

The Value of Open Source Information: Two Military Intelligence Coups by the Web

Matthew Burton

matthew.s.burton@ugov.gov | www.matthewburton.org

October 3, 2008

Recently, I was a panelist at the Director of National Intelligence's Open Source Conference. The title of my panel was "Young Analysts Talk About the Value of Open Source." The intelligence field's definition of "open source" is different from what you might think: all it means is "information derived from public sources": newspaper articles, television broadcasts, Web sites, etc.

To outsiders, it might seem odd to have a conference about this: doesn't everyone understand the value of information? But when your desk has piles of secrets stolen from the enemy, it's understandably difficult to spend time reading about things the whole world already knows. And because those secrets are transmitted over private, physically impenetrable networks, the Web isn't available at everyone's desk. So the Intelligence Community is slow to realize the power of publicly available information in anticipating threats.

Judging by the product booths in the exhibit hall, the conference revolved around helping analysts navigate the Web, and harnessing Web content en masse so that it can be delivered to analysts on a non-Web network. (This is an important point to emphasize to readers outside the intelligence world. Analysts have access to infinite stores of foreign newspapers and news broadcasts, but it is stored privately, copied from a public medium to a private intranet with an interface that looks more like a card catalog than the Web.)

What to do with all of that information? Is open source intelligence somehow special or different from classified sources? The title of the panel implied that it is, and that I would have a unique take. I do.

I think it's wrong to think of "open source" as a type of information. Instead, we should view it as a manner of analysis: even if your original source is public, you can't do anything special with the information if you treat it just like all your other classified sources. And how you treat the information begins and ends with your network. There are open networks and closed networks, and when public information is placed on a closed network, it is no longer possible to do open source analysis. Whether classified or unclassified, your network is the only thing that can make information more or less valuable, as the network is the key to allowing open source analysis.

Let's take an old piece of classified satellite imagery as an example. It's been sitting in a file cabinet for years, and nobody has looked at it. Today, it is declassified: take it out of the file cabinet, slap "UNCLASSIFIED" on it, and put it back into the filing cabinet. Maybe it eventually winds up in another filing cabinet at the National Archives, whereupon it officially becomes "open source." When that happens, the image is no more valuable as it was the day before.

Wait: take that image back out, scan it, and put it on the Web. Now you've got something interesting: people will see it, talk about it, post it in other places, compare it with other images, and even alter it. All of these activities will make that image much more valuable than it ever was sitting in the file cabinet. The more open your information network is, the more valuable your information will become.

This rule also applies in reverse: take a video from YouTube and put it onto a closed network (pick any of the several agency-level networks that are only available to select intelligence analysts), and it instantly loses a lot of value. If intelligence agencies really want to use open source intelligence, *they should open their networks instead of simply offering infinite amounts public information in a closed environment.* After all, it is the network that decides whether information is open or closed.

What constitutes an "open" network? Here are two stories that demonstrate how the Web's openness adds value to information. I picked these stories precisely because they have clear military intelligence applications.

Singularity: The Iran Missile Photoshop Snafu

There is only one Web. All Web users are in the same place, and so is all of the Web's data. But that's not the case in the Intelligence Community: many CIA analysts work on one network, NSA analysts on another, etc. So if an NSA analyst posts a comment in response to an intelligence assessment (this is not always possible, but that's another matter), their CIA counterpart probably won't see it. This fractures the community, and it needs to stop. It's as if your city's phone network only allowed you to speak with people in the same area code. (For more about this, read about Metcalfe's Law.) Here's a real-world example of how multiple, fractured communities can keep you from making the most of information:

On July 9, 2008, the Iranian military test-fired some missiles. Then they posted photos of the test on their Web site:



Agence France-Pressé distributed this photo, and the next morning, it ran above the fold all over the world:



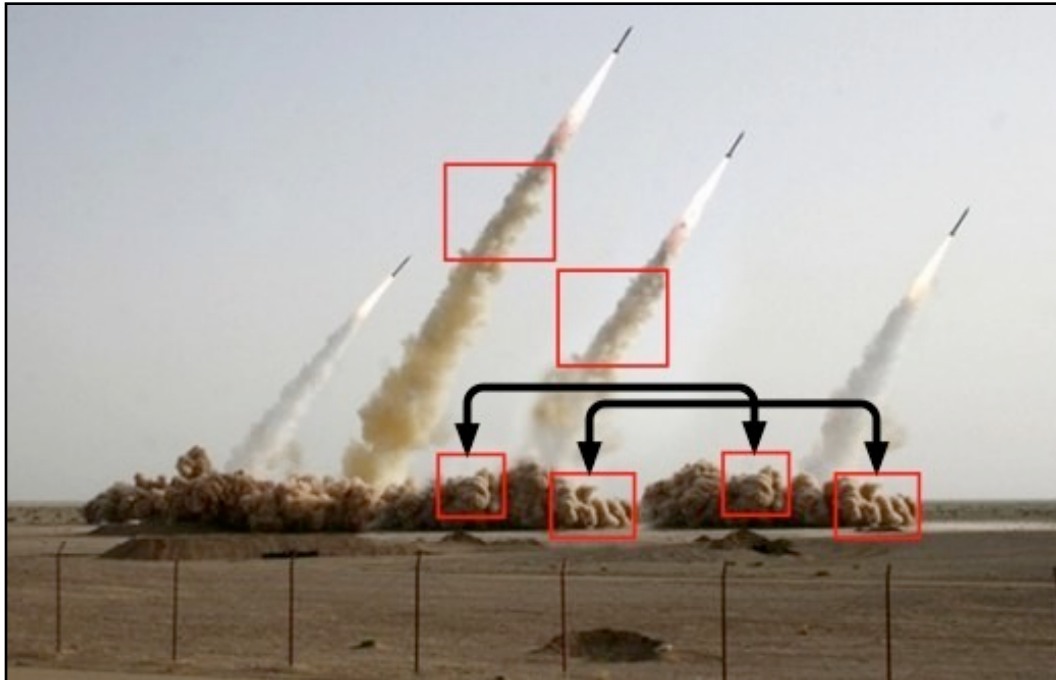
About the same time, TV networks were showing video of the launch:



It's easy to spot the difference when I put them right next to each other: the photo was altered to show four missiles, when only three were launched. The Web caught the mistake on Wednesday evening, long before the photo was on newsstands the next morning. Why didn't the press catch their mistake in time? Because "The Press"--which we refer to as a single community--was instead acting like multiple, disconnected communities relying on mutually exclusive information. The newspapers were paying attention to photos, the broadcasters were paying attention to video, and neither of them were watching the Web. Had all the journalists been working with all of the available information (just like that Web user who caught the error), the newspapers would have run a different photo. Maybe this:



Or maybe even this, which was posted on the blog Little Green Footballs at 6pm Pacific. The *Times* had not even settled on its incorrect front page, and the Web had already found the real story and done a full analysis:



The behavior of the journalists in this case is more or less how intelligence analysts work: some people have access to some documents, and other people have access to other documents. The behavior of Web users, on the other hand, exemplifies open source analysis. Their achievement should teach us that when people work as a single community and share access to the same information, good things happen. People feed off of one another's insights, and the information becomes more valuable. By sharing your data, you learn things you never would have learned by hiding it.

Freedom: The Jim Gray Rescue Effort

The Web is fast and free: fast, in the sense that you can make things happen overnight; and free, in the sense that you don't need the boss's permission to implement new ideas. This was no more apparent than in the case of Jim Gray.

Jim Gray was a database pioneer based in San Francisco. He disappeared while sailing in the Bay area last January. He was never found.

But his friends, many of whom were Silicon Valley magnates, did their best to try. One of these friends retasked DigitalGlobe imagery satellites to shoot fresh photos over 3,500 square miles of sea. The result was 1400 gigantic images. So coders automatically split the images up into 560,000 smaller ones that could be easily reviewed; someone else contributed a computer program that automatically sharpened all of the images.

Now that they had over a half million pictures, they needed a way to solicit volunteers to look at them. Amazon.com already had a ready-made solution, so the images were posted on a special page that let virtual volunteers cull through these images in search of Gray's sailboat. 12,000 volunteers reviewed all 560,000 images three times each.

All of that happened in three days. They didn't find Gray's sailboat. But it was the largest search party in history, on top of being simply amazing.

The Web's speed and freedom made it possible. Had this job been given to an organization with a closed network, this never would have happened.

What about your organization's intranet? Could your IT team split up those images into easy-to-consume portions? Could they write a script that automatically sharpened all 560,000 images? Could they create a tool that lets the rest of your workforce review the images from their own desks? Could your servers handle the load? And could all of that be accomplished in three days? In order to do that, you need to have the technical talent on hand.

Once you have the talent, those people need permission to work freely and create solutions on the fly; improvisation is essential during emergencies. Do they have that permission? Or would your org have to submit justification statements, get contract managers to approve personnel reassignments, and run each line of code past the infosec people before finally getting something live?

On the Web, if you have an idea, you just do it. This is a philosophy and capability the Intelligence Community (and any large, data-driven organization) should keep in mind if it wants to fully exploit its data. Analysts will always have too much information to process and too little time in which to do it (and focusing the "open source intelligence" conversation on culling Web content contributes to that problem). To have any hope, they need an IT and bureaucratic infrastructure that will accommodate improvised solutions.

Again: if you want to make the most of open source, get an open network that lets analysts contribute their own material. Put all of your analysts onto one large network instead of letting them work on lots of small ones. Hire some developers and familiarize them with the analysts' information problems. Then, set them free and let them experiment.

Thanks to DanF for the pointer on the Iran story.