Internet Exploitation – Springboard to a New Open Source Business Model
Thinking About the Internet in an Internet Way

By [redacted] ——— CIA b3, ODNI b6

The Internet has figured prominently in Community open source collection planning since the mid-1990's. In one form or another, strategic objectives aimed at bringing the cyberworld into the intelligence orbit have employed such verbs as "harness," "capture" and "exploit," all of which indicate the intention to "get a handle" on the Internet as if it were a medium like other media which, according to a long-established business model, collectors selectively cull into a controlled environment for processing, transformation, and redistribution.[1] This process, which is broadly defined as "open source exploitation," rests on two major premises:

1. The open source arena, though incredibly vast, is segmented in a way that makes the culling process ("selection" in the usual parlance) possible. Television programming is channeled, with known starting and ending times for individual offerings. Newspapers appear at predictable intervals, with discrete articles. Specific radio frequencies are associated with specific broadcasts at specific times. The content offered by these media in specific contexts (dates, times, frequencies) can be characterized (source description). Moreover, the information flow is regulated by recognized mediators (columnists, anchorpersons, reporters) who act as gatekeepers for us, the audience. The lanes are clearly marked. Thus it is possible to know where to look for open source information.

2. The open source exploitation process is "driven by requirements." That is, the objective of open source collection and production is to answer certain information needs that are articulated (through formal and informal processes) into actionable chunks, the more specific and detailed the better. Thus it is possible to know what to look for.

The premises that one can know what to look for in open sources and where to look for it underlie the Community's basic open source business model, which is **monitoring.** We identify those locales in the open source universe which we have determined are likely to contain the discrete pieces of information we seek, keep our eye on them, and pluck out the information as it shows up (not unlike setting out crab pots in areas where these crustaceans are known to congregate). What we do with this information once we capture it may vary (translate, analyze, repost, synthesize, summarize, etc.), but the monitoring model of information capture is constant. To use a recently popular metaphor, monitoring is a **gathering** (as opposed to **hunting**) activity. Significantly, it is also **reactive** (we speak in terms of "responding" to requirements, "answering" intelligence needs, "following" topics, or reporting on media "reaction" to events).

[1] This model has been applied for decades to foreign media collection by the Foreign Broadcast Information Service, the Intelligence Community's leading open source collector.

Enter the Internet, a whole new world of open source. The Community expects open source collectors to deal with the Internet as with other publicly available information sources. At first glance, the main challenge posed by the Internet is one of volume; in terms of the basic monitoring model, the Internet would seem to fit in as simply another medium. It is segmented (domain names, URL's), has describable content ("jihadist website"), and can be regularly accessed, i.e. monitored. Unsurprisingly, the Internet category that received earliest and most thorough attention from collectors was online newspapers. Experience with using the Internet to track down intelligence information, however, leads us to conclude that the monitoring model is not optimal in this case. This conclusion rests on several factors, including 1) the nature of the Internet as an information system; 2) in a larger context, the ongoing global transformation of the way in which news and other types of information is generated and propagated; 3) an altered concept of the way in which open source can be used to address intelligence questions. The following remarks explain this reasoning in some detail and suggest a new model for Internet exploitation as a component of the open source discipline, commensurate with its vastness and extreme fluidity. This model, as the subtitle of this paper indicates, requires that we as intelligence collectors and analysts approach the Internet in an Internet way[2] as opposed to, say, a newspaper way.

1. What is the Internet? Good question. The issue here plays on many levels (technology, epistemology, legality, economy, etc.). Let us assume that we can conceive of the Internet as a discrete entity. As an "open source" the Internet has many features that set it apart from other open sources.

a. It is interactive. Not everyone can appear on television or publish a newspaper article, but anyone can engage in discussion on the Internet. From almost anywhere.
b. It is instantaneous. The speed of light is a very important factor in the way information travels around the Internet.
c. It is nonlinear. Whereas traditional media outlets process information in station-to-station mode ("the review process") before it is released, information on the Internet travels in all directions at once.
d. It is unrestricted. Whereas in many places it is necessary to stage a coup in order to get access to television broadcasting, one can always find somewhere on the Internet from which to send one's message. Regulation, filtering and censorship exist, but can be evaded more easily than is the case with traditional media.
e. It is low-cost. Traditional media, especially in the commercial arena, carry infrastructure costs that have an impact on the nature of the content carried by these media. The cost of content delivery on the Internet approaches zero by comparison.

One could continue in this vein for quite some time, but the point is evident. The Internet is not so much a medium (in the sense of an information conveyance) as it is an

[2] Thanks to FBIS manager [redacted] for this pithy expression.

CIA b3, ODNI b6

environment, and one does not so much monitor an environment as one navigates it. The fact that the Internet does not share the conduit structure typical of traditional media makes the information transmission and exchange process on the Internet quite different, and compels us as open source collectors and analysts to approach it differently.

2. A New Kind of Information Sharing. As the Internet has increasingly become the milieu of choice for communication and information transmission, it comes as no surprise that a new paradigm for this kind of activity has arisen, especially within the realm of news and journalism, that reflects the capabilities offered by the Internet. One manifestation of this paradigm has been called "participatory journalism,"[3] for which web logs, or blogs, have become the leading vehicle. Another manifestation of the new information-sharing paradigm is the wiki, simply described as an online encyclopedia to which multiple authors contribute (see http://en.wikipedia.org for a well-known example). Wikis are similar to blogs in that they are updated in real time and feature contributions from throughout the cyberworld. Wikis differ from blogs in that they aspire to be permanent records of agreed-upon knowledge rather than somewhat ephemeral flashes of opinion or insight.[4] Other familiar forms of Internet information exchange include bulletin boards, chat rooms, newsgroups, discussion forums, etc. Two features most prominently differentiate this information sharing paradigm from that of traditional media: 1) Its content is not media-defined, and 2) It is **disintermediated.** Both of these features have a bearing on how open source collectors and analysts should approach the Internet.

Traditional media were described above as information channels or conduits, each with its own distinctive features and attendant infrastructure, including designated mediators, such as reporters and "talking heads," who define for the audience the content that flows through the channel ("All the news that's fit to print"). Especially in the case of commercial media, the nature of these media themselves and the costs of infrastructure determine the content. Commercial television stations, for example, require content that is suitable to the medium (moving images as opposed to text or audio) and that will attract a sufficient audience to offset infrastructure costs. The cost driver makes it imperative to offer content of any sort, which is often recycled since there is only so much capacity to produce original content. This type of medium-defined content landscape is easily mapped, and open source collectors can focus their attention on (i.e., **monitor**) selected areas while ignoring others known to be unproductive or redundant. By contrast, the low-cost Internet does not impose such strictures on content. The cost of delivering content on the Internet approaches zero. There are no conduits to maintain, and with such capabilities as hyperlinking all content formats are equally appropriate to all contexts. Though there is obviously a technical infrastructure supporting the web, it is largely transparent to the information exchange process and impacts little on content (except in such areas as bandwidth, which is becoming less of a constraint as time and

[3] See Rebecca MacKinnon, *The Worldwide Conversation: Online Participatory Media and International News* (The Joan Shorenstein Center on the Press, Politics and Public Policy Working Paper Series #2004-2), 2004.

[4] See [redacted], *The Wiki and the Blog: Toward a Complex Adaptive Intelligence Community*, 2004 (Galileo Award winner) for a discussion of wikis and blogs in the intelligence context.

CIA b3, ODNI b6

technology progress).

Related to the absence of traditional information channeling on the Internet is the disintermediation of the mass information transmission process. In economics disintermediation means simply the elimination of the middleman, and is often used to characterize the advantages of e-commerce over traditional retail sales, stock transactions, etc. The same applies to the Internet news and information marketplace. Traditional media have brokers, or gatekeepers, who determine what information gets through to the audience and how this information is "spun." These mediators exist as both behind-the scenes actors (editors, program directors) and as visible authorities (anchorpersons, columnists, hosts). The Internet has largely done away with such mediation. Internet users have direct access to a virtually unlimited set of news and information providers. Blogs, wikis, bulletin boards, discussion forums and chat rooms provide opportunity for individual citizens to exchange views and information, and hence to form opinions, directly with one another on issues both large and small to an extent that was not possible in the pre-Internet age. In this environment a collection strategy of "monitoring the gatekeepers" is inadequate to the task of coping with the unchanneled information traffic pattern. In other words, it is fruitless for intelligence collectors to act as gatekeepers (monitors) ourselves for the Community in an open source environment without gates.

3. The Internet and Intelligence Questions. The Internet is an effectively boundless, constantly changing, and wholly participatory environment in which information is crafted and molded by the very action of public interchange, rather than by editorial ensilage. One cannot effectively participate in this environment in a static manner, and monitoring is an essentially static activity. One "surfs" or "navigates" the Internet, following the flow of information as it leads from place to place, changing shape all the while. To return to our aquatic metaphor, traditional media as a whole can be likened to a network of rivers or canals, in which the information flows within well defined channels. One can set one's monitorial "traps" within these channels at strategic spots and be fairly certain that all of the information streaming within the banks will pass through one's filter. The Internet, however, is a global information ocean. There are certain known or predictable currents (like the Gulf Stream) that one can regularly fish, but the vast majority is unknown, among other reasons because it is constantly changing. In order to maximize one's catch of fish, or one's capture of information, one has to venture into this unknown and follow the schools where they lead. To set out one's monitorial crab pots in just a few locations would be to risk missing much indeed.

This strongly suggests that the monitoring model of open source exploitation practiced by Community collectors is at best poorly suited to the Internet. The mutability and fluidity of the Internet as an information matrix, and the cacophony of participatory voices that can drive the formulation and flow of information, dictate that one must actively pursue information (hunt) rather than lie in wait for its passage through a media channel (gather). This in turn suggests that the "requirements driven process" that underlies the current monitoring model is also not optimal when applied to Internet exploitation. This is not to say that exploiting the Internet for open source intelligence should not be linked to intelligence requirements. The point is rather that instead devoting our entire Internet effort to monitoring in search of elusive discrete pieces of information (pieces of the

puzzle) we should concentrate on formulating broad intelligence questions (or adopting such questions as posed by others), identifying Internet trails that might shed light on these questions, and navigating these trails until they lead to answers. Too often the "requirements" that intelligence collectors endeavor to meet are expressed as nouns or short noun phrases, digestible but disconnected chunks that we can catch in our media monitoring nets. To cast these same nets in the open Internet is much less effective, like panning for gold in the ocean. This is not purely a question of technology, though technology can certainly aid in the hunt. It is a question of methodology. If we approach intelligence requirements proactively as broad questions that need answering (How will China meet its increasing energy needs over the next decade? Will the rising generation of Palestinians approach issues with Israel differently? Is Cuban anti-Americanism a model for Venezuela?) we can intelligently navigate the Internet (assisted by data mining and visualization technologies) *in search of entire answers rather than puzzle pieces.*

4. What To Do? The main conclusion derived from the preceding analysis is that the Intelligence Community's open source program must apply a new business model to the Internet, one that accounts for the nature in which information is generated and propagated in this environment. This new model, it was suggested, should be based on a broad research agenda rather than a monitor-and-react posture. To set this agenda and effectively follow it requires that we make certain adjustments in how we, as open source exploiters, think about products, requirements, "turf," and knowledge sharing. These adjustments do not need to be costly. The point that this is a matter primarily of methodology, not technology, cannot be overstated. Technology aside, *we can make significant improvements in large-scale Internet exploitation just by changing our approach.* The following recommendations are made from this standpoint.

5. Recommendations.

A. Set a Top-Level Open Source Research Agenda. The prevailing view in the Intelligence Community today is that open source is first and foremost a collection (gathering) discipline. Intelligence research and analysis is seen as an all-source activity, with the collection requirements process designed to provide fodder for this activity.[5] This requirements process deconstructs top-level intelligence questions into component pieces of information that collectors can identify, obtain, and feed back into the analytic process for reconstruction.[6] In current practice open

---

[5] This way of thinking is evident in the influential WMD Commission Report (*The Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction Report to the President of the United States, March 31, 2005).* The recommendation on creating an Open Source Directorate is found in the chapter on Collection and comprises four pages of discussion (pp. 377-80). Open Source analysis, on the other hand, is briefly treated with the following recommendation in the chapter on Analysis: "The DNI should create a small cadre of all-source analysts—perhaps 50—who would be experts in finding and using unclassified, open source information." (p. 395)

[6] The atomization of intelligence questions into discrete "collection requirements" and "intelligence priorities" that can be arrayed in regional/topical cellular matrices, and the deleterious effect of this process on the ability effectively to answer these questions, is a topic for another paper.

source (especially foreign media) analysis is too often framed only by what is obtained by collectors according to this requirements process. The preceding sections argued that the open source arena as embodied in the Internet is too vast and fluid to be exploited effectively by piecemeal collection. It is also commonly acknowledged that open sources by themselves contain answers to a large percentage of today's intelligence questions.[7] A Community-level **open source research program**, unobstructed by stovepipes built around regional areas of responsibility, sources and methods, mission specificity and disciplinary tradecraft, would enable us to use open sources more effectively to address important intelligence questions directly, without going through the encumbrances of the current disaggregated collection tasking and requirements model, which lengthens the analysis production process and promotes "information leakage" as collected bits of open source information move along the processing chain.

B. **Use Our Internet Exploitation Capacity To Focus More on the "Unknown Unknown."** The whole point of having an intelligence apparatus is to prevent harm to the national interest. When the intelligence apparatus fails to do so, this most often results in some kind of embarrassing or damaging surprise. Intelligence success, therefore, prevents unwelcome surprise. To prevent surprise one must find and explore the "unknown unknown," i.e. that arena of reality that we are not even aware of in which potentially damaging surprises reside. The Internet epitomizes the unknown unknown. Therefore, Internet exploitation should involve discovering and analyzing potential threats that we do not yet see and which, by definition, are not specifically articulated in any existing requirements framework. To confine Internet exploitation to preexisting requirements frameworks would be to restrict it to the realm of the "known unknown" (gaps) or even the "known known" (corroboration) and diminish its capacity to help avoid damaging surprise. Therefore, the open source intelligence collection, research and analysis cycle should accommodate some degree of non-reactive exploratory Internet research, guided by overarching intelligence questions but not confined to narrow and specific intelligence requirements as they are usually understood to pertain to open source. Among other things, this would mean that "responsiveness to existing requirements" or some such formulation would not be a good metric for assessing Internet-based open source analysis. A new and more forward-looking criterion for evaluation would be needed.

---

[7] This has been the case for almost 60 years: "A proper analysis of the intelligence obtainable by these overt, normal and aboveboard means would supply us with over 80 percent, I should estimate, of the information required for the guidance of our national policy." (Memorandum Respecting Section 202 (Central Intelligence Agency) of the Bill To Provide for a National Defense Establishment, Submitted by Allen W. Dulles, April 25, 1947. *Hearings Before the Committee on Armed Services, United States Senate, Eightieth Congress, First Session on S. 758 Part 3, April 30, May 2, 6, 7, 9, 1947.* United States Government Printing Office, Washington, D. C. 1947)

6. Conclusions. The Intelligence Community comprehends the Internet as the business of the open source discipline, but by its nature the Internet breaks the existing open source paradigm, which is based on traditional mass information media. This paper has argued that in order effectively to use the Intemet to answer intelligence questions, and to anticipate as-yet unseen threats, a similar paradigm break in open source exploitation is needed. Specifically, the monitor-and-react model, adopted for traditional media but poorly suited to the Intemet, would usefully cede its place to a research-oriented model that proceeds directly from fully formed intelligence questions rather than from piecemeal intelligence requirements deconstructed from these same questions. There is no better time than the present, on the eve of the creation of a new Open Source Center, to implement this new paradigm.