

UNCLASSIFIED // For Official Use Only

# Federal Bureau of Investigation

## Investigative Data Warehouse (IDW) Program



### IDW Status Update

(21 September 2005)

Prepared for

[Redacted]

(FBI IDW PM)

[Redacted]

(SPAWAR)

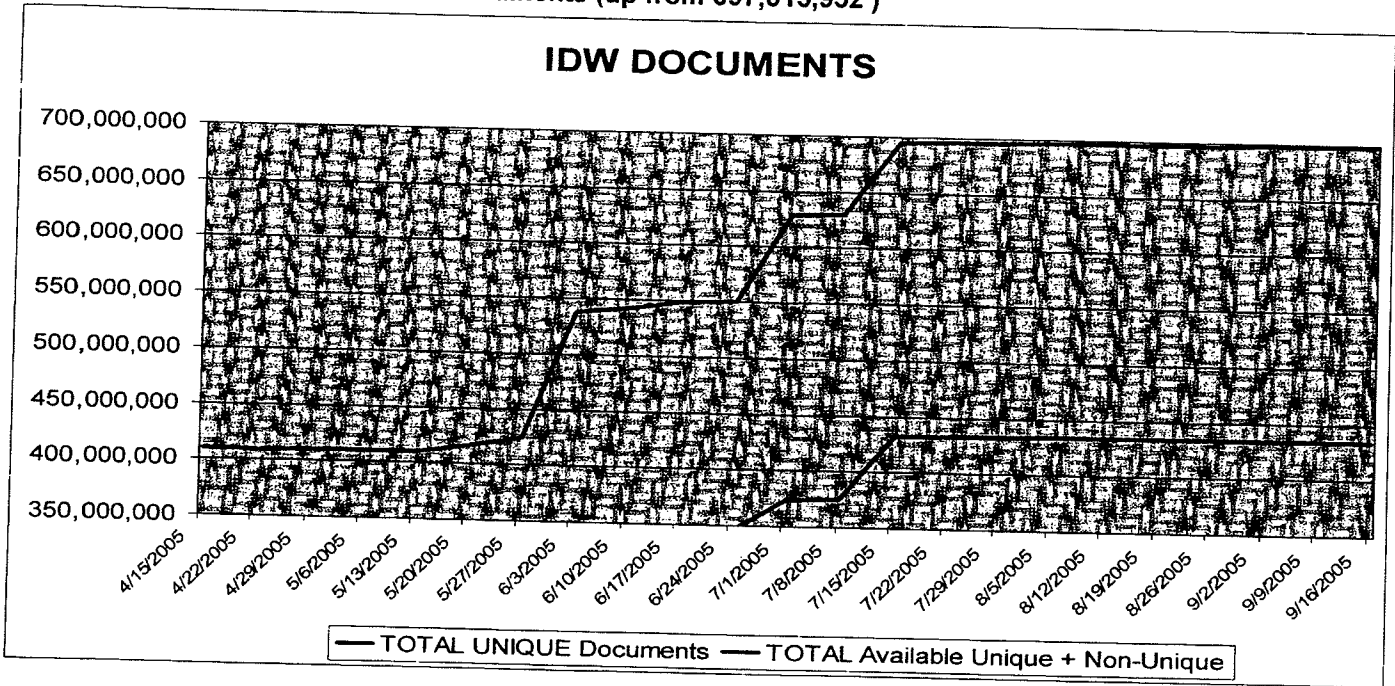
b6  
b7c

ALL INFORMATION CONTAINED  
HEREIN IS UNCLASSIFIED  
DATE 08-02-2007 BY 65179 DMH/BJA/CAL

IDW O&M Status Update  
UNCLASSIFIED // FOUO

### Other Program Updates: IDW Statistics

- Documents in IDW
  - 434,638,628 "unique" documents (up from 434,548,739)
  - 697,728,261 available documents (up from 697,615,952 )



## IDW-S V1.2 Data Collections as of August 27, 2005

### Change Notes:

- (i) The [redacted] Reports Collection was released on the IDW-S V1.2 Core and taken off the IDW-S SPT Subsystem on August 27, 2005. b2 b7E
1. Automated Case System (ACS) Electronic Case File (ECF) – This data set contains copies of the ECs, FD-302s, Facsimiles, FD-542s, Inserts, Transcriptions, Teletypes, Letter Head Memorandums (LHM), Memorandums and other FBI documents contained within ACS/ECF for all but the following FBI case classification violations:  
[redacted] b2  
[redacted] The files are updated daily.
  2. CIA Intelligence Information Reports (IIR-CIA): Electronic copies of all CIA non-TS/non-SCI Intelligence Information Reports (IIRs) and Technical Disseminations (TDs) from 1978 to MAY 2004. These files are not updated.
  3. IntelPlus [redacted] Fileroom – This data set contains copies of scanned pages and their corresponding OCR text files related to an FBI [redacted] investigation of [redacted] Updated periodically. b2 b6 b7C b7E
  4. IntelPlus [redacted] Fileroom – This data set contains copies of scanned pages and their corresponding OCR text files related to multiple FBI field office data regarding the terrorist activities involving [redacted] Updated periodically.
  5. IntelPlus [redacted] Fileroom – This data set contains copies of scanned pages and their corresponding OCR text files related to the [redacted] investigation. Updated periodically. b2 b7E
  6. IntelPlus [redacted] Fileroom – This data set contains copies of scanned pages and their corresponding OCR text files related to an FBI [redacted] investigation of [redacted] Updated periodically. b2 b7E
  7. IntelPlus [redacted] Fileroom – This data set contains copies of scanned pages and their corresponding OCR text files related to [redacted] and provided to the FBI. Updated periodically.
  8. IntelPlus [redacted] Fileroom – This data set contains copies of scanned pages and their corresponding OCR text files related to acts of domestic terrorism, such as the April 19, 1995 Oklahoma City Bombing (Okbom). Updated periodically. b2 b7E

9. IntelPlus [redacted] Fileroom – This data set contains copies of scanned pages and their corresponding OCR text files collected by the FBI's [redacted]. Updated periodically.
10. IntelPlus [redacted] Fileroom – This data set contains copies of scanned pages and their corresponding OCR text files related to Al Qaeda [redacted]. Updated periodically.
11. IntelPlus [redacted] Fileroom – This data set contains copies of scanned pages and their corresponding OCR text files related to an FBI [redacted]. Updated periodically.
12. IntelPlus [redacted] Fileroom – This data set contains copies of scanned pages and their corresponding OCR text files related to an FBI [redacted] investigation of the [redacted]. The data, once received will be updated periodically, preferably daily.
13. IntelPlus [redacted] Fileroom – This data set contains copies of scanned pages and their corresponding OCR text files related to [redacted] from all 56 FBI field divisions and other agencies. Updated periodically.
14. Joint Intelligence Committee Investigation (JICI) – Scanned copies and their corresponding OCR text files of all FBI documents related to extremist Islamic terrorism between 1993 and 2002. These files are not updated.
15. 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. These files are updated daily.
16. Secure Automated Messaging Network (SAMNet) – SAMNet consists of messaging traffic that is sent between different U.S. Government agencies through the Automated Digital Information Network (AutoDIN), including Intelligence Information Reports (IIRs) and Technical Disseminations (TD) from the FBI, Central Intelligence Agency (CIA), Defense Intelligence Agency (DIA), and others from November of 2002 to present. IDW receives copies of these classified messages up to SECRET with no SCI caveats. These files are updated daily.
17. [redacted] – This [redacted] data set contains copies of reports regarding [redacted].
18. 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). This file is updated monthly.

b2  
b7E

b2  
b7E

b2  
b7E

b2  
b7E

# Investigative Data Warehouse Program

Presented by

b6  
b7c

ALL INFORMATION CONTAINED  
HEREIN IS UNCLASSIFIED  
DATE 08-02-2007 BY 65179 DMH/EJA/CAL

September 12, 2005

# Concept of Operation - Data

- CORE library documents
  - ~ 42.3 M unstructured documents in 17 collections
    - ACS ECF: Updated daily
    - IntelPlus (11 Filerooms): Updated when provided
    - SAMNet: Updated daily
    - JICI: Static
    - VGTOF: Updated monthly
    - Open Source News: Updated thrice weekly

- [Redacted]

- SPT documents

Outside the Scope

- Sources

- FBI
- [Redacted]
- Financial Crimes Enforcement Center (FinCen)
- [Redacted]
- [Redacted]
- [Redacted]

b2  
b7E

**Automated Case System (ACS)**  
The ACS supports Investigative Case Management, Document Management, and Case File Management.

**InfoPlus (many components)**  
Foreign Disposition

**Secure Automated Messaging Network (SAMRES) (now FBI Automated Messaging System) (FAMS)**

**Joint Intelligence Computer Center (JC3)**  
All Intelligence Computer Environment, Day to Day, Central Intelligence Unit

**Violent Gang (Foreign) Offender (FROG)**  
Violent Gang (Foreign) Offender (FROG)

**Open Source (Intelligence) (OSINT)**  
Open Source (Intelligence) (OSINT)

**Intelligence Information Resources (IIR)**

September 12, 2005

# Core External Data Sources

13

b2  
b7E  
Outside the Scope

# Planned Major Changes to Services

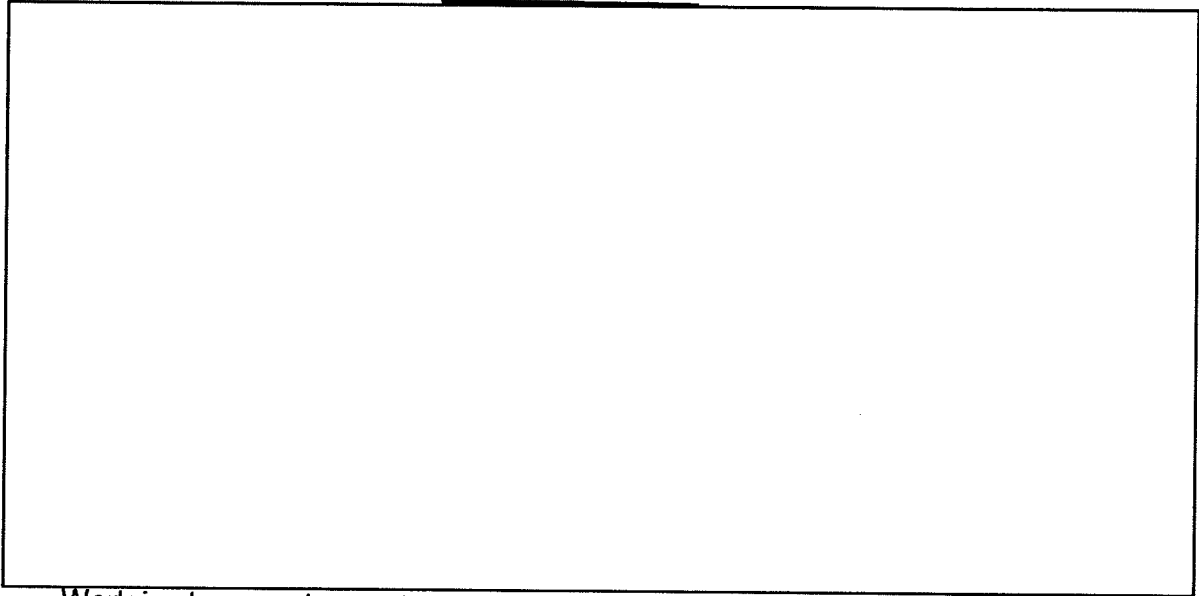
- 
- 
- 
- 
- Data Ingest
  - Add new sources of data as requested & as approved by ISPG
  -
- 
-





# Program QA & CM

Outside the Scope



- Work is also ongoing to develop and populate an Access database to support the property management and configuration management aspects of IDW.

September 12, 2005

Federal Bureau of Investigation  
Information Resources Division  
Data Management Section



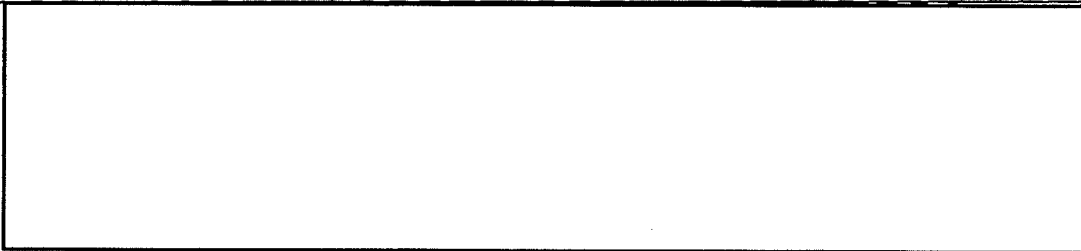
ALL INFORMATION CONTAINED  
HEREIN IS UNCLASSIFIED  
DATE 06-02-2007 BY 65179 DMH/BJA/CAL

Investigative Data Warehouse (IDW)  
Performance Measurement & Monitoring  
Baseline Version 1.0  
March 15, 2004

---

Note: Hardcopy versions of this document must be verified for correct version number with the IDW O&M Document Control Library.

**PMM NG-OM-009.BL1**  
**FOR OFFICIAL USE ONLY**



## 2 IDW Overview

Outside the Scope

IDW is composed of three distinct sections, ingest, user query and control

### 2.1 Ingest

The ingest function accepts information from external sources, validates the data, and formats the data for data storage.

#### 2.1.1 External Sources

IDW receives data from,

- The Automated Case System (ACS) is composed of several applications. These applications are Investigative Case Management (ICM), Electronic Case File (ECF) and the Universal Index (UNI). IDW uses a subset of the Electronic Case File (ECF).
- Intelplus (CT Filerooms) formerly called the [redacted]
- Secure Automated Messaging Network (SAMNET) cable traffic messages received by the FBI
- Violent Gang Terrorist Offender File (VGTOF) from Criminal Justice Information Systems (CJIS) Division
- Joint Intelligence Committee Inquiry (JICI) - Paper counter-terrorism files
- The Open Source Library is collected from global news sources that are either in English or have been translated into English
- [redacted] data received from the [redacted]
- Various paper documents including
  - handwritten notes
  - official FBI correspondence
  - intelligence analysis
  - other terrorism related information

b2  
b7E

#### 2.1.2 Data Validation

IDW has been certified to handle data at a Secret level. Any data classified higher than Secret cannot exist in IDW. The data validation process verifies that all data to be stored in

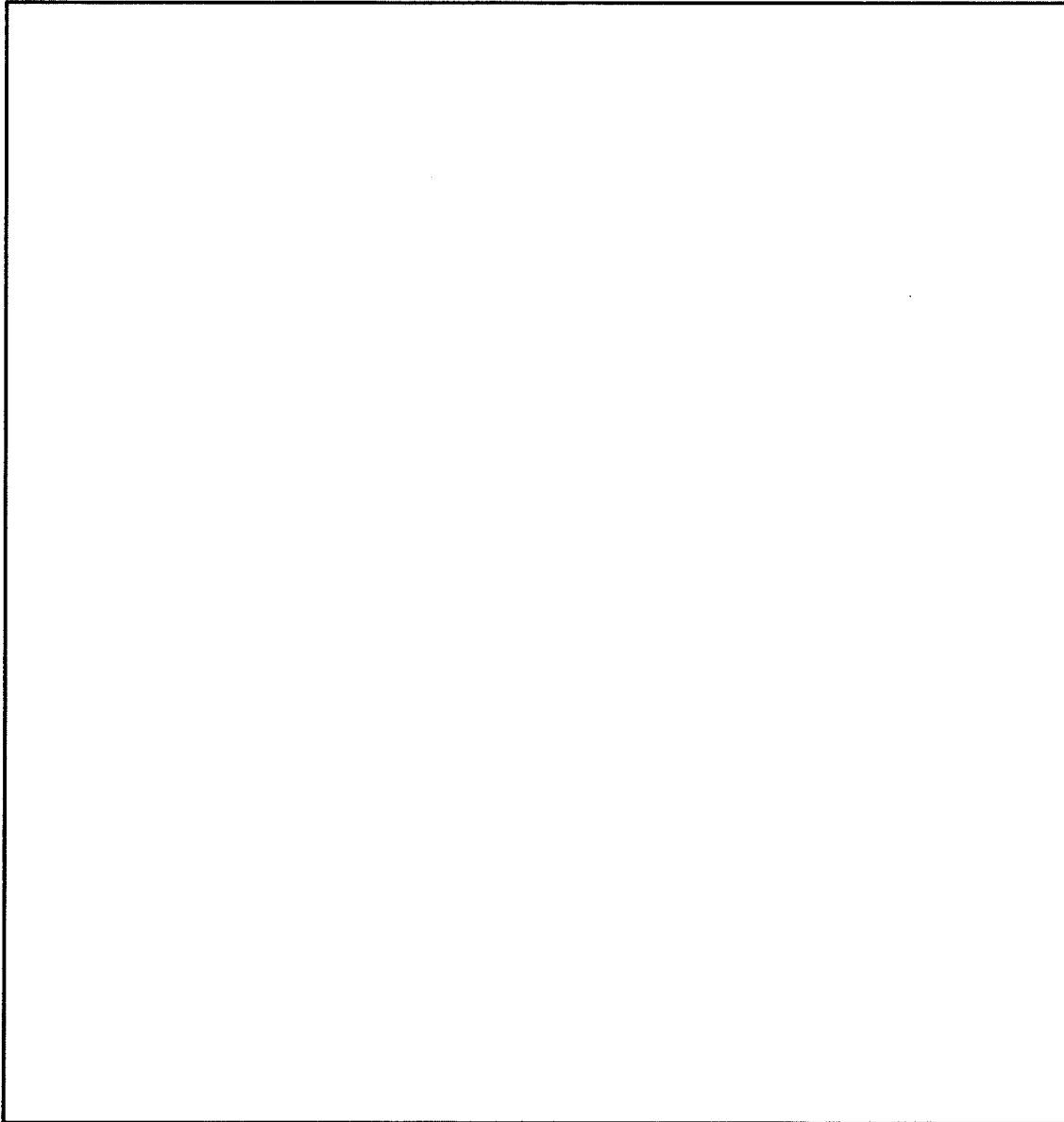
---

Note: Hardcopy versions of this document must be verified for correct version number with the IDW O&M Document Control Library.

**PMM NG-OM-009.BL1**  
**FOR OFFICIAL USE ONLY**

---

IDW meets this security guideline. Data that is questionable is removed from the system, and is reviewed. After the reviewer certifies that the data is appropriate for IDW it is placed into the IDW databases.



Outside the Scope

---

Note: Hardcopy versions of this document must be verified for correct version number with the IDW O&M Document Control Library.



**Investigative Data Warehouse (IDW) Training  
Federal Bureau of Investigation  
Washington, D.C.**

[Redacted]

[Redacted]

1. [Redacted] Data Sources, [Redacted]

[Redacted]

ALL INFORMATION CONTAINED  
HEREIN IS UNCLASSIFIED  
DATE 08-02-2007 BY 65179 DMH/BJA/CAL

**OUTLINE:**

Outside the Scope

1. [Redacted] Data Sources [Redacted]

[Redacted]

b. Data Sources: IDW draws its information from the following data sources:

- **ACS:** Automated Case Support. (**CT Case Files**) Archived Investigative dating back through calendar year 1995. Updates pushed (FTP) to IDW once per day.
- **Intel Plus:** Updates pushed (FTP) to IDW as needed (usually once per day).

- [Redacted]
- [Redacted]
- [Redacted] Domestic terrorism incidents
- [Redacted] Al-Qaeda [Redacted]
- [Redacted] Significant international terrorism incidents

b2  
b7E

- **JICI:** Archived data from the Joint Intelligence Committee Investigation (JICI) regarding 9/11. (Static file system)
  - **ORCON:** Parsed data from SAMNet and ACS data collections at this caveat. Pushed to IDW once per day.
  - **SAMNet:** SECRET and below cable traffic to/from FBI Headquarters. Pulled (FTP) from SAMNet server (IDW chronological job runs every 10 minutes; SAMNet posts new data 3 times per day).
  - **TFRG:** Terrorist Financial records. Updates pushed (FTP) to IDW once per day.
  - **VGTOF:** Violent Gang & Terrorist Offense File. Updated data file is provided in entirety by CD on a weekly basis.
  - **Open Source Research (OSINT):** Current from the MiTAP server.
- c. Handling caveats can apply (here are some examples) –

ORCON – Originator Controlled

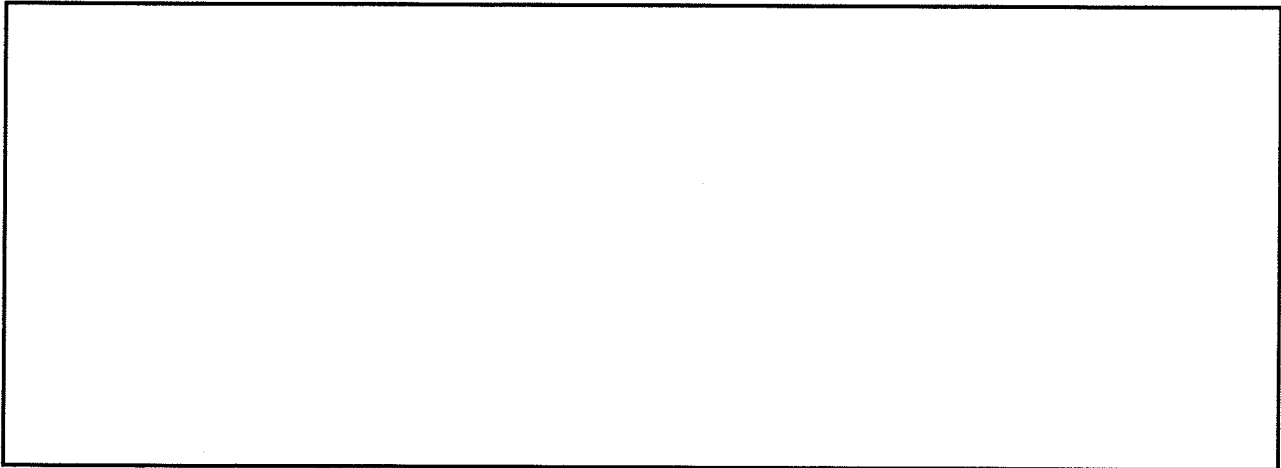
[Redacted]

NOFORN – Not Releasable To Foreign Nationals

LES – Law Enforcement Sensitive

COMINT – Communications Intelligence

b2  
b7E



Outside the Scope



**IDW draws its information from the following data sources:**

**ACS:** Automated Case Support. Archived Investigative dating back through calendar year 1995. **(CT Case Files)**  
Updates pushed (FTP) to IDW once per day.

**Intel Plus** (Updates pushed (FTP) to IDW as needed (once per day)

[REDACTED]

[REDACTED]

[REDACTED] Domestic terrorism incidents

[REDACTED] Al Qaeda [REDACTED]

[REDACTED] Significant international terrorism incidents

**JICI:** Archived data from the Joint Intelligence Committee Investigation (JICI) regarding 9/11. (Static file system)

b2  
b7E





**ORCON:** parsed data from SAMNet and ACS data collections at this caveat. Pushed to IDW once per day

**SAMNet:** SECRET and below cable traffic to/from FBI Headquarters. Pulled (FTP) from SAMNet server (IDW chronological job runs every 10 minutes; SAMNet posts new data 3 times per day)

**TFRG:** Terrorist Financial records

Updates pushed (FTP) to IDW once per day.

**VGTOF:** Violent Gang & Terrorist Offense File Updated datafile is provided in entirety by CD on a weekly basis

**Open Source Research (OSINT):** Pulled (FTP) from MiTAP server daily

**Collections to search:** Select 'Browse' to view a collection's frequently occurring concepts.

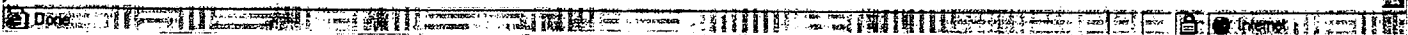
[\(Select all collections\)](#)    [\(Unselect all collections\)](#)

- ACS [\(Browse\)](#)
- Intel+  [\(Browse\)](#)
- Intel+  [\(Browse\)](#)
- Intel+  [\(Browse\)](#)
- Intel+  [\(Browse\)](#)
- Intel+  [\(Browse\)](#)
- JICI [\(Browse\)](#)
- ORCON [\(Browse\)](#)
- Opensource [\(Browse\)](#)
- SAMNET [\(Browse\)](#)
- VGTOF [\(Browse\)](#)

b2  
b7E  
Outside the Scope

b2  
b7E

ALL INFORMATION CONTAINED  
HEREIN IS UNCLASSIFIED  
DATE 08-02-2007 BY 65179 DMH/BJA/CAL



**IDW**

## Open Source Data Sources



Addis Ababa Tribune –  
Agencia Brasilia -  
Al-Ahram (Egypt, weekly English  
version) –  
AllAfrica.com –  
Arabic News –  
Asahi Shimbun -  
Asia Times (Hong Kong) –  
Bangkok Post -  
Christian Science Monitor -  
Crescent International -  
Daily Telegraph (London,  
England) -  
Dawn (Karachi, Pakistan) -  
Debka (Israel) -  
East Africa Daily Nation -  
Gulf News (UAE) -  
Ha'aretz (Israel) -  
IFRC International Federation of  
the Red Cross -  
IRIN Integrated Regional  
Information Network -  
Iraq Press News Agency -  
Islamic Republic News Agency  
(Iran) -  
Jakarta Post -  
Janes -  
Jordan Times -  
L'Osservatore Romano -  
Lagos (Nigeria) Guardian -  
Lahore (Pakistan) Nation -  
Lebanon Daily Star (Beirut) -  
Malaysian Star -  
Manila Bulletin -  
Manila Times -  
Miami Herald -



Moscow Times -  
National Post and CP -  
New Straits Times (Kuala Lumpur) -  
PETRA (Jordanian News Agency) -  
Pakistan Observer (Islamabad) -  
Palestine Chronicle -  
People's Daily (China) -  
Philippine Star -  
Pravda -  
ProMed - epidemiology mail list  
Russian Information Agency  
Novosti -  
Russian Issues (Misc. Russian news) -  
Russian Observer -  
SABA (The News Agency of Yemen) -  
Saudi Gazette -  
South African Dispatch -  
Sydney Morning Herald -  
Tehran Times -  
Times of India -  
UNHCR UN High Commissioner on Refugees -  
Ummah News -  
Uzbekistan Report -  
Washington Post -  
XinHua News Service -  
Yemen Times (weekly) -  
Yomiuri Shimbun (Japan) -

ALL INFORMATION CONTAINED  
HEREIN IS UNCLASSIFIED  
DATE 08-02-2007 BY 65179 DMH/BJA/CAL

IDW-NG-PM-001.01.00

# **Federal Bureau of Investigation (FBI) Investigative Data Warehouse (IDW)**

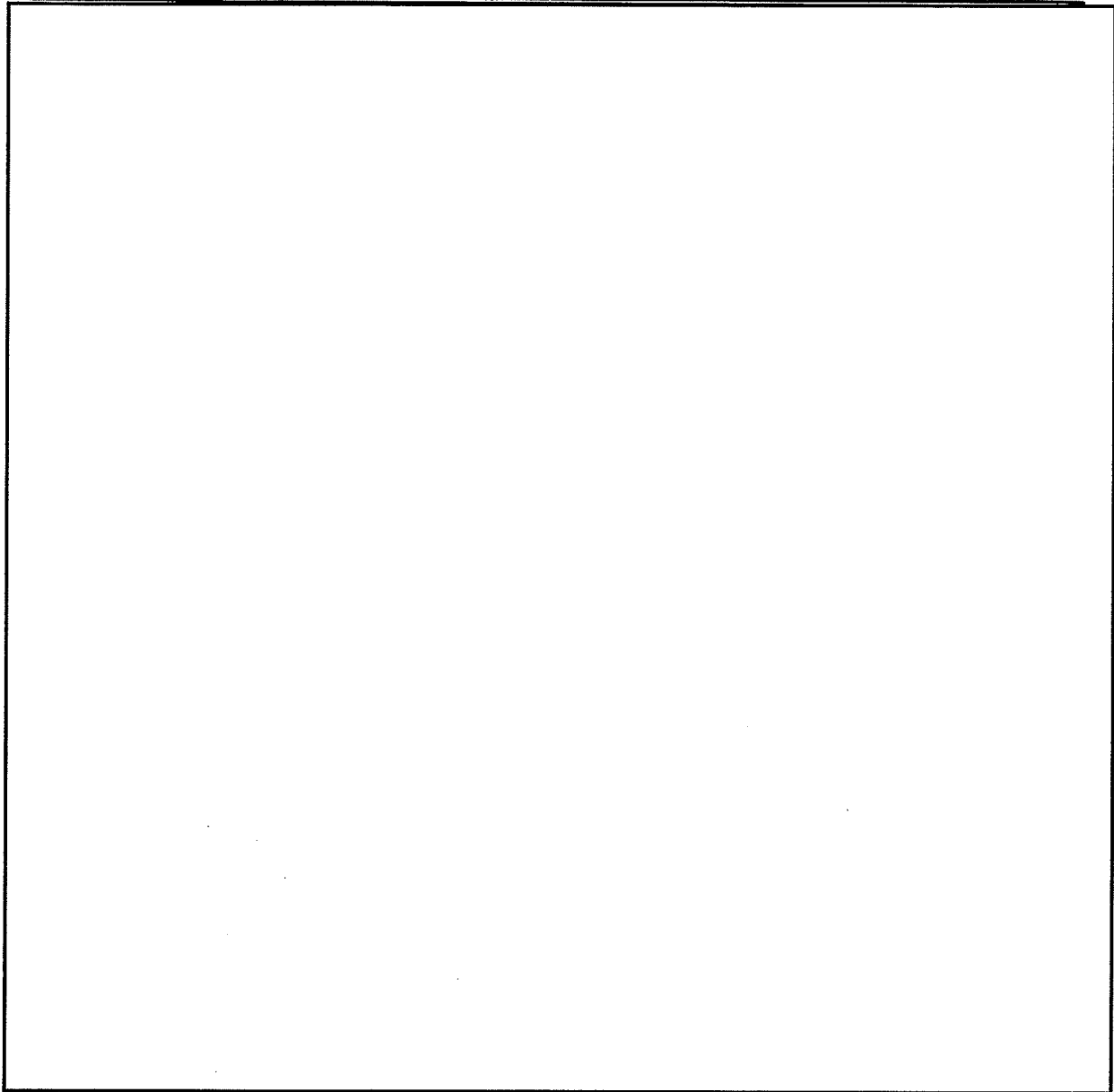
## **Project Management Plan**

**June 30, 2005**



**Prepared under Subcontract 04-6844-27  
(Prime Contract N65236-04-D-6844)  
Task Order 0093**

***NORTHROP GRUMMAN***  
*Information Technology*



Data ingest functions will be the most frequently performed data administration tasks. As this function is primarily tasked to another subcontractor, O&M personnel are not deeply knowledgeable of the specific administrative tasks. The table below indicates the source and frequency of periodic data ingest operations that are performed; other operations may be needed on an ad hoc basis.

Data Type	Media	Frequency
ACS	FTP	Daily

Data Type	Media	Frequency
Intel Plus	FTP	Weekly
JICI	N/A	One-time
SAMNet	FTP	3X Daily
Open Source	CD	Daily
VGTOF	CD	Weekly

Table 5-1 IDW Data Ingest Type, Media and Frequency

# Jarmel-sanchez

 HighView Image  
 06/16/08 07:10





IDW-NG-CA-002.01.06

**Federal Bureau of Investigation (FBI)  
Investigative Data Warehouse (IDW)**

**System Security Plan**

**June 30, 2005**



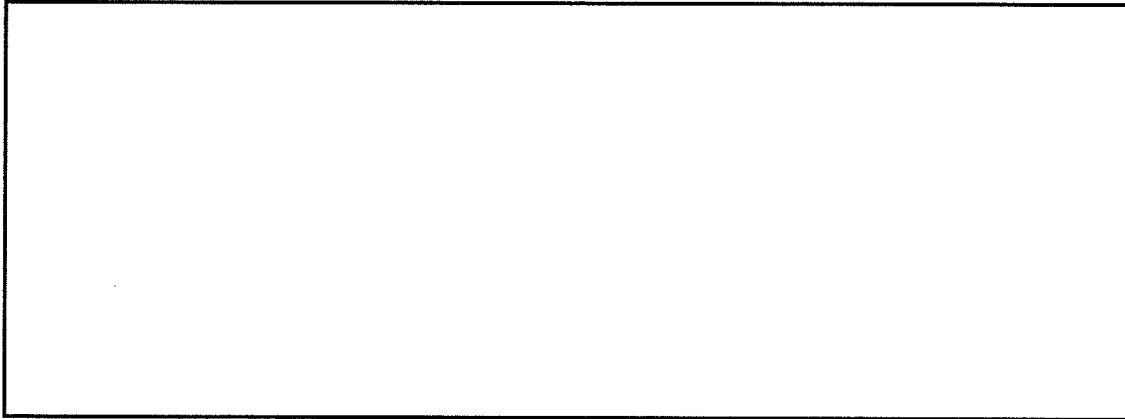
**Prepared under SubContract 04-6844-27  
(Prime Contract N65236-04-D-6844)  
Task Order 0093**

***NORTHROP GRUMMAN***

*Information Technology*

IDW System Data Ingest Flow

Outside the Scope

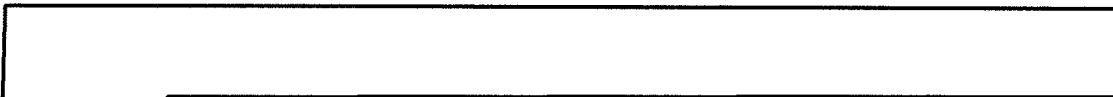


**Automated Transfer of Raw Data to IDW-S**

Raw data is automatically transferred to the IDW-S system from approved data sources in accordance with an Interconnection Agreement, Memorandum of Understanding, Memorandum of Agreement or Interface Control Document. IDW-S V1.2.9 receives update data by the following means:

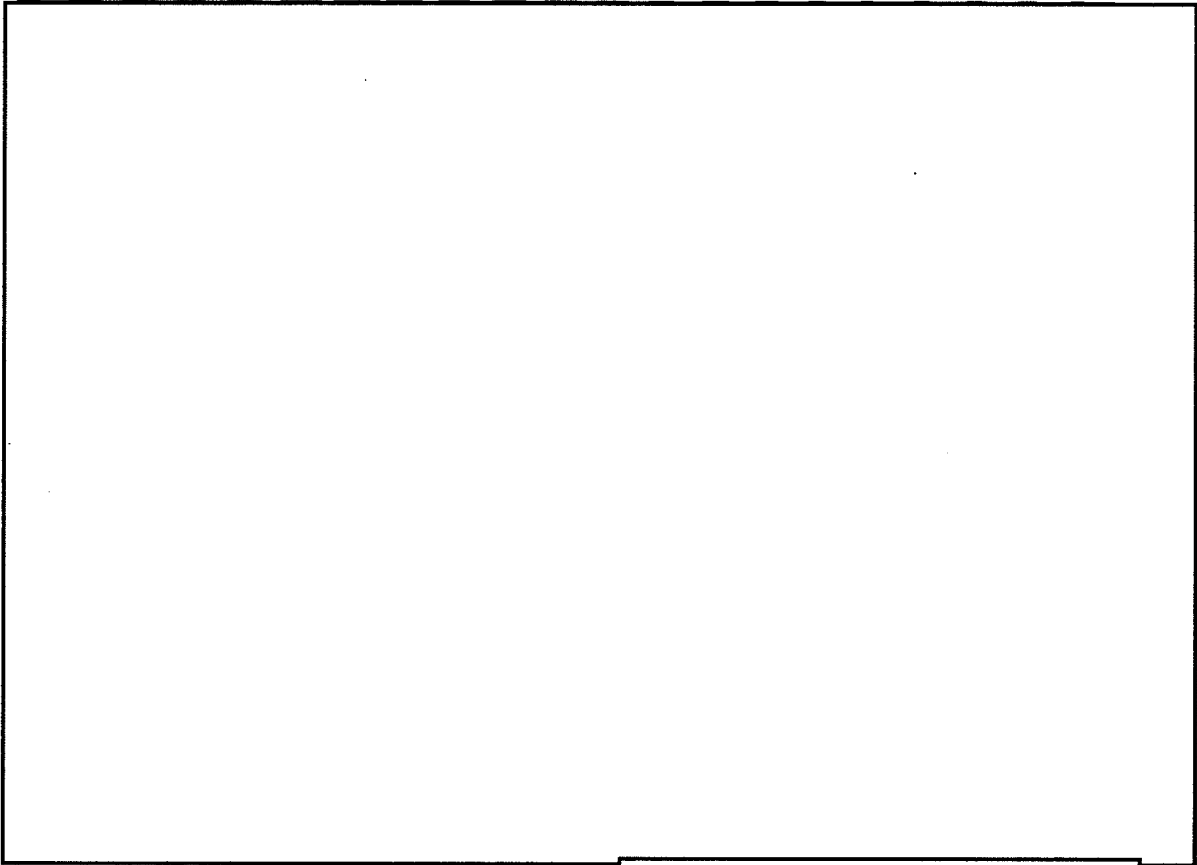
- ACS update data is transmitted to the IDW-S system via File Transfer Protocol (FTP) sessions over FBINet. ACS initiates the FTP session.
- IntelPlus update data is transmitted to the IDW-S system via File Transfer Protocol (FTP) sessions over FBINet. IntelPlus initiates the FTP session.
- SAMNet update data is transmitted to the IDW-S system via File Transfer Protocol (FTP) sessions over FBINet. For SAMNet, the IDW-S system initiates the FTP session.
- VGTOF data is transmitted to the IDW-S system via CD/DVD.
- Open Source News data is transmitted to the IDW-S system via CD/DVD.
- The JICI library is static and is not updated on IDW-S.
- The IIR library is static.

Outside the Scope



data into individual serial records (in the case of ACS ECF), documents (IntelPlus, JICI, Open Source News, IIR), messages (SAMNet), or files (VGTOF)





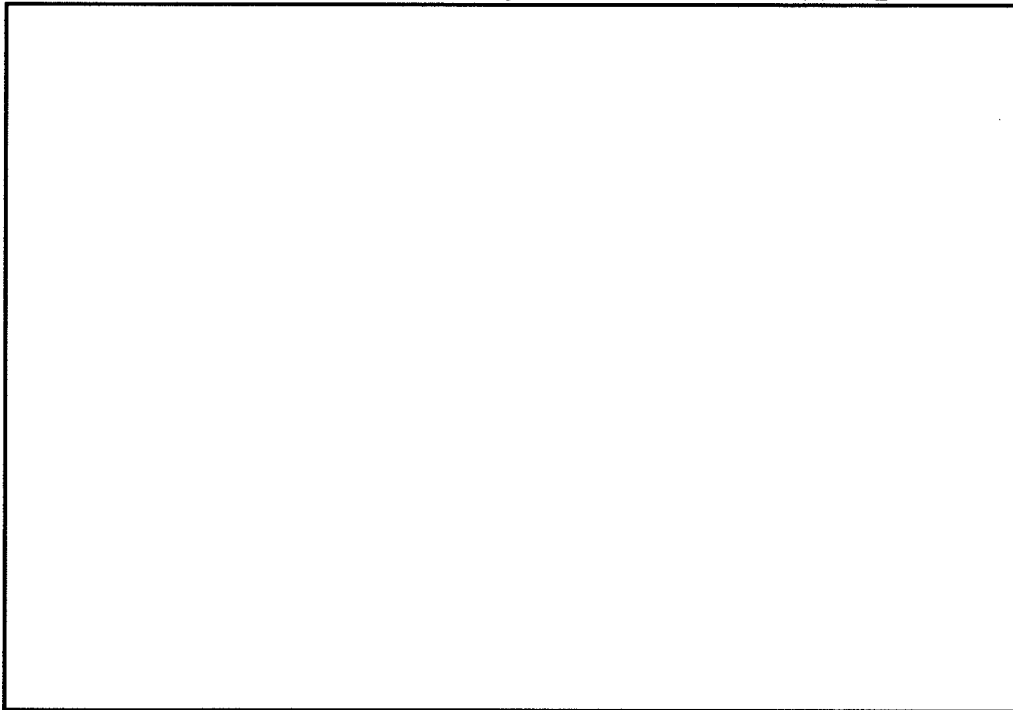
DATASET NAME	DATA SOURCE	CLASSIFICATION
ACS Electronic Case Files (ECF)	FBI Automated Case System (ACS)	Secret and below
IntelPlus Filerooms	FBI IntelPlus	Secret and below

b2  
b7E

DATASET NAME	DATA SOURCE	CLASSIFICATION
SAMNet	FBI Secure Automated Messaging Network (SAMNet)	Secret and below
VGTOF	FBI National Crime Information Center (NCIC)	SBU
JICI	FBI Records Management Division (RMD), Document Laboratory (DocLab), FBIHQ	Secret and below
Open Source News	MiTAP	Unclassified
Intelligence Information Reports	CIA	Secret

b2  
b7E

Table 3.5.1.1 IDW-S Data Sets



Outside the Scope

### 3.9.9 Indirect Connections

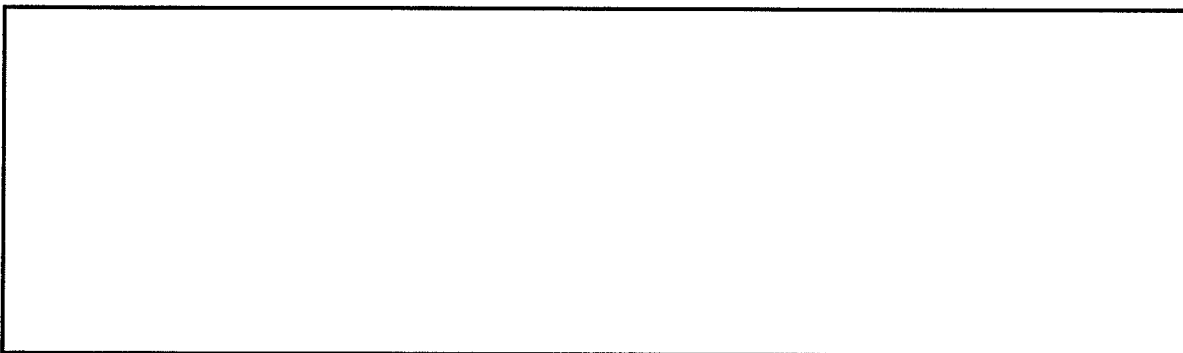
#### 3.9.9.1 IDW-S Core Subsystem Indirect Connections

SYSTEM NAME	CLASSIFICATION & COMPARTMENTS	ACCREDITED BY	TRANSFER METHOD
FBI Automated Case System (ACS)	Secret	FBI	FTP
FBI Intel Plus	Secret	FBI	FTP

b2  
b7E

SYSTEM NAME	CLASSIFICATION & COMPARTMENTS	ACCREDITED BY	TRANSFER METHOD
FBI Secure Automated Messaging Network (SAMNet)	Secret	FBI	FTP
FBI National Crime Information Center (NCIC)	Unclassified/SBU	FBI	CD-ROM
FBI Document Conversion Laboratory (DocLab), Records Management Division (RMD), FBIHQ	Secret	FBI	CD/DVD
MiTAP – Open Source News	Unclassified	NA	CD-ROM
IIR	Secret	CIA	CD-ROM

b2  
b7E



Outside the Scope

# Jarmel-sanchez

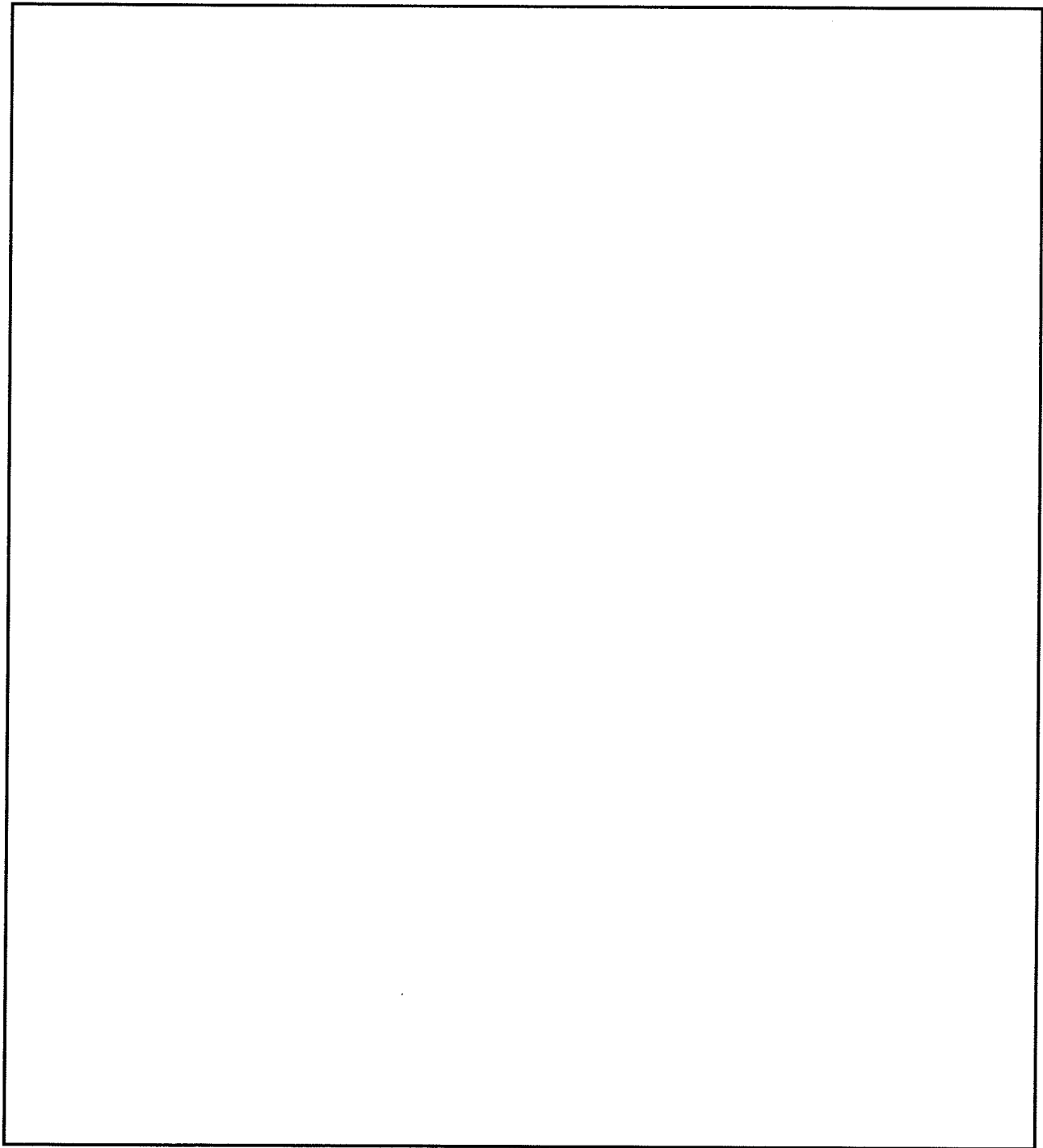
 **HighView Image**  
 **06/16/08 07:10**



b6  
b7c

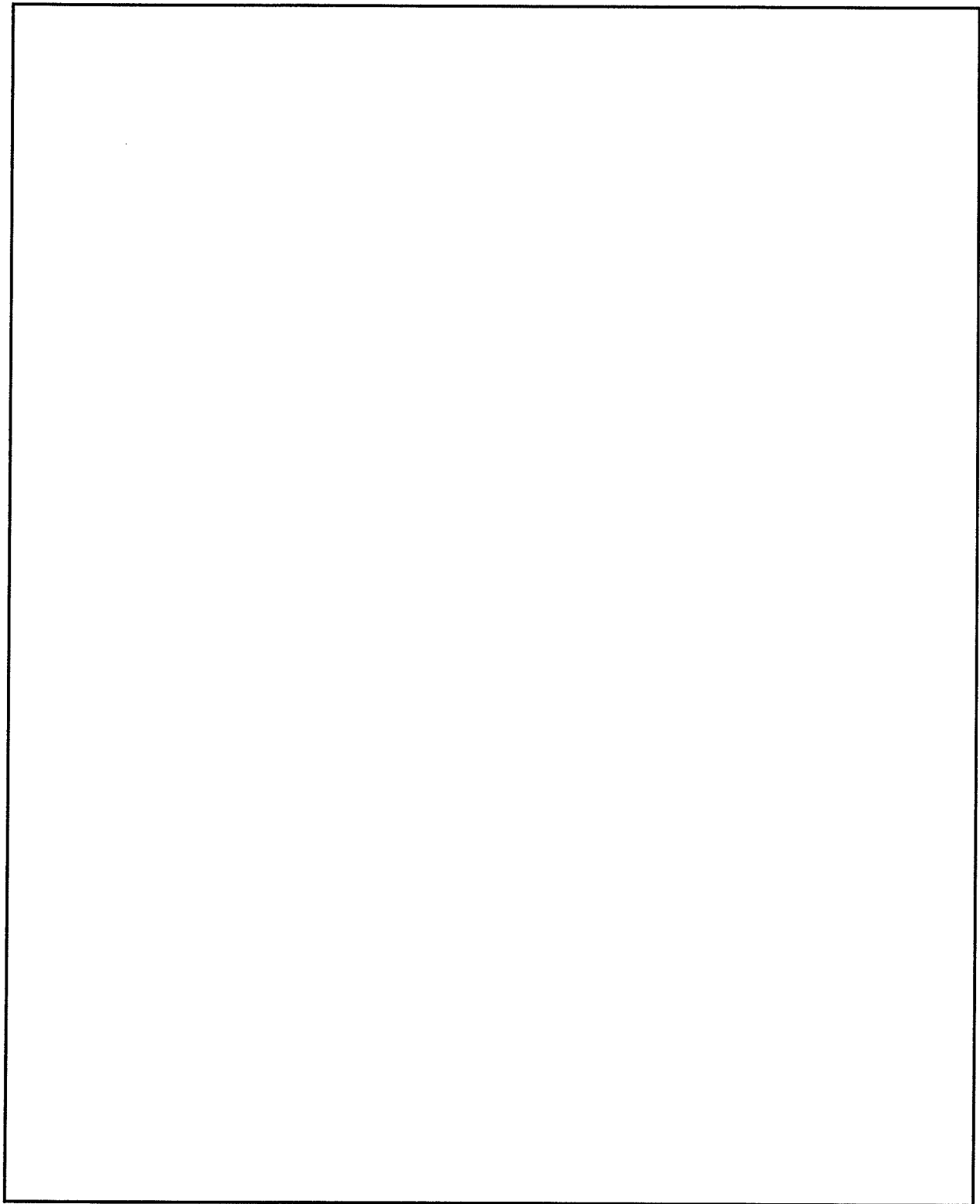
**From:** [REDACTED] (OGC)(FBI)  
**Sent:** Tuesday, September 28, 2004 9:58 AM  
**To:** [REDACTED] (OGC) (FBI)  
**Cc:** MILLER, WILLIAM A. (OGC) (FBI)  
**Subject:** IDW data sources

**UNCLASSIFIED**  
**NON-RECORD**



b5  
b2  
b7E





b5  
b2  
b7E

[Redacted]

Assistant General Counsel  
Administrative Law Unit (Room 7338)  
Office of the General Counsel  
(202) 324-[Redacted]

b6  
b7C



*Federal Bureau of Investigation*

# **FBI Data Warehousing, Data Mining & Collaboration: An Enterprise View of Data**

**Public Brief**

ALL INFORMATION CONTAINED  
HEREIN IS UNCLASSIFIED  
DATE 08-27-2007 BY 65179 DMH/BJA/CAL

**by**

***Ken Ritchhart,***

(202) 324-7781; [KRITCHHART@FBI.GOV](mailto:KRITCHHART@FBI.GOV)  
[KRITCHHA@LEO.GOV](mailto:KRITCHHA@LEO.GOV)

***Section Chief, Data, Engineering & Integration  
Program Management Office***

May 30, 2003

