

[redacted] OGC) (FBI)

From: [redacted] OGC)(FBI)  
Sent: Friday, October 08, 2004 9:06 AM  
To: [redacted] (ITOD)(FBI)  
Cc: [redacted] OGC) (FBI)  
Subject: RE: open source data in IDW

b6  
b7C

UNCLASSIFIED  
NON-RECORD

What do they mean by "named entities?" What we are trying to determine is how these articles are chosen. Obviously, most articles from these news sources are filtered out - I assume we are not getting the local news or the sports section from the Pakistani Observer, etc. So how do they determine which articles go into the database?

Thanks,

Elizabeth N. Jones  
x1778

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

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

From: [redacted] (ITOD)(FBI)  
Sent: Friday, October 08, 2004 8:57 AM  
To: [redacted] (OGC)(FBI)  
Cc: [redacted] (ITOD)(FBI)  
Subject: RE: open source data in IDW

b6  
b7C

UNCLASSIFIED  
NON-RECORD

Here is what I received from MITRE:

The open source data collected for the FBI comes from the MiTAP system run by San Diego State University (SDSU). MiTAP is a complex system written by MITRE that collects raw data from the internet, standardizes the format, extracts named entities, and routes documents into appropriate newsgroups. Although the system was designed to collect foreign language data and process it with machine translation, right now all of the data collected is from English language sources.

The MiTAP system at SDSU collects the data, processes it and makes the data available via a nntp news server. MITRE has a script that checks the server for new news, collects it, packages it into a format suitable for transport and post the data to a web site for (password controlled) download. The server is checked frequently for new data and new packages are posted for download three times a day.

As agreed to by our Office of General Counsel, this process was intended to be temporary – really more of a proof of concept. A better long term solution would be for the FBI to run its own copy of MiTAP and manage its data collections directly. We could certainly help you do that.

Does this answer your questions? If not, let me know.

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

From: [redacted] OGC)(FBI)  
Sent: Tuesday, October 05, 2004 8:14 AM  
To: [redacted] (ITOD)(FBI)

b6  
b7C

**Subject:** open source data in IDW

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

Have you had any luck finding information about the IDW open source data, such as the filters Mitap employs, why it compiles this list, or why it is giving it to the FBI for free?

Thanks,



b2  
b6  
b7C

**UNCLASSIFIED**

**UNCLASSIFIED**

**UNCLASSIFIED**