of the person who furnished the material pursuant to a promise that his/her identity would be held in confidence.

Freedom of Information and Privacy Acts

FOIPA # 1058805

Subject: Investigative Data Warehouse

File Number:

Section: CID

Federal Bureau of Investigation

From: [redacted]CID)(FBI)
Sent: Thursday, February 24, 2005 3:43 PM
To: [redacted]CID) (FBI)
Subject: FW: IDW searches - restricted files

b6
b7C

SENSITIVE BUT UNCLASSIFIED
NON-RECORD

[redacted]asked me to send him an official e-mail. This is just FYI, in case there was a systemic problem, and not just a mistake about the one file.

b6
b7C

Hope you're enjoying the snow :)

[redacted]

-----Original Message-----
From [redacted]CID)(FBI)
Sent: Thursday, February 24, 2005 3:39 PM
To: [redacted](ITSD)(FBI)
Subject: IDW searches - restricted files

b6
b7C

SENSITIVE BUT UNCLASSIFIED
NON-RECORD

Dea[redacted]
My name is [redacted]and I'm an intelligence analyst currently receiving IDW [redacted]training during the Quantico ACES I course. While learning the system, I pulled up a public corruption file which I understand should be restricted from general access. The idw_doc_id is [redacted] The file classification is [redacted]194-0, and because the Data Sources page states that only files 194A-Z are restricted, I wondered if that accounted for the release of the file for general access. If so, you may need to add the zero files to your filter.

b6
b7C
b2

Thank you very much for your time, and please let me know if you have any questions.

[redacted]

SENSITIVE BUT UNCLASSIFIED

SENSITIVE BUT UNCLASSIFIED

From: [redacted] (ITSD)(CON)
Sent: Friday, March 18, 2005 7:22 AM
To: [redacted]
Subject: RE: IDW searches - restricted files -phone number

b6
b7C

SENSITIVE BUT UNCLASSIFIED
NON-RECORD

Excellent.....and that is what we all live for.....deadines (as opposed to Deadwood)

-----Original Message-----
From: [redacted] (CID)(FBI)
Sent: Thursday, March 17, 2005 3:47 PM
To: [redacted] (ITSD)(CON)
Subject: RE: IDW searches - restricted files -phone number

b6
b7C

SENSITIVE BUT UNCLASSIFIED
NON-RECORD

I have, and I got in just fine. I'll let you know about the other when I get a chance - deadlines are such fun!

Thanks again,
[redacted]

-----Original Message-----
From: [redacted] (ITSD)(CON)
Sent: Thursday, March 17, 2005 2:30 PM
To: [redacted] (CID)(FBI)
Subject: RE: IDW searches - restricted files -phone number

b6
b7C

SENSITIVE BUT UNCLASSIFIED
NON-RECORD

[redacted]

We have removed the file, thanx again for coming by to help make sure we got the right one. It is SO nice to get some user feedback directly.

Two other questions.....Have you tried to log back into IDW since you got back and Is there anything else I can do to help you work the issue of additional files you wanted to include in IDW?

b6
b7C

Incase you need some other names.....FBI PM [redacted] and the IDW O&M Lead [redacted]

Thanx again
R
[redacted]

-----Original Message-----
From: [redacted] (CID)(FBI)
Sent: Monday, March 14, 2005 11:10 AM
To: [redacted] ITSD)(CON)
Subject: RE: IDW searches - restricted files -phone number

b6
b7C

SENSITIVE BUT UNCLASSIFIED

NON-RECORD

b6
b7C

I'm sorry I didn't respond before (it was my last week at Quantico, and we turned in our access cards) - I'm back at HQ now, and I no longer have access to IDW. What kind of info do you need?

My number here is 202-324-

Thanks,

-----Original Message-----
From: (ITSD)(CON)
Sent: Thursday, March 10, 2005 10:56 AM
To: (CID)(FBI)
Subject: FW: IDW searches - restricted files -phone number

b6
b7C

SENSITIVE BUT UNCLASSIFIED
NON-RECORD

If you tried to call on the 202 number, please dont leave a message.....somehow we cant retrieve any of them

R

-----Original Message-----
From: (ITSD)(CON)
Sent: Wednesday, March 09, 2005 8:12 AM
To CID)(FBI)
Cc (ITSD)(FBI)
Subject: FW: IDW searches - restricted files

b6
b7C

SENSITIVE BUT UNCLASSIFIED
NON-RECORD

My name is and I am the ISSO for the IDW system. Thank you for sending the information on this to us. I searched for the document you mentioned below, and couldn't find it. Could you please send me some more information from your search so we can locate and remove it?

b6
b7C

I can also be reached either 202-324- or (cell)

Thanx again

-----Original Message-----
From ITSD)(FBI)
Sent: Wednesday, March 09, 2005 7:39 AM
To ITSD)(CON)
Subject: FW: IDW searches - restricted files

b6
b7C

SENSITIVE BUT UNCLASSIFIED
NON-RECORD

[redacted]
IDW Lead for O&M, SPT, and DE
202-[redacted]
[redacted]P)
-----Original Message-----
From: [redacted]CID)(FBI)
Sent: Thursday, February 24, 2005 3:39 PM
To: [redacted](ITSD)(FBI)
Subject: IDW searches - restricted files

SENSITIVE BUT UNCLASSIFIED
NON-RECORD

Dear [redacted]
My name is [redacted] and I'm an intelligence analyst currently receiving IDW/Chiliad training during the Quantico ACES I course. While learning the system, I pulled up a public corruption file which I understand should be restricted from general access. The idw_doc_id is [redacted] The file classification is [redacted] 94-0, and because the Data Sources page states that only files 194A-Z are restricted, I wondered if that accounted for the release of the file for general access. If so, you may need to add the zero files to your filter.

b6
b7C
b2

Thank you very much for your time, and please let me know if you have any questions.

[redacted]

SENSITIVE BUT UNCLASSIFIED

SENSITIVE BUT UNCLASSIFIED

SENSITIVE BUT UNCLASSIFIED

SENSITIVE BUT UNCLASSIFIED

SENSITIVE BUT UNCLASSIFIED

Freedom of Information and Privacy Acts

FOIPA # 1058805

Subject: Investigative Data Warehouse

File Number:

Section: ITSD Processes

Federal Bureau of Investigation

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

b2
b7E

b2
b7E

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

b2
b7E

The procedure for secure deletion of individual files 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] 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. b2 b7E
- [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. b2 b7E
- [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 o[redacted] is [redacted] b2 b7E
- [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.

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. b2 b7E
- 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.

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.

[redacted]

b2
b7E

Audit reports can be customized and provided upon request.

Freedom of Information and Privacy Acts

FOIPA # 1058805

Subject: Investigative Data Warehouse

File Number:

Section: ITSD Audit

Federal Bureau of Investigation

IDW Data Contention and Audit Inventory for 2006

315C	1 and 4	ACS		11-Sep-06	2006-11SEP-01A	Provided Audit for these case id's on 28Sept06
315C	38 and 39	ACS		11-Sep-06	2006-11SEP-01A	Provided Audit for these case id's on 28Sept06
		ACS		21-Aug-06	2006-21AUG-01	Removed from Quarantine
66F		ALL		10-Aug-06	2006-10AUG-01A	No Results for requested audit.
315C	ALL	ACS		21-Jul-06	2006-21JUL-01	Removed all docs from collection on 21JUL06.
n/a		ALL		15-Jul-06	2006-15JUL-01A	Provided Audit for this issue.
315H	ALL	ALL		21-Jun-06	2006-21JUN-01A	Provided audit for specific documents.
100 311A	24	ACS		9-Jun-06	2006-09JUN-01	Removed this document on 12JUN06
n/a		ALL		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.
(S) [redacted]		[redacted]	27-Apr-06	2006-27APR-01 & 01A	Removed 7 documents on 28APR06 and provided requested audit on 01MAY06.
		[redacted]	27-Apr-06	2006-27APR-01 & 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

Freedom of Information and Privacy Acts

FOIPA # 1058805

Subject: Investigative Data Warehouse

File Number:

Section: OCA

Federal Bureau of Investigation

FEDERAL BUREAU OF INVESTIGATION
FOIPA
DELETED PAGE INFORMATION SHEET

Serial Description ~ COVER SHEET

Total Deleted Page(s) ~ 29
Page 2 ~
O/S
Page 4 ~
O/S
Page 6 ~
O/S
Page 9 ~
O/S
Page 10 ~
O/S
Page 11 ~
O/S
Page 12 ~
O/S
Page 13 ~
O/S
Page 14 ~
O/S
Page 15 ~
O/S
Page 16 ~
O/S
Page 18 ~
O/S
Page 19 ~
O/S
Page 21 ~
O/S
Page 23 ~
O/S
Page 24 ~
O/S
Page 26 ~
O/S
Page 27 ~
O/S
Page 28 ~
O/S
Page 29 ~
O/S
Page 31 ~
O/S
Page 32 ~
O/S

XXXXXXXXXXXXXXXXXXXXXXX
X Deleted Page(s) X
X No Duplication Fee X
X for this Page X
XXXXXXXXXXXXXXXXXXXXXXX

Page 33 ~
O/S
Page 34 ~
O/S
Page 35 ~
O/S
Page 50 ~
O/S
Page 51 ~
O/S
Page 52 ~
O/S
Page 53 ~
O/S

XXXXXXXXXXXXXXXXXXXXXXX
X Deleted Page(s) X
X No Duplication Fee X
X for this Page X
XXXXXXXXXXXXXXXXXXXXXXX

Congressional Affairs Office
Congressional Contacts

Date Entered: 05/21/2004 | (•) Briefing ( ) Hearing ( ) Other

2004-736 **Event Date:** 5/12/2004

b6
b7C

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:

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

DOJ Notification: None **DOJ Date/Time:**

b6
b7C

FBI Participants: CIO Zalmai Azmi (Briefer), AD Eleni Kalisch, SSA (OCIO)

Other Participants:

Committees /Subcommittees: Senate Select Committee on Intelligence

Members/Staff:

b6
b7C

Details of Briefing:

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

OTHER

O/S

Follow Up Action:

None

109th

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

CAO Contact Person:

b6 b7C **DOJ Notification:** **DOJ Date/Time:**

FBI Participants: Zal Azmi (ACS demo)

Other Participants:

Committees /Subcommittees: House Appropraitions

b6 b7C **Members/Staff:**

Details of Briefing:

The staff were provided a demo and briefing on IDW and ACS. 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. 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/iniatives.

b6 b7C OTHER O/S

Follow Up Action:

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:

b2
b6
b7C
b7E

CAO Contact Person: SSA

DOJ Notification: **DOJ Date/Time:** 1:00:00 PM

FBI Participants: SC Mike Morehart (TFOS) (TFOS, observer)

Other Participants:

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

Members/Staff:

Details of Briefing:

b6
b7C
OTHER O/S
b2
b7E

Follow Up Action:

b6
b7C
OTHER O/S

Congressional Affairs Office
Congressional Contacts

Date Entered: 10/03/2005 | (•) Briefing ( ) Hearing ( ) Other ( ) FOC

OTHER O/S

2005-366 Event Date: 8/26/2005

Subject: IDW

b6 b7C

CAO Contact Person: SSA

DOJ Notification: DOJ Date/Time:

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

Other Participants:

Committees /Subcommittees: Senate Appropriations

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 thorugh 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:

b6
b7C
OTHER O/S **DOJ Notification:** **DOJ Date/Time:**

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 availbility; weaknesses and improvements needed; data composition; cooperation with outside agencies and DNI; intelligence products; Beta version; batch queries; training; financial resources.

OTHER O/S

Follow Up Action:

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:

CAO Contact Person: SSA [redacted] b6 b7C OTHER O/S

DOJ Notification: **DOJ Date/Time:**

FBI Participants: None

Other Participants:

Committees /Subcommittees: Senate Banking, Housing and Urban Affairs

Members/Staff: Shelby, Hagel, Martinez, Allard

Details of Briefing:

[redacted] b2 b6 b7C b7E 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.

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.

Freedom of Information and Privacy Acts

FOIPA # 1058805

Subject: Investigative Data Warehouse

File Number:

Section: FD

Federal Bureau of Investigation

FEDERAL BUREAU OF INVESTIGATION
FOIPA
DELETED PAGE INFORMATION SHEET

Serial Description ~ COVER SHEET

Total Deleted Page(s) ~ 4
Page 4 ~ b2, b7E
Page 5 ~ b2, b7E
Page 19 ~ b2, b7E
Page 20 ~ b2, b7E

XXXXXXXXXXXXXXXXXXXXXXX
X Deleted Page(s) X
X No Duplication Fee X
X for this Page X
XXXXXXXXXXXXXXXXXXXXXXX

Federal Bureau of Investigation
Response to Investigative Data Warehouse (IDW) Press Article for Senate Appropriations Committee
September 7, 2006

There are two concerns being expressed about IDW in the article. One deals with whether the FBI has complied with the Privacy Act's requirement to publish a "systems notice" in the Federal Register and the other is whether the FBI has complied with the privacy impact analysis requirements of the "E-Government Act."

The answer to the first question is "yes." We consider IDW to be part of the FBI's Central Record System, an "umbrella" system that is comprised of all of the FBI's investigative files. While it is true that "IDW" isn't specifically mentioned in the CRS Privacy Act System Notice, we don't believe that is necessary. The system notice does state: "In recent years . . . the FBI has been confronted with increasingly complicated cases, which require more intricate information processing capabilities. Since these complicated investigations frequently involve massive volumes of evidence and other investigative information, the FBI uses its computers, when necessary to collate, analyze, and retrieve investigative information in the most accurate and expeditious manner possible." The system notice describes in reasonable detail what information we obtain, what routine uses we make of it, the authorities for maintaining the system and so forth. This notice is published in the Federal Register and is publicly available. In our view, we are compliant with both the letter and spirit of the Privacy Act in this regard.

The answer to the second question is also "yes." In fact, since IDW has been categorized as a "national security system," the E-Government Act does not require it to undergo a privacy impact analysis (PIA) at all. Even so, FBI and DOJ policy requires a PIA to be conducted. For IDW, the FBI has done several PIA's. We did one for the original system and did others as significant data sets were added to IDW. None of these systems were published since the law does not require them to be conducted in the first place. The point is that we have done far more to analyze the privacy implications of IDW than the law requires. Yes, the analyses have not been conducted in the public domain but Congress weighed the costs and benefits of conducting such an analysis in public and chose to exclude national security systems from that requirement when it passed the E-Government act.

Investigative Data Warehouse Program Overview

Presented by [redacted] b6 b7C

Overview

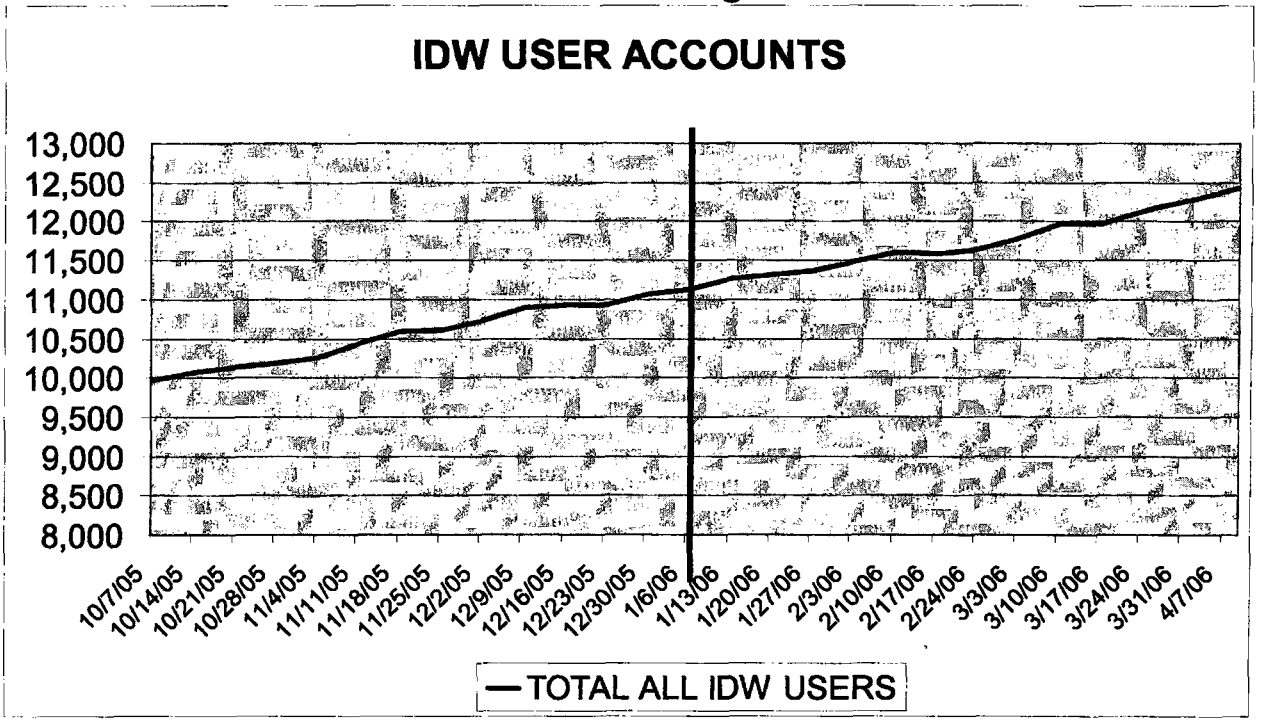
- Access tools and data through standard desktop computer, FBI Intranet, & a browser session
- Interactive search services
  - Search based on free-form entered phrase
  - Search for subject terms such as name, DOB, SSN, telephone number
  - Search on metadata such as case number, document date, author, etc.
  - Capability to refine searches & set of search results
- Batch search services – search for subject terms
- More than 12,000 users – including Joint Terrorism Task Force (JTTF) members
- Collection of 53 government multi-source datasets including data from FBI, CIA, DOS, DHS, FinCEN
- Authority to Hold Data:
  - OGC and Privacy Impact Assessments (PIA) done for all data
  - Information Sharing Policy Board (ISPB) approves all data
  - Audit and Security Log Protocols are maintained across all dat

Background

- Program started in FY02
- Program goals in FY03 OMB 300 Exhibit
  - Data from 40 sources
  - 5,000 users
  - 100 – 200 M records
  - 3-5 second response time for interactive searches
- Program achievements
  - Data from 53 sources
  - 12,432 user accounts
  - Data from ~ 500 M documents
  - 3-8 second median response time for interactive searches
  - Added Batch Search service
  - Special Project Team provided services to 5 task forces or operations
- Authority to Operate through 2008 for Integration/Test & Operational Environments
- Authority to Hold Data
  - OGC and Privacy Impact Assessments (PIA) done for all data
  - Information Sharing Policy Board (ISPB) approves all data

Growth in User Accounts

12,432 total users as of 4/12/06
All Field Offices & 23 Legal Attache Offices

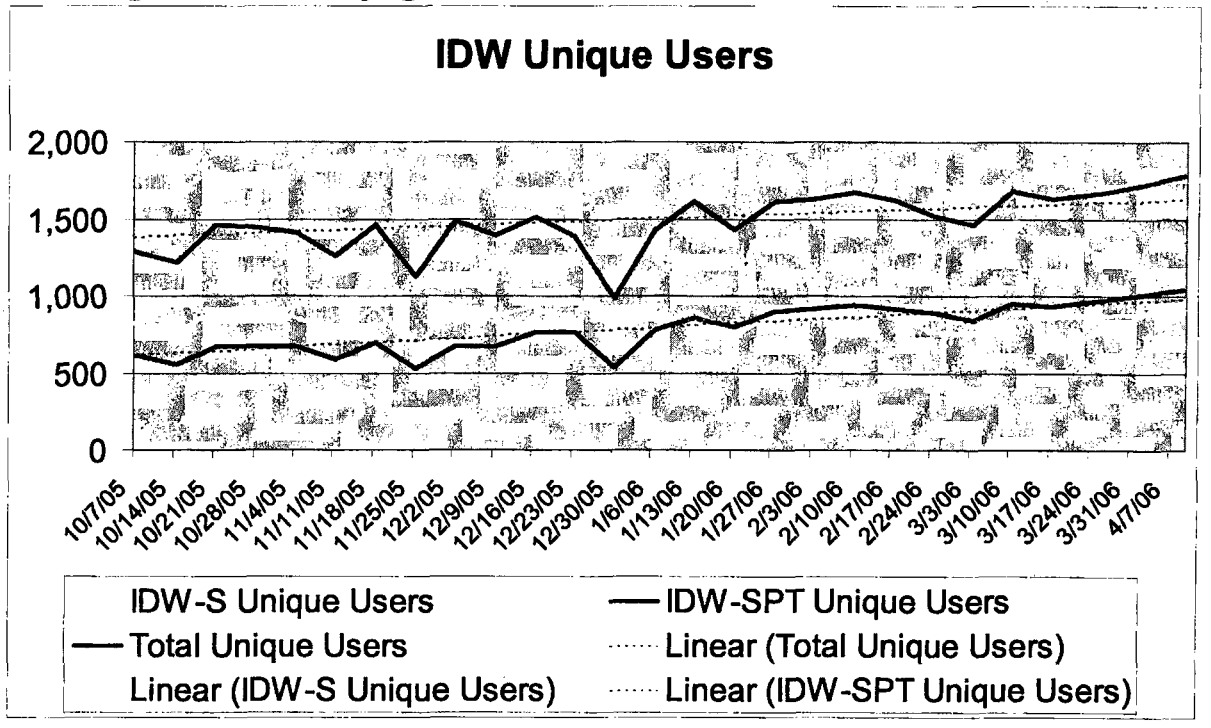


Users

Breakdown of IDW Accounts Account Description	No. of Accts as of 02/22/2006	Percent
FBI Agent	3,858	32%
FBI Analyst/Support	3,044	25%
Federal Agency (Non-FBI)	1,541	13%
State Agency	313	3%
Local Agency	975	8%
Other LE & Analyst users	2,248	19%
Total	11,979	100%

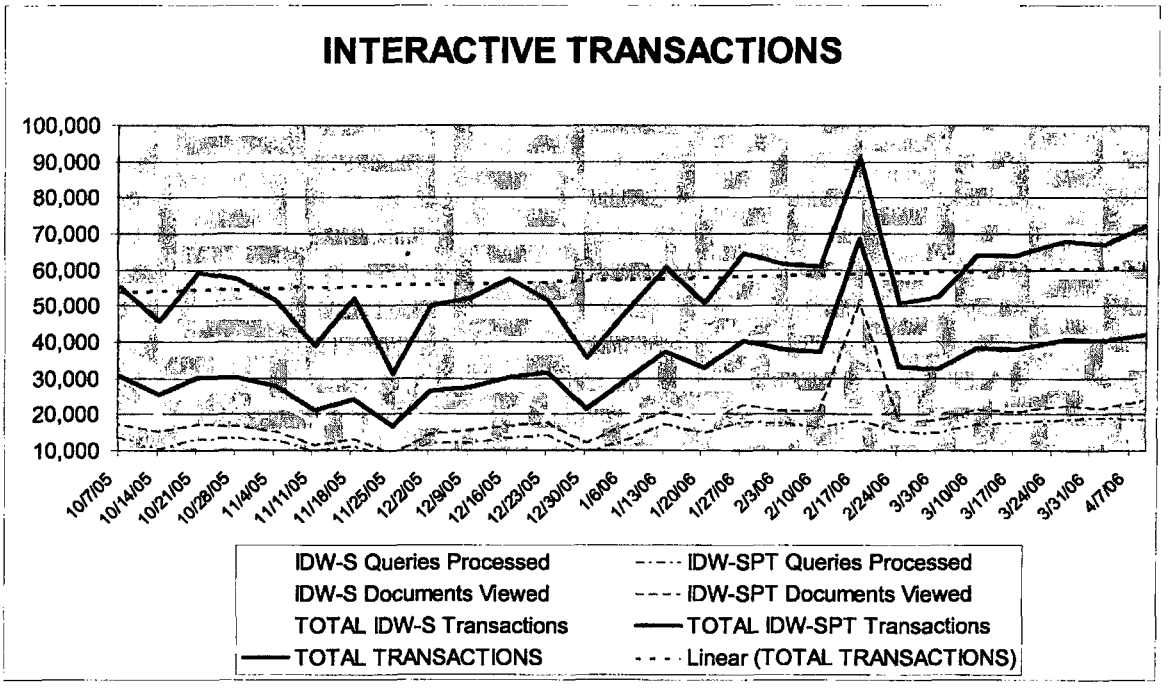
IDW Unique Users per Week

- Between ~ 1,200 & ~ 1,600 unique users sign on in any given week



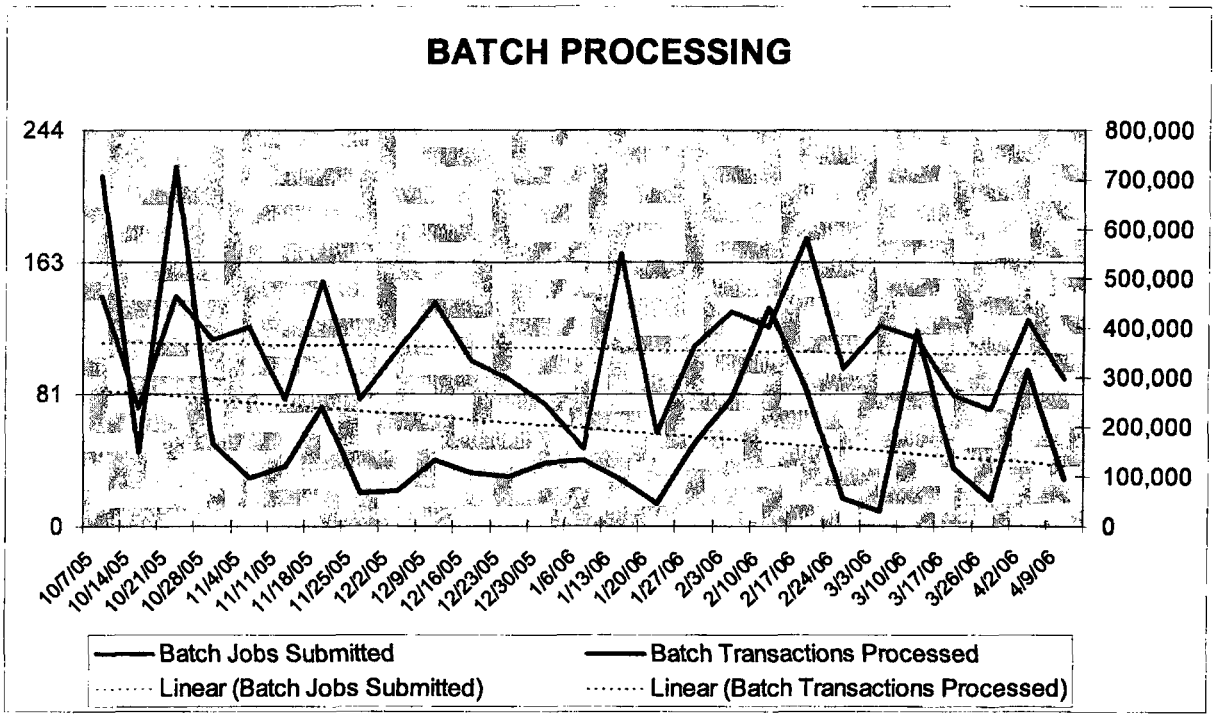
Interactive Transactions (Weekly)

- Between ~ 40,000 & ~ 60,000 interactive transactions in any given week

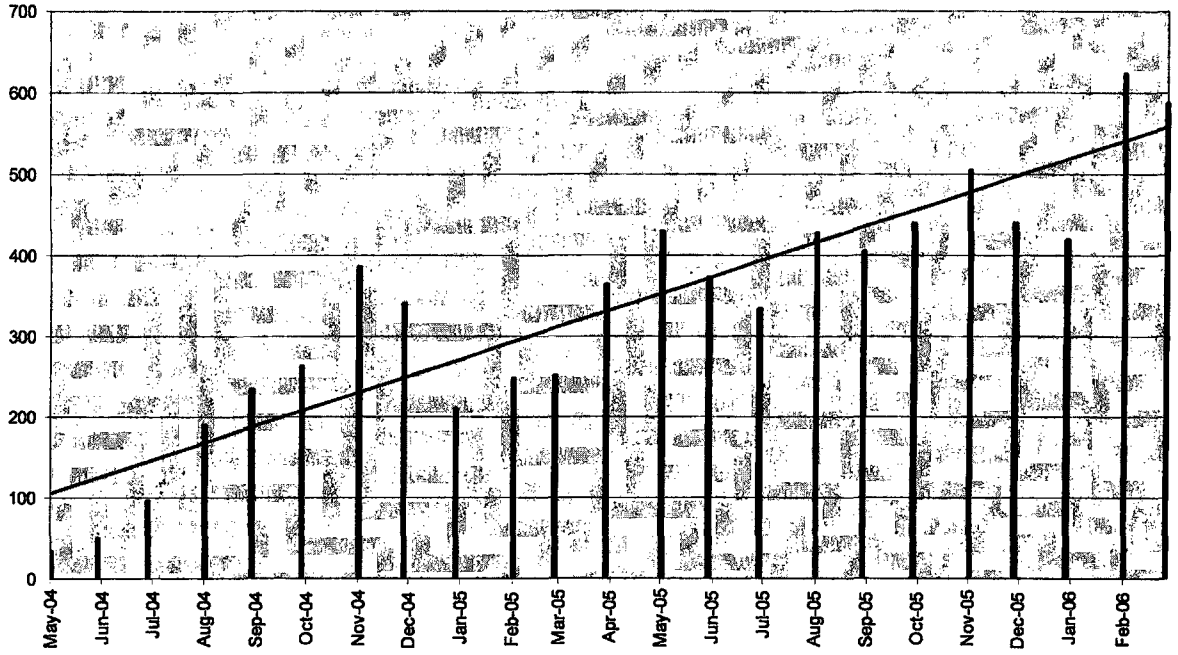


Batch Processing

- Between ~ 50 & ~ 150 batch jobs submitted in any given week



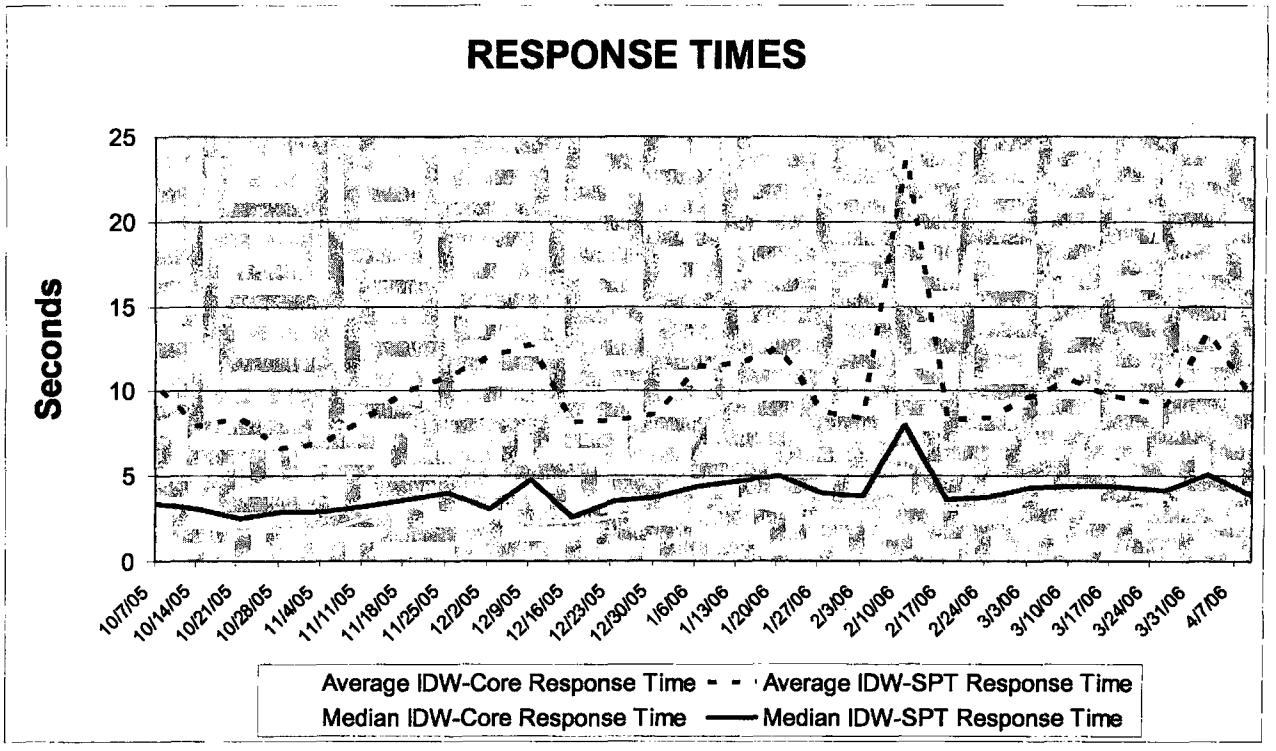
Number of IDW Batch Jobs



7,642 Batch jobs were submitted between 05/25/2004 and 3/17/2006 by a total of 891 different users for an average of 9 batch jobs per user.

Interactive Search Response Times (Weekly)

Average response time ~ 8 to 12 secs, Median response time ~ 3-8 secs



Composition of Data

Document Source	No. of Docs as of 03/17/2006	Percent
FBI Sources	164,271,091	28%
Non-FBI Sources	422,915,362	72%
Total	587,186,453	100%
Document Type	No. of Docs as of 03/17/2006	Percent
Structured	542,939,167	92%
Un-Structured	44,247,286	8%
Total	587,186,453	100%

Case References to IDW Services

- 23,759 references to IDW in ACS/ECF as of 3/17/2006
  - 87% referred to NSB related Case Classifications
  - References to IDW in FCI cases increasing
  - IDW contains only limited case data from Cyber, Civil Rights and Public Corruption
  - CID requests for IDW accounts has increased dramatically in last 2 months
- Sample of batch service job titles
  - Vetting for special events
  - Specific investigations
  - Threat assessments

Areas for Improvement

- Need capability for continuous operation given loss of primary site or capability at primary site
- Additional resources needed to meet for projected growth in number of users and volume of data
- Need more robust (responsiveness and capacity) backup and restore capability
- Need more robust test capability

Status

- Core capability completed
  - Interactive search
  - Batch search
  - Monitor & control capability to manage Quality-of-Service
- Improving service
  - Optimizing capacity & responsiveness given resource constraints
  - Transitioning to Extraction/Transformation/Load service
- Demand for service growing beyond expectations & base resources
  - Host more data (4x-8x current 4 TB)
  - Provide service to more users (~ 20K)
  - Provide continuous monitoring for updates
  - Provide geo-coded data
  - Integrate with emerging Enterprise services

Conclusion

- IDW has managed growth in users & data while maintaining Quality-of-Service
- IDW users are obtaining operational utility from the services provided by IDW
- Additional resources needed to meet greater than anticipated demand for services
- IDW services should be enhanced to provide data-related services
  - Provide services to reduce effort of integration & data migration
  - Provide functions needed to manage data quality

Freedom of Information and Privacy Acts

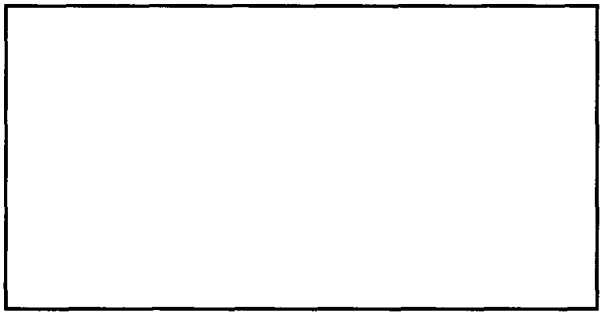
FOIPA # 1058805

Subject: Investigative Data Warehouse

File Number:

Section: 66F-HQ-C1434337

Federal Bureau of Investigation



OTHER O/S

IDW, Version 1.0 is scheduled for release January 2004 and will provide access to six primary data sources:

IDW Version 1.0

- o Electronic Case File (ECF)/Automated Case Support (ACS) /Virtual Case File - FBI investigative and intelligence information.
- o SAMNET - wire traffic to the FBI from members of the intelligence community.
- o JICI - counterterrorism files that were scanned into a database to accommodate the Joint Intelligence Committee Investigation (JICI) on the September 11, 2001, terrorist attack on the World Trade Center.

b2
b7E

- o VGTOF - Violent Crime and Terrorist Offender File information that includes biographical data pertaining to members of the identified groups.
- o Open Source - Fifty-seven news sources from around the world that are either in English or have been translated into English.