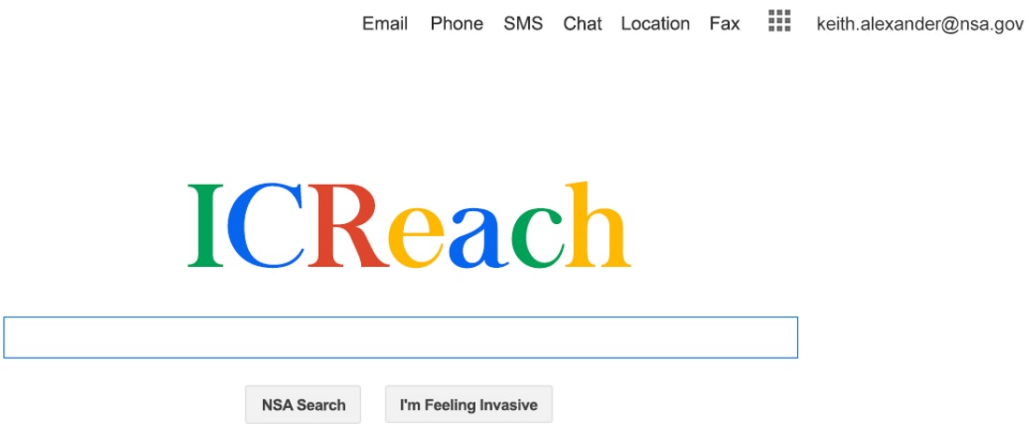
The Intercept_

THE SURVEILLANCE ENGINE

How the NSA Built Its Own Secret Google

Ryan Gallagher

August 25 2014, 7:09 p.m.



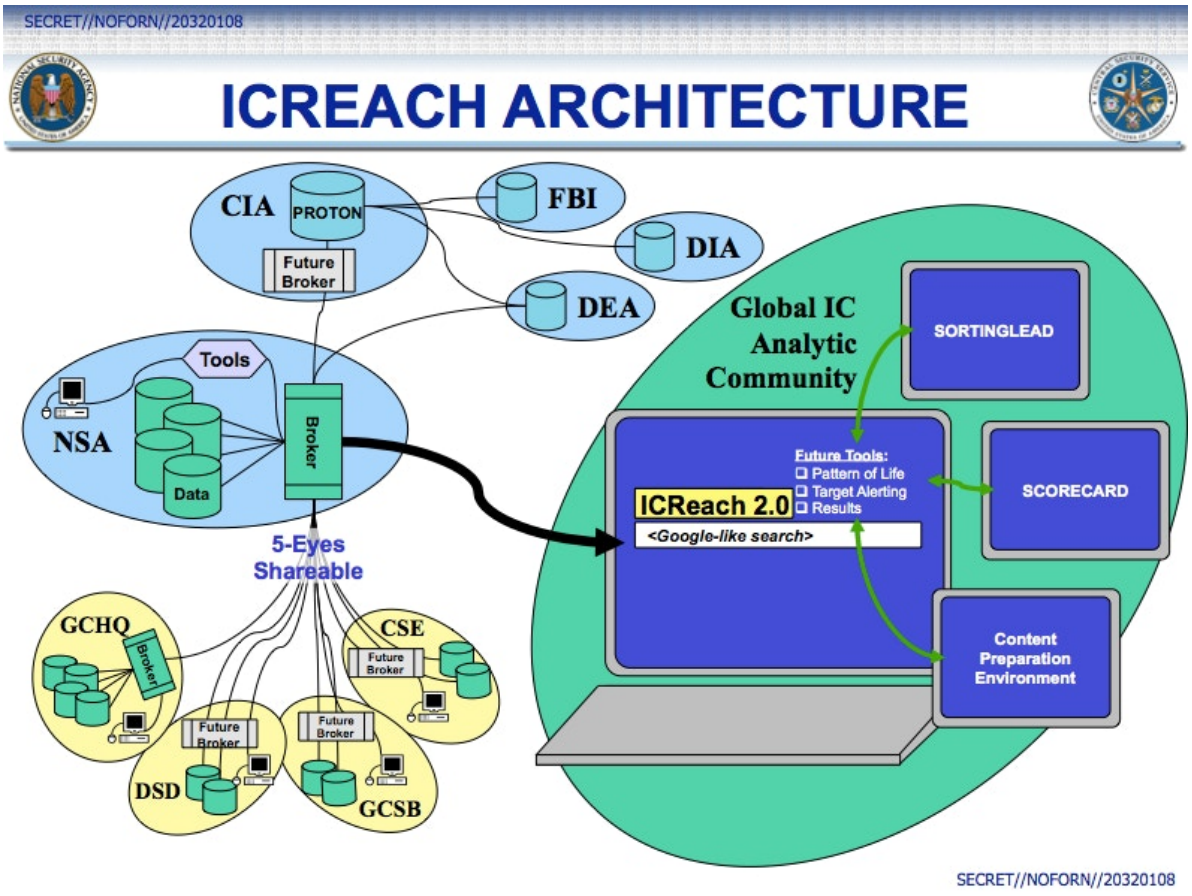
The National Security Agency is secretly providing data to nearly two dozen U.S. government agencies with a "Google-like" search engine built to share more than 850 billion records about phone calls, emails, cellphone locations, and internet chats, according to classified documents obtained by *The Intercept*.

The documents provide the first definitive evidence that the NSA has for years made massive amounts of surveillance data di-

rectly accessible to domestic law enforcement agencies. Planning documents for ICREACH, as the search engine is called, cite the Federal Bureau of Investigation and the Drug Enforcement Administration as key participants.

ICREACH contains information on the private communications of foreigners and, it appears, millions of records on American citizens who have not been accused of any wrongdoing. Details about its existence are contained in the archive of materials provided to *The Intercept* by NSA whistleblower Edward Snowden.

Earlier revelations sourced to the Snowden documents have exposed a multitude of NSA programs for collecting large volumes of communications. The NSA has acknowledged that it shares some of its collected data with domestic agencies like the FBI, but details about the method and scope of its sharing have remained shrouded in secrecy.



ICREACH has been accessible to more than 1,000 analysts at 23 U.S. government agencies that perform intelligence work, according to a 2010 memo. A planning document from 2007 lists the DEA, FBI, Central Intelligence Agency, and the Defense Intelligence Agency as core members. Information shared through ICREACH can be used to track people's movements, map out their networks of associates, help predict future actions, and potentially reveal religious affiliations or political beliefs.

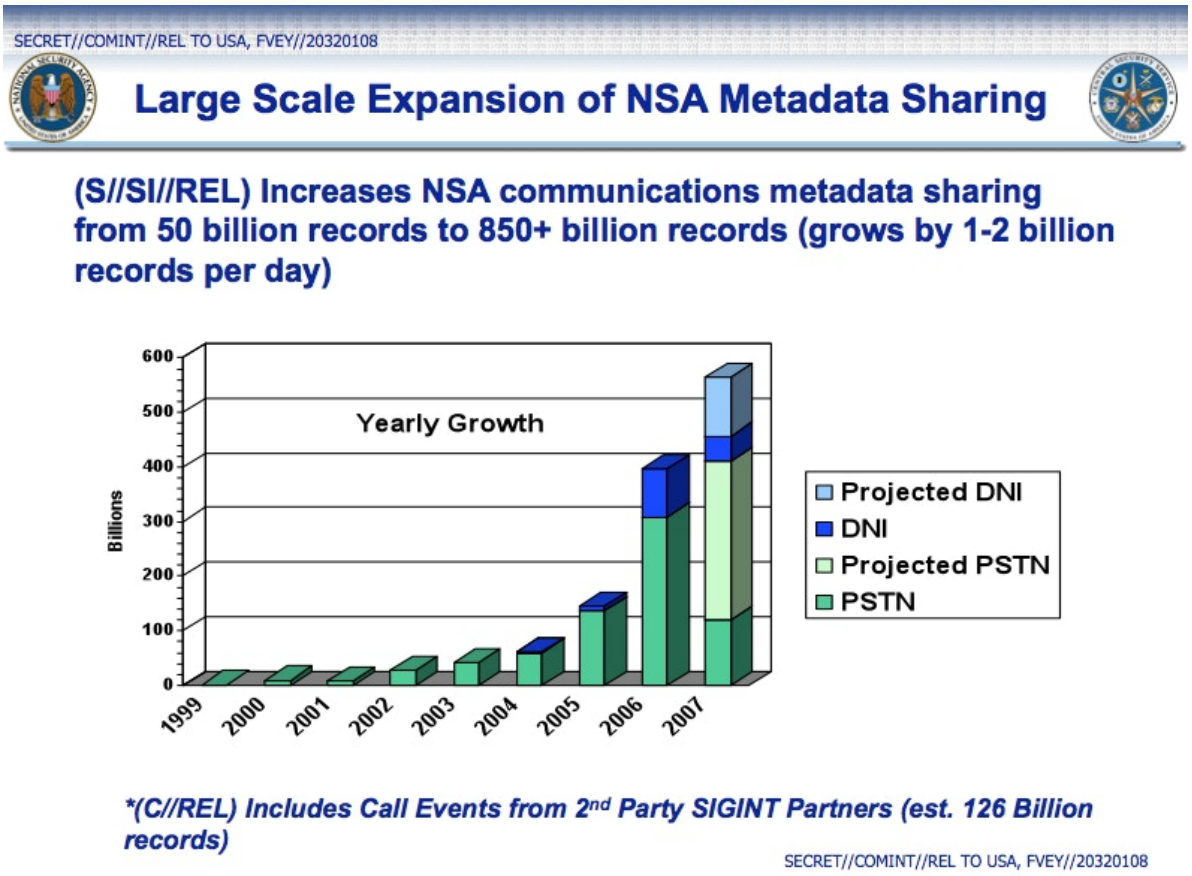
The creation of ICREACH represented a landmark moment in the history of classified U.S. government surveillance, according to the NSA documents.

"The ICREACH team delivered the first-ever wholesale sharing of communications metadata within the U.S. Intelligence Community," noted a top-secret memo dated December 2007. "This team began over two years ago with a basic concept compelled by the IC's increasing need for communications metadata and NSA's ability to collect, process and store vast amounts of communications metadata related to worldwide intelligence targets."

The search tool was designed to be the largest system for internally sharing secret surveillance records in the United States, capable of handling two to five billion new records every day, including more than 30 different kinds of metadata on emails, phone calls, faxes, internet chats, and text messages, as well as location information collected from cellphones. Metadata reveals information about a communication – such as the "to" and "from" parts of an email, and the time and date it was sent, or the phone numbers someone called and when they called – but not the content of the message or audio of the call.

ICREACH does not appear to have a direct relationship to the large NSA database, previously reported by *The Guardian*, that

stores information on millions of ordinary Americans' phone calls under Section 215 of the Patriot Act. Unlike the 215 database, which is accessible to a small number of NSA employees and can be searched only in terrorism-related investigations, ICREACH grants access to a vast pool of data that can be mined by analysts from across the intelligence community for "foreign intelligence" – a vague term that is far broader than counterterrorism.



Data available through ICREACH appears to be primarily derived from surveillance of foreigners' communications, and planning documents show that it draws on a variety of different sources of data maintained by the NSA. Though one 2010 internal paper clearly calls it "the ICREACH database," a U.S. official familiar with the system disputed that, telling *The Intercept* that while "it enables the sharing of certain foreign intelligence metadata," ICREACH is "not a repository [and] does not store events or

records." Instead, it appears to provide analysts with the ability to perform a one-stop search of information from a wide variety of separate databases.

In a statement to *The Intercept*, the Office of the Director of National Intelligence confirmed that the system shares data that is swept up by programs authorized under Executive Order 12333, a controversial Reagan-era presidential directive that underpins several NSA bulk surveillance operations that monitor communications overseas. The 12333 surveillance takes place with no court oversight and has received minimal Congressional scrutiny because it is targeted at foreign, not domestic, communication networks. But the broad scale of 12333 surveillance means that some Americans' communications get caught in the dragnet as they transit international cables or satellites – and documents contained in the Snowden archive indicate that ICREACH taps into some of that data.

Legal experts told *The Intercept* they were shocked to learn about the scale of the ICREACH system and are concerned that law enforcement authorities might use it for domestic investigations that are not related to terrorism.

"To me, this is extremely troublesome," said Elizabeth Goitein, co-director of the Liberty and National Security Program at the New York University School of Law's Brennan Center for Justice. "The myth that metadata is just a bunch of numbers and is not as revealing as actual communications content was exploded long ago – this is a trove of incredibly sensitive information."

Brian Owsley, a federal magistrate judge between 2005 and 2013, said he was alarmed that traditional law enforcement agencies such as the FBI and the DEA were among those with access to the NSA's surveillance troves.

"This is not something that I think the government should be doing," said Owsley, an assistant professor of law at Indiana Tech Law School. "Perhaps if information is useful in a specific case, they can get judicial authority to provide it to another agency. But there shouldn't be this buddy-buddy system back-and-forth."

Jeffrey Anchukaitis, an ODNI spokesman, declined to comment on a series of questions from *The Intercept* about the size and scope of ICREACH, but said that sharing information had become "a pillar of the post-9/11 intelligence community" as part of an effort to prevent valuable intelligence from being "stove-piped in any single office or agency."

Using ICREACH to query the surveillance data, "analysts can develop vital intelligence leads without requiring access to raw intelligence collected by other IC [Intelligence Community] agencies," Anchukaitis said. "In the case of NSA, access to raw signals intelligence is strictly limited to those with the training and authority to handle it appropriately. The highest priority of the intelligence community is to work within the constraints of law to collect, analyze and understand information related to potential threats to our national security."

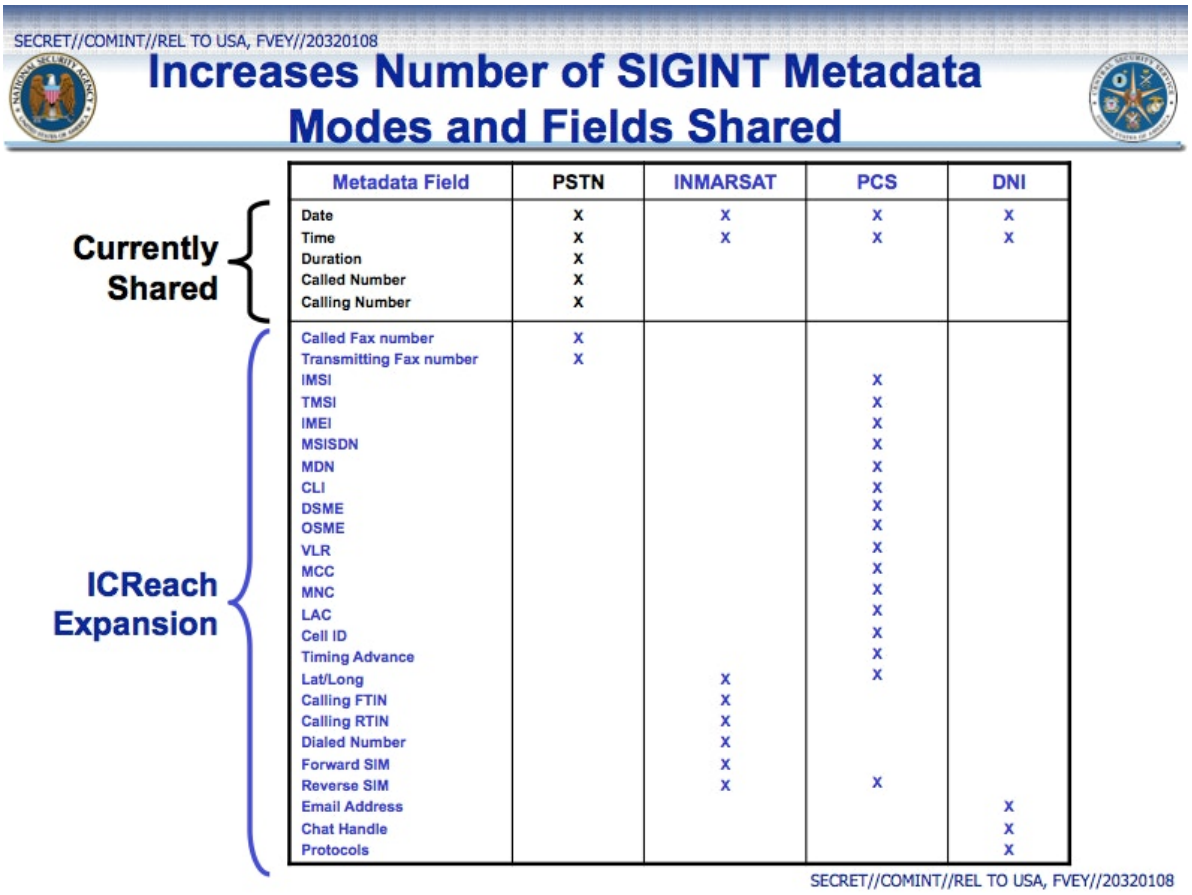
One-Stop Shopping

The mastermind behind ICREACH was recently retired NSA director Gen. Keith Alexander, who outlined his vision for the system in a classified 2006 letter to the then-Director of National Intelligence John Negroponte. The search tool, Alexander wrote, would "allow unprecedented volumes of communications metadata to be shared and analyzed," opening up a "vast, rich source of information" for other agencies to exploit. By late 2007 the NSA reported to its employees that the system had gone live as a pilot program.

The NSA described ICREACH as a "one-stop shopping tool" for analyzing communications. The system would enable at least a 12-fold increase in the volume of metadata being shared between intelligence community agencies, the documents stated. Using ICREACH, the NSA planned to boost the amount of communications "events" it shared with other U.S. government agencies from 50 billion to more than 850 billion, bolstering an older top-secret data sharing system named CRISSCROSS/PROTON, which was launched in the 1990s and managed by the CIA.

To allow government agents to sift through the masses of records on ICREACH, engineers designed a simple "Google-like" search interface. This enabled analysts to run searches against particular "selectors" associated with a person of interest – such as an email address or phone number – and receive a page of re-

sults displaying, for instance, a list of phone calls made and received by a suspect over a month-long period. The documents suggest these results can be used reveal the "social network" of the person of interest – in other words, those that they communicate with, such as friends, family, and other associates.



Metadata Field	PSTN	INMARSAT	PCS	DNI
Date	X	X	X	X
Time	X	X	X	X
Duration	X			
Called Number	X			
Calling Number	X			
Called Fax number	X			
Transmitting Fax number	X			
IMSI			X	
TMSI			X	
IMEI			X	
MSISDN			X	
MDN			X	
CLI			X	
DSME			X	
OSME			X	
VLR			X	
MCC			X	
MNC			X	
LAC			X	
Cell ID			X	
Timing Advance			X	
Lat/Long		X	X	
Calling FTIN		X		
Calling RTIN		X		
Dialed Number		X		
Forward SIM		X		
Reverse SIM		X	X	
Email Address				X
Chat Handle				X
Protocols				X

The purpose of ICREACH, projected initially to cost between $2.5 million and $4.5 million per year, was to allow government agents to comb through the NSA's metadata troves to identify new leads for investigations, to predict potential future threats against the U.S., and to keep tabs on what the NSA calls "worldwide intelligence targets."

However, the documents make clear that it is not only data about foreigners' communications that are available on the system. Alexander's memo states that "many millions of…minimized communications metadata records" would be available through ICREACH, a reference to the process of "minimization,"

whereby identifying information – such as part of a phone number or email address – is removed so it is not visible to the analyst. NSA documents define minimization as "specific procedures to minimize the acquisition and retention [of] information concerning unconsenting U.S. persons" – making it a near certainty that ICREACH gives analysts access to millions of records about Americans. The "minimized" information can still be retained under NSA rules for up to five years and "unmasked" at any point during that period if it is ever deemed necessary for an investigation.

The Brennan Center's Goitein said it appeared that with ICREACH, the government "drove a truck" through loopholes that allowed it to circumvent restrictions on retaining data about Americans. This raises a variety of legal and constitutional issues, according to Goitein, particularly if the data can be easily searched on a large scale by agencies like the FBI and DEA for their domestic investigations.

"The idea with minimization is that the government is basically supposed to pretend this information doesn't exist, unless it falls under certain narrow categories," Goitein said. "But functionally speaking, what we're seeing here is that minimization means, 'we'll hold on to the data as long as we want to, and if we see anything that interests us then we can use it.'"

A key question, according to several experts consulted by *The Intercept*, is whether the FBI, DEA or other domestic agencies have used their access to ICREACH to secretly trigger investigations of Americans through a controversial process known as "parallel construction."

Parallel construction involves law enforcement agents using information gleaned from covert surveillance, but later covering

up their use of that data by creating a new evidence trail that excludes it. This hides the true origin of the investigation from defense lawyers and, on occasion, prosecutors and judges – which means the legality of the evidence that triggered the investigation cannot be challenged in court.

In practice, this could mean that a DEA agent identifies an individual he believes is involved in drug trafficking in the United States on the basis of information stored on ICREACH. The agent begins an investigation but pretends, in his records of the investigation, that the original tip did not come from the secret trove. Last year, Reuters first reported details of parallel construction based on NSA data, linking the practice to a unit known as the Special Operations Division, which Reuters said distributes tips from NSA intercepts and a DEA database known as DICE.

Tampa attorney James Felman, chair of the American Bar Association's criminal justice section, told *The Intercept* that parallel construction is a "tremendously problematic" tactic because law enforcement agencies "must be honest with courts about where they are getting their information." The ICREACH revelations, he said, "raise the question of whether parallel construction is present in more cases than we had thought. And if that's true, it is deeply disturbing and disappointing."

Anchukaitis, the ODNI spokesman, declined to say whether ICREACH has been used to aid domestic investigations, and he would not name all of the agencies with access to the data. "Access to information-sharing tools is restricted to users conducting foreign intelligence analysis who have the appropriate training to handle the data," he said.

 CIA headquarters in Langley, Virginia, 2001.

Project CRISSCROSS

The roots of ICREACH can be traced back more than two decades.

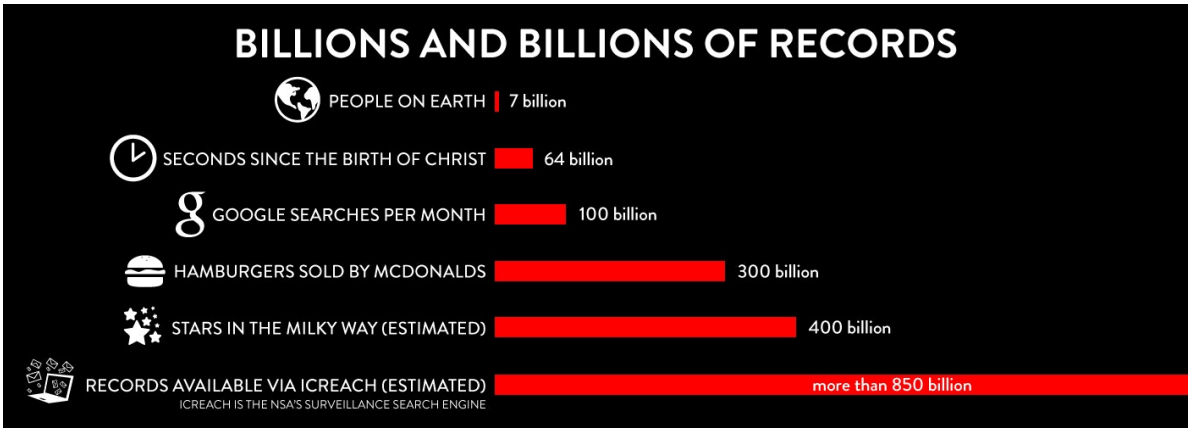
In the early 1990s, the CIA and the DEA embarked on a secret initiative called Project CRISSCROSS. The agencies built a database system to analyze phone billing records and phone directories, in order to identify links between intelligence targets and other persons of interest. At first, CRISSCROSS was used in Latin America and was "extremely successful" at identifying narcotics-related suspects. It stored only five kinds of metadata on phone calls: date, time, duration, called number, and calling number, according to an NSA memo.

The program rapidly grew in size and scope. By 1999, the NSA, the Defense Intelligence Agency, and the FBI had gained access to CRISSCROSS and were contributing information to it. As CRISSCROSS continued to expand, it was supplemented with a system called PROTON that enabled analysts to store and examine additional types of data. These included unique codes used to identify individual cellphones, location data, text messages, passport and flight records, visa application information, as well as excerpts culled from CIA intelligence reports.

An NSA memo noted that PROTON could identify people based on whether they behaved in a "similar manner to a specific target." The memo also said the system "identifies correspondents in common with two or more targets, identifies potential new phone numbers when a target switches phones, and identifies networks of organizations based on communications within the group." In July 2006, the NSA estimated that it was storing 149 billion phone records on PROTON.

According to the NSA documents, PROTON was used to track down "High Value Individuals" in the United States and Iraq, investigate front companies, and discover information about foreign government operatives. CRISSCROSS enabled major narcotics arrests and was integral to the CIA's rendition program during the Bush Administration, which involved abducting terror suspects and flying them to secret "black site" prisons where they were brutally interrogated and sometimes tortured. One NSA document on the system, dated from July 2005, noted that the use of communications metadata "has been a contribution to virtually every successful rendition of suspects and often, the deciding factor."

However, the NSA came to view CRISSCROSS/PROTON as insufficient, in part due to the aging standard of its technology. The intelligence community was sensitive to criticism that it had failed to share information that could potentially have helped prevent the 9/11 attacks, and it had been strongly criticized for intelligence failures before the invasion of Iraq in 2003. For the NSA, it was time to build a new and more advanced system to radically increase metadata sharing.



A New Standard

In 2006, NSA director Alexander drafted his secret proposal to then-Director of National Intelligence Negroponte.

Alexander laid out his vision for what he described as a "communications metadata coalition" that would be led by the NSA. His idea was to build a sophisticated new tool that would grant other federal agencies access to "more than 50 existing NSA/CSS metadata fields contained in trillions of records" and handle "many millions" of new minimized records every day – indicating that a large number of Americans' communications would be included.

The NSA's contributions to the ICREACH system, Alexander wrote, "would dwarf the volume of NSA's present contributions to PROTON, as well as the input of all other [intelligence community] contributors."

Alexander explained in the memo that NSA was already collecting "vast amounts of communications metadata" and was preparing to share some of it on a system called GLOBALREACH with its counterparts in the so-called Five Eyes surveillance alliance: the United Kingdom, Australia, Canada, and New Zealand.

ICREACH, he proposed, could be designed like GLOBALREACH and accessible only to U.S. agencies in the intelligence community, or IC.

A top-secret PowerPoint presentation from May 2007 illustrated how ICREACH would work – revealing its "Google-like" search interface and showing how the NSA planned to link it to the DEA, DIA, CIA, and the FBI. Each agency would access and input data through a secret data "broker" – a sort of digital letterbox – linked to the central NSA system. ICREACH, according to the

presentation, would also receive metadata from the Five Eyes allies.

The aim was not necessarily for ICREACH to completely replace CRISSCROSS/PROTON, but rather to complement it. The NSA planned to use the new system to perform more advanced kinds of surveillance – such as "pattern of life analysis," which involves monitoring who individuals communicate with and the places they visit over a period of several months, in order to observe their habits and predict future behavior.

The NSA agreed to train other U.S. government agencies to use ICREACH. Intelligence analysts could be "certified" for access to the massive database if they required access in support of a given mission, worked as an analyst within the U.S. intelligence community, and had top-secret security clearance. (According to the latest government figures, there are more than 1.2 million government employees and contractors with top-secret clearance.)

In November 2006, according to the documents, the Director of National Intelligence approved the proposal. ICREACH was rolled out as a test program by late 2007. It's not clear when it became fully operational, but a September 2010 NSA memo referred to it as the primary tool for sharing data in the intelligence community. "ICREACH has been identified by the Office of the Director of National Intelligence as the U.S. Intelligence Community's standard architecture for sharing communications metadata," the memo states, adding that it provides "telephony metadata events" from the NSA and its Five Eyes partners "to over 1000 analysts across 23 U.S. Intelligence Community agencies." It does not name all of the 23 agencies, however.

The limitations placed on analysts authorized to sift through the

vast data troves are not outlined in the Snowden files, with only scant references to oversight mechanisms. According to the documents, searches performed by analysts are subject to auditing by the agencies for which they work. The documents also say the NSA would conduct random audits of the system to check for any government agents abusing their access to the data. *The Intercept* asked the NSA and the ODNI whether any analysts had been found to have conducted improper searches, but the agencies declined to comment.

While the NSA initially estimated making upwards of 850 billion records available on ICREACH, the documents indicate that target could have been surpassed, and that the number of personnel accessing the system may have increased since the 2010 reference to more than 1,000 analysts. The intelligence community's top-secret "Black Budget" for 2013, also obtained by Snowden, shows that the NSA recently sought new funding to upgrade ICREACH to "provide IC analysts with access to a wider set of shareable data."

In December last year, a surveillance review group appointed by President Obama recommended that as a general rule "the government should not be permitted to collect and store all mass, undigested, non-public personal information about individuals to enable future queries and data-mining for foreign intelligence purposes." It also recommended that any information about United States persons should be "purged upon detection unless it either has foreign intelligence value or is necessary to prevent serious harm to others."

Peter Swire, one of the five members of the review panel, told *The Intercept* he could not comment on whether the group was briefed on specific programs such as ICREACH, but noted that the review group raised concerns that "the need to share had

gone too far among multiple agencies."

– – –

Photo credit: Alexander: Carolyn Kaster/AP Photo; CIA Headquarters: Greg Mathieson/Mai/Mai/The LIFE Images Collection/Getty Images

– – –

Documents published with this article:

- CIA Colleagues Enthusiastically Welcome NSA Training
- Sharing Communications Metadata Across the U.S. Intelligence Community
- CRISSCROSS/PROTON Point Paper
- Decision Memorandum for the DNI on ICREACH
- Metadata Sharing Memorandum
- Sharing SIGINT metadata on ICREACH
- Metadata Policy Conference
- ICREACH Wholesale Sharing
- Black Budget Extracts