

UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

RYAN NOAH SHAPIRO,)	
)	
PLAINTIFF)	Civil Action No. 1:13-cv-729 (PLF)
vs.)	
)	
DEPARTMENT OF JUSTICE,)	
)	
DEFENDANT)	
_____)	

**PLAINTIFF’S RESPONSE TO DEFENDANT’S SUPPLEMENTAL
MEMORANDUM**

Introduction

When an FBI Special Agent needs to locate records referring or relating to an individual, they have a variety of options at their disposal. These options include the use of two brand new, state-of-the-art computer systems known as Sentinel and the Investigative Data Warehouse. Additionally, in recognition of the fact that not all relevant records reside in centralized computer databases, Special Agents personally speak to other FBI employees in relevant departments or field offices to assist in locating records, many of which remain in paper format.

In contrast, when a FBI FOIA analyst needs to locate records referring or relating to an individual, the search is typically limited to looking for main file (or sometimes cross-reference) index entries in the Automated Case Support (ACS) system, a decades-old system used to access the FBI’s Central Records System (CRS). In this case, the FBI analyst used only one of the three programs in the ACS system, a search called the Universal Name Index (UNI) to search for main and cross-reference index entries.

There was apparently no attempt to speak with anyone at FBI to determine whether further responsive records existed or to pursue the leads contained the cross-referenced entries.

As a result of the FBI's deficient search for records on Aaron Swartz in this case, it failed to locate the voluminous set of records which would have been easily have been found by employing one or more obvious, minimally burdensome, and readily-available search techniques. The Court should therefore grant summary judgment to Plaintiff, order the FBI to conduct a new search on an expedited basis, and set a new briefing schedule for arguments over the propriety of the FBI's redactions to the records that will be produced as a result of the new search.

Procedural History and Statement of Facts

Plaintiff Ryan Noah Shapiro, a doctoral candidate at MIT,¹ filed a FOIA request with the FBI seeking records on a recently deceased computer programmer and activist named Aaron Swartz. (First Hardy Decl. Ex. A) [ECF dkt: 5-3 at 20-37.] Mr. Shapiro sought and received approval for expedited processing of his request. *Id.*; (First Hardy Decl. Ex. C) [ECF dkt: 5-3 at 41.] The FBI sent a response to Mr. Shapiro stating that it had conducted a search of its Central Records System and located 23 pages of documents, 21 of which were being released. (First Hardy Decl. Ex. D) [ECF dkt: 5-3 at 43-44.] Contending that the "FBI performed an inadequate search for records," Mr. Shapiro submitted an administrative appeal. (First Hardy Decl. Ex. E) [ECF dkt: 5-3 at 47-48.]

¹ This Court's Memorandum Opinion dated March 31, 2014 refers to Aaron Swartz as a doctoral candidate at MIT. It is actually Mr. Shapiro who is a doctoral candidate at MIT; Aaron Swartz was a graduate student at Harvard at the time of his death.

While the administrative appeal was pending, Mr. Shapiro filed the present lawsuit. On March 31, 2014, this Court issued a Memorandum Opinion holding the parties' cross-motions for summary judgment in abeyance pending a new search or further explanation from the FBI about the adequacy of its completed searches. The FBI elected not to conduct a new search, and instead filed a supplemental memorandum detailing why it believed its previous search efforts were adequate.

During the pendency of Mr. Shapiro's lawsuit, the FBI processed FOIA requests from another individual, Mr. John Greenwald, Jr., and released several hundred additional pages of records relating to Aaron Swartz. (Ex. 10; Ex. 11.) The release of documents about Aaron Swartz to Mr. Greenwald did not occur until earlier this month, and Plaintiff only learned of their existence a week ago.

Records Management at the FBI

In its Memorandum Opinion, this Court directed the FBI to "consider whether responsive records would reasonably reside outside the CRS" and to "either conduct a full-text search of ECF or provide further explanation as to why such a search is unnecessary in this particular case." In response, the FBI has provided a further declaration to the Court describing its search efforts in this case and the reasons that it did not conduct a full-text search through ECF. (Second Hardy Decl.) [ECF dkt: 22-1.] The Second Hardy Declaration, however, focuses too narrowly on the options available to the FBI for searches both inside and outside of the CRS. Accordingly, Mr. Shapiro provides additional facts in this section to paint a more complete picture of records management at the FBI.

1. Automated Case Support (ACS)

“Deployed in 1995, ACS is one of several applications residing on the Bureau’s investigative mainframe and is intended to contain information ranging from unclassified to Secret. ACS is the FBI’s investigative system of records and is comprised of three subsystems: a case indexing system; a case management system; and a system to store and retrieve text documents.” Commission for the Review of FBI Security Programs (hereinafter “Webster Report”) (Ex. 15 at 48.) The ACS is one of the systems that allows FBI personnel to access the Central Records System (CRS). (First Hardy Decl. ¶ 15) [ECF dkt: 5-3 at 5] (“Because the CRS cannot electronically query the case files for data such as an individual’s name or social security number, the required information is duplicated and moved to the ACS so that it can be searched.”)

But while the ACS contains a large number of documents, not all records relevant to investigations are contained in it or the CRS. As noted in the preceding paragraph, it is only designed to information classified at up to the Secret level. The lack of controls in ACS means that some users are not able to submit data, in order to protect sources. (Ex. 8 at 23.) Additionally, “[w]orking papers and drafts relating to subjects covered in the case files, as well as less important administrative records, are maintained within divisions, units, squads, and other subordinate groups in local filing locations.” (Ex. 16 at 28.)

Further, records that are *supposed* to be contained in ACS are often not there for a variety of reasons. First, an evaluation of the FBI’s Records Management Architecture revealed that “FBI employees generally do not understand exactly what constitutes a record even though most systems provide for the capability to designate information as

records (either manually or automatically). When dealing with electronic information, FBI employees have even more difficulty understanding what constitutes a record and, if declared a record, what to do with it. The process for capturing and declaring email messages as records is too complicated.” (Ex. 3 at 9-10.) As a result, “[m]ost external media and data are not captured and then managed as an FBI record. Records from desktop applications (e.g., Word documents and emails) often are not properly declared and managed. *This creates vulnerabilities from records management, FOIA, and discovery perspectives.*” (Ex. 3 at 10) (emphasis added).

According to the National Commission on Terrorist Attacks Against the United States (“9/11 Commission”) Staff Statement Number 12, “For a variety of reasons, significant information collected by the FBI never gets uploaded into the Automated Case Support system, or it gets uploaded long after it is learned. One of the reasons for this is the traditional approach to cases, in which information is treated as ‘owned’ by the case agent and maintained in a paper case file. One official told us that headquarters personnel visiting the field have been amazed at the information they found in the paper files.” (Ex. 1 at 6.)

Another problem is with the ACS system itself. According to the Webster Report, “Many FBI agents avoid ACS, often by delegating ACS functions to support staff. Many agents distrust ACS, and, in defiance of Bureau policy, refuse to upload into the system the most sensitive information in their possession.” (Ex. 15 at 48.) The avoidance of ACS is unsurprising, given how archaic the system is. Nearly ten years ago, the Department of Justice (DOJ) Office of the Inspector General (OIG) was already describing the FBI’s ACS system as “antiquated” and “obsolete and limited[.]” (Ex. 14

at 3, 5.) As one former FBI executive explained at a Congressional hearing in 2002, “there’s no mouse, there’s no icon, there’s no year 2000 look to it, it’s all very keyboard intensive.” (Ex. 14 at 15.) According to the FBI’s former Chief Technology Officer, Jack Israel, ACS is “based on old technology . . . you’re dealing with the old IBM green screens. You’re not dealing with a web-based environment, which every one is used to from the Internet.” (Ex. 5 at 3.)



Example of an IBM 3270 “Green Screen” Terminal

The DOJ OIG concluded that there were “national security implications because the FBI is continuing to rely on the ACS and paper files, *which hampers FBI agents and analysts from adequately searching* and sharing information from investigative files.” (Ex. 14 at 6)(emphasis added). The FBI has therefore developed and implemented more modern systems for case management and searching, which are described in Part 3 of this section.

2. Electronic Case File (ECF)

The Electronic Case File (ECF) system is part of the ACS. The “ECF serves as the central electronic repository for the FBI’s official investigative textual documents. ECF provides the capability of uploading word processing documents to the mainframe where they are filed and serialized, parsing uploaded documents for structured document fields and lead information, searching documents by both structured (i.e., formatted fields such as From/To) and unstructured (i.e., full text) means, and downloading documents in their original word processing format. ECF also handles the serialization of non-textual records.” (Ex. 16 at 30.)

As Mr. Hardy admits, searching for keywords in text documents via ECF may locate records that would not be found through an index search of ACS. (Second Hardy Decl. ¶ 8.) This is because the ACS’s index is not a “Google-like index, what the FBI does is manually build an index of all the people, places and things.” (Ex. 5 at 4.) “Under the ACS process, Special Agents marked paper documents with the information they wanted to be indexed and OSTs indexed the documents in ACS.” (Ex. 2 at 15.) Thus, if Special Agents are not sufficiently diligent in marking names on each sheet of paper, a search of the Universal Name Index will not locate responsive records.

3. Sentinel

Sentinel is an “electronic information and case management system that includes records management, workflow management, evidence management, search and reporting capabilities, and information sharing with other law enforcement agencies and

the intelligence community.” (Ex. 2 at 2.) Implemented in July 2012, FBI employees now routinely use Sentinel to perform their daily investigative activities. (Ex. 2 at 2.) According to the FBI, Sentinel has made it easy for Bureau personnel to retrieve and share information: “Sentinel allows the FBI’s law enforcement, intelligence, and support personnel to take advantage of the new opportunities created by the arrival of a modernized global case management system. Information now flows from person to person without the need to generate or mail paper records.” (Ex. 4.)

Users are able to search Sentinel with a search function: “According to a July 2012 FBI report, the search function is both flexible and powerful enough to accommodate the substantial volume and wide variety of information available for retrieval in Sentinel.” (Ex. 2 at 2.) Sentinel currently has the ability to limit searches to specific cases and subfiles contained within a case. (Ex. 2 at 27.)

4. Investigative Data Warehouse

With the exception of a small number of extremely sensitive documents, electronic documents and materials kept in an FBI case file are available through the IDW. (Ex. 8 at 14.) This includes almost all of the data in ACS, as well as dozens of other FBI databases. (Ex. 8 at 16-19; Ex 12.) The IDW also contains millions of scanned paper records which have been converted into computer text, and a large amount of information ingested from other agencies. (Ex. 8 at 16.)

The IDW has advanced search capabilities, including the ability to do multi-word searches. (Ex. 8 at 21.) For example, a user could do a search for “‘flight school’ NEAR/10 ‘lessons’,” which would return all documents where the phrase “flight school”

occurred within 10 words of the word “lessons.” (Ex. 8 at 21.) Typing the name “Mohammad Atta” (one of the 9/11 hijackers) and “flight training” into IDW pulls up 250 articles relating to Atta. (Ex. 13.)

The software can also search for variants of words and names by using the partially incorporated Language Analysis Services (LAS), and can search for variations of birth dates. (Ex. 8 at 21, 24; Ex. 13.) The system is so fast that it can run 1,000 names across 50 databases in 30 minutes, a process which previously would have taken over 32,000 hours. (Ex. 13.)

The IDW can conduct “structured queries” in which the results are limited to a certain date range or FBI classification code. (Ex. 8 at 21.) It also has the ability to extract names from unstructured text, which allows concept-based searched. For example, a user could query information about a terrorist organization and retrieve a list of names extracted from documents about the organization. The user could then select one of the names on the list and look at whether the document also contains a phone number for that individual. (Ex. 8 at 21.)

Records on Aaron Swartz

The FBI reviewed 23 pages of records about Aaron Swartz and produced 21 pages to Mr. Shapiro. Each of the documents is captioned “UNSUB(S); US COURTS – VICTIM; COMPUTER INTRUSION – OTHER” and lists the Case ID # of “288A-WF-238943”. (Hardy Decl. Ex. B) [ECF dkt: 22-1.] Because Mr. Swartz’s name was not in the caption, the records were not part of a “main file” on him. Rather, the caption indicates that the subject of the investigation is unknown (“UNSUB(S)” is an abbreviation for “Unknown

Subject(s)"). The records that were produced were found by conducting a cross-reference search for Aaron Swartz, but only those records in which the Special Agent included Swartz's name as a cross-reference came back as responsive records.

There are, however, hundreds of pages of responsive records that are about Aaron Swartz in the same case file, 288A-WF-238943. (Ex. 10; Ex. 11.) These records were produced to another requester, Mr. Greenwalde, because he had not only asked for the file on Aaron Swartz (as Mr. Shapiro did), but he also asked for the file on Aaron Swartz by its case number.

Given that there were 23 pages of records released to Mr. Shapiro, all coming from the same case file, the FBI should have looked at the entire case file to determine if there were any further responsive records. Had it done so, the FBI would have produced the same records to Mr. Shapiro that it had produced to Mr. Greenwalde.

The failure of the FBI to locate the hundreds of other records about Aaron Swartz in response to Mr. Shapiro's request illustrates how important it is for the FBI to go beyond its reliance on searching the indexes to the ACS. The FBI had numerous, obvious leads it should have followed in this case: reviewing the entire case file 288A-WF-238953;² searching for Swartz's name in ECF, IDW, and Sentinel; and contacting the relevant offices conducting the investigation (Chicago, Washington, and Cyber).

Mr. Hardy contends that such a search is time- and resource-intensive and is typically unlikely to return additional responsive records. Therefore, the FBI is willing to undertake an ECF text search only in two situations: "(a) a search of the CRS locates no

² Although Aaron Swartz is not referred to as a subject of the investigation in the caption, it is clear that he was a focus of the investigation and therefore that other documents in the file would likely relate to him.

records but the FBI has reason to believe that responsive records likely exist; or (b) a search of the CRS results in some records being located, but there is information indicating that more specific responsive records likely exist.” (Second Hardy Decl. ¶ 8.)

Specifically with respect to ECF searches of names, Mr. Hardy avers, “[N]ames in the CRS that are not indexed are generally those deemed to have no continuing significance to the FBI. Such names are generally incomplete and are not accompanied by any other identifying information, such as a date of birth, social security, address, phone number, etc. They may only consist of a first or last name. A full-text search would generally return these unidentifiable names as results. . . . RIDS usually cannot identify the vaguely-referenced individuals whose incomplete names are returned as hits.” (Second Hardy Decl. ¶ 7.)

The problem with Mr. Hardy’s rationale is apparent in this case. In the documents about produced to Mr. Greenwalde, Aaron Swartz’s name was not cross-referenced, but it cannot be said that Swartz would be “deemed to have no continuing significance to the FBI,” given the evidence in the documents of attempts to interview him, his family, and his associates; to subpoena material about him from a grand jury; and to research his background.

There is reason to believe that the FBI has still further records on Aaron Swartz. One of the pages of records released to Mr. Greenwalde notes, “A search in ACS for aaronsw.com [Aaron Swartz’s web and email host] was positive for case ID 315T-HQ³-C1475879-IP, serial 91.” (Ex. 10 at 88.) This suggests that the FBI can and does use ACS to search for information about individuals with search terms other than a name.

³ It is difficult to tell whether it is “HQ” or “HO” because of a redaction box.

Mr. Shapiro's FOIA request provided enough background information about Swartz such that the FBI could not have reasonably limited its search to just the name Swartz and just the name index search. While Mr. Shapiro appreciates that ECF full-text search, IDW search, and Sentinel search might not be appropriate in every case, such searches are necessary here.

Mr. Hardy's explanation of ECF full-text searches is littered with words like "usually," "generally," and "may," but lacks any (even approximate) quantification to back up these characterizations. As this case demonstrates, searches beyond the ACS's UNI index are appropriate in more than extraordinary situations.

Conclusion

Mr. Shapiro renews his request for summary judgment in his favor on the adequacy of the search. The FBI should be ordered to conduct a new search on an expedited basis (as expedited processing was granted by the agency) which includes a search of Sentinel, a full-text search of ECF, a search of IDW, and searches of the Washington, Chicago, and Cyber offices; consultation with agents in the relevant offices who worked on case involvig Swartz to ask for additional responsive records (include paper records and emails); searching the entire file 288A-WF-238943 for responsive records; searching the file 315T-HQ-C1475879 (or "HO") for responsive records. In conducting a new search, search terms which should be included are variations on Swartz's name and other personal information (email address, date of birth, social security number, website domain) as well as keywords likely to return records about investigations into his

activities, such as “PACER” and “JSTOR.” The search of Swartz’s name, date of birth, etc. in the CRS using the UNI feature of ACS was simply insufficient and unreasonable.

Respectfully Submitted,

/s/ Jeffrey Light

Jeffrey L. Light
D.C. Bar #485360
1712 Eye St., NW
Suite 915
Washington, DC 20006
(202)277-6213
Jeffrey.Light@yahoo.com

Counsel for Plaintiff