Open-source Intelligence Gathering and Open-analysis Intelligence for Biosecurity

Geoff Grossel[1], Aidan Lyon[2] and Mike Nunn[3]

[1]Department of Agriculture, Fisheries and Forestry, Canberra, 2601, Australia

[2]University of Maryland, College Park, Maryland, 20742, USA

[3]Australian Centre for International Agricultural Research, Canberra, 2601, Australia

Introduction

Open-source intelligence is the analysis product of publicly available information that has been collected, sorted and archived. *Open-analysis intelligence* is the analysis of freely available information that is done out in the open by an engaged user community. A national security agency of some country may mine Twitter for data and analyse this information to predict social uprisings. This would be an example of open-source intelligence, but not of open-analysis intelligence, because the analysis of the information is confidential. An example of open-analysis intelligence is rdtn.org, a website that analysed crowdsourced measurements of radioactivity in Japan during the 2011 earthquake to map actual levels of radioactivity across the nation.

We have developed an online system, AquaticHealth.net, that performs both open-source intelligence and open-analysis intelligence with the goal of tracking and forecasting aquatic animal diseases. The result is a an application that provides real-time and relevant information to decision-makers in a practical way with the ultimate goal of providing early warning and improving response times to aquatic health issues and disease threats. This intelligence platform is also being expanded

to cover animal and plant biosecurity and will eventually integrate to form the International Biosecurity Intelligence System – IBIS.

Public health intelligence gathering systems already exist on the Internet, for example HealthMap and ProMED (Madoff 2004). These websites have a "One Health" perspective (One Health 2011) that encompasses diseases of humans, animals and plants. Others like WAHID (World Animal Health Information Database) operated the World Organisation for Animal Health (OIE) focus on animal diseases. Such websites are emerging as vast and invaluable biosecurity intelligence resources.

These systems offer broadly similar information aggregation services with subtle differences around a central theme of aggregating both open-source and contributed content. They are heavily biased toward news aggregation, and with the exception of Google Flu Trends and HealthMap's Flu Near You and Outbreaks Near Me app, they are noticeably light on useful analysis output. None of the systems provide tools (other than discussion threads, etc.) freely available to the user community to conduct any collaborative or open-intelligence analysis of the information gathered.

Furthermore, there are no systems solely devoted to aquatic animal diseases. The systems are slowly converging, and there is increasingly more collaboration—particularly between the two most prominent public health systems, ProMED and HealthMap. However, there will always exist an opportunity for boutique intelligence websites that are tailor-made to fill analysis gaps and offer much needed services, services that take the next step beyond simply aggregating and organising information and offer an open-analysis intelligence application with some effective yet easy to use analysis tools.

Our approach in designing our first model website, AquaticHealth.net, was to integrate the existing systems' components that would have the most utility for aquatic animal biosecurity, adding new tools where the demands required them, or where the opportunities emerged during development. Our resuting IBIS website, which hosts aquatic, plant and animal intelligence systems, represent a departure from pre-filtered and owner-controlled systems. Rather than a broad approach to open-source intelligence gathering, we give the user community as much control as possible over the gathering and organisation of information, and then also provide the means to produce meaningful outputs with open-analysis intelligence tools. This allows for a collegiate community to concentrate directly on issues relevant to them. We have called this approach *open-source intelligence gathering and open-analysis intelligence for biosecurity* with the aim of promoting crowdsourcing and active engagement in initiating and contributing to collaborative intelligence analysis, and control over scanning, gathering and content. Not all analysis needs to be crowdsourced or conducted openly online. The open-source information can be used at the discretion of decision makers for confidential analysis. This can be especially important when information is commercially sensitive, or there are trade implications, or when unfounded public perceptions and expectations need to be managed appropriately e.g., in the national interest.

Governments need to be mindful that responsive and effective analysis of rapidly reported information already in the public domain, with or without official approval, is a more transparent mechanism for responsive decision making than is private analysis of classified information—which is often criticized as resulting in avoidance of critical issues and delays in responsive action against disease risks. It is for this very reason that independently managed applications like ProMED exist today. Governments are also beginning to understand that transparency and being held to account for critical emergency decision-making are directly the result of prompt and proactive disclosure of facts. This understanding stems from the realization that the risks associated

with complex problems are greater when you know something but do nothing (for whatever reason), than if you publicly disclose the facts as they are known at the time and immediately anticipate, plan or initiate an appropriate response. For example, a competent authority may, for managing public perceptions, sensitivities and for strategic trade reasons, not release information associated with an emerging disease problem, such as oyster herpes virus, as the disease may not be nationally or internationally reportable, and because unfavorable public perception of the word "herpes" may seriously impact critical trade periods—i.e., the release of conditioned oysters during a short distribution window over the Christmas or summer holiday period. In the absence of prompt public disclosure the competent authority will not be held to account and thus may delay a critical and timely emergency response in favor of a short-term trade gain. Such a strategic delay is a false economy, because in the aquatic environment disease spreads rapidly and all too often results in establishment and spread of the disease and serious consequences for the industry at the regional and potentially the national level. Both sustained, healthy trade and responsive action to an emerging disease threat can co-exist when managed responsibly.

With nationally and internationally reportable diseases, most competent authorities are obliged to promptly report nationally and to the World Organisation for Animal Health (OIE) respectively via their established reporting mechanisms. Notwithstanding such obligations, if we scan the data held in the OIE's disease notification recording system WAHID, it is clearly evident this is not always the case. The risks associated with the spread of diseases in globally traded aquatic animals and their products from non-reporting member countries alone justifies the need for effective open-source, open-intelligence applications to detect risk early and take appropriate responsive action.

Designing AquaticHealth.net

Our staring point was to develop a dedicated aquatic biosecurity intelligence information gathering system by examining existing web-based systems to identify features that may be particularly important for aquatic biosecurity. We examined a range of criteria including search strategies, language search and translation capabilities, mapping capabilities and the structure and logic of information filters (Lyon et al 2011). The most significant ways in which the systems differ is their reliance on automated software vs. human beings for content gathering and analysis. Some like BioCaster (biocaster.nii.ac.jp) are fully automated, whereas ProMED is at the other end of the information collection spectrum and is completely human–based. We have opted for a balanced mixture of both automated and human crowdsourcing—an open call to an undefined group, usually composed of people appropriate for a kind of task, to contribute to an analysis or to solve a problem (Brabham 2008). We believe that this is the best approach to take to generate human–mediated analysis.[1] For a comprehensive analysis of the functionality and technology behind AquaticHealth.net and the subsequent IBIS application see Lyon et al (2012).

The technology for developing open-source, open-analysis intelligence applications is not new. What separates applications like AquaticHealth.net from, say, HealthMap is not the ability to collect and organize open-source data, it is the capacity to offer the user complete and un-filtered control of searching, and then provide a set of easy to use vertically integrated applications to generate outputs in the form of intelligence reports from which health management decisions relating to disease risks can be based. An essential feature of an open-source intelligence application is the ability to keep abreast of the latest technology in the intelligence-reporting domain. To achieve this, the applications must resource and sustain a dynamic research community. Integrated analysis tools must continually be researched, constructed and uploaded as early as possible so that feedback from

[1] See also Floridi (2009) for a more theoretical assessment of ‘Web 2.0’ and ‘Web 3.0’ technologies.

the user community can direct continual improvement of the systems or determine the usefulness, relevance and life span of the tool.

Wiki-Based Open Intelligence Analysis

Our research and development strategy is based around the adage *fail early and often*. We initially attempted to encourage user engagement by contributing to wiki pages as a simple form of open-analysis intelligence. A range of diseases were set up and seeded with rudimentary information around which narratives could be built, they also included extensive reference to the site's reports. Topics were originally selected for placement into two categories, 1) specific diseases and, 2) other topics. We seeded approximately 20 wiki entries using aquatic diseases of concern as subjects, mostly diseases listed by the OIE. Each wiki also included a forecasting section, in which users could record educated guesses about where diseases are likely to spread next. The open source nature of the wiki was intended to allow users to have fluid debates about the forecasts. All revisions of the wiki entries were recorded and made viewable by all users and topics were only limited to the imagination of the system users.

By tracking the use of the wiki's over time it became apparent that the registered user community, now growing substantially in number, did not fully engage and contribute, and generally showed little interest. We concluded the wiki's were not working in the way we had anticipated. Nielsen (2011), an advocate of open science, proposes that open science wiki's often fail because people, particularly those with careers in science, do not want to contribute to something of little or no benefit to the advancement of their careers. People would rather write articles, even substandard articles, for peer-reviewed journals, as this kind of publication is far more beneficial to them individually. Nielsen proposes this is not due to some kind of selfishness of the individuals, but rather the consequences of the wrong incentive structures being in place. A large component of

hiring decisions, tenure reviews, and promotions are the number of publications an individual has accumulated. Furthermore, open contribution may compromise contractual research obligations and intellectual property rights.

Nielsen's explanation for failed open science wikis may partly account for the failure of our wiki's. However, the failure cannot so much be attributed to the wiki's themselves, but to the original "seed" topics we chose. It should be noted that the failure of the wiki's was not complete and some aspects of them were successful. By testing the various wiki topics, we found community engagement was successful if the topic, or the subject within a specific disease wiki, was consistent with what we were trying to accomplish—i.e., active and collaborative open- analysis intelligence of emerging disease issues, capturing emerging disease trends, forecasting, and capturing other interesting hot issues and topics that can be tracked and analysed with the goal of improving early warning and responsive action to aquatic animal health issues, threats and disease outbreaks. Trying to engage registered users to update general disease information in the disease-focused topics was never the intention of how an intelligence gathering and analysis website was designed to function, hence the failure of wiki topics that were out of scope, and the need to develop an alternative online application that is appealing, collaborative, yet simple. Nielsen (2011) argues that one of the challenges in making open science succeed is the creation of a more user-friendly online tool that encourages participation throughout the network, such as the open science website (http://www.openscience.org).

Blog-Based Open Intelligence Analysis

Like Nielsen, we believe our challenge is to find the most appropriate online format to create engagement in open-analysis intelligence. We have therefore moved to a new, hybrid wiki-blog analysis platform called "Emerging Issues". It has the same goal of promoting open- analysis

intelligence, but does so by focusing on what the users of the site care about most: disease outbreaks or events that might be disease outbreaks, emerging issues, and forecasting disease outbreaks. Emerging Issues was born out of the failure of the original out-of-scope wiki topics with the exception of a wiki titled “Emerging Diseases for 2011” and from the forecasting sections of the disease-specific wiki’s. The simple and attractive core function of these wiki features was that they achieved our primary goal; they successfully captured events for further analysis and resulted in user engagement.

Entries in the Emerging Issues wiki-blog feature a title and date, a window for all users to provide content, a forecasting section, links to related reports, share functions, comments, and a map. The systems have now stabilized on a functional and meaningful model that provides an easy to use analysis tool for our registered user community to construct intelligence reports. The forecasting applications retained from some of the older wiki’s in AquaticHealth.net proved to be accurate and potentially very useful for improving biosecurity planning and health management in aquaculture and fisheries. These forecasts have been transferred to the most relevant Emerging Issues wiki-blog. For example, oyster herpes–related summer mortality in Pacific oyster culture was used as one of the original wiki topics. As emerging patterns in the global spread of the disease became apparent, the information was fed into the “OsHV–1” wiki. The resulting prediction was that oyster herpes presents the greatest single disease risk to the Australian Pacific oyster industry. Eight months after this prediction was made, Australia and New Zealand Pacific oyster industries both incurred outbreaks of oyster herpes related summer mortality. In more examples, similar predictions were made prior to disease outbreaks in bivalve mollusc aquaculture in the US and Vietnam.

To date, all four forecast made during 2010-11 have proven to be correct. The latest Emerging Issues wiki-blog forecasts are: the forecast that the marine fish disease known as VNN (viral

nervous necrosis) will cause significant problems in the emerging grouper aquaculture sector throughout the Southeast Asia region during 2012-2013. This forecast is scheduled to play-out over the long-term, however, recent scientific reports from China indicate this forecast was correct and has now run its course. Another forecast made in February 2012 was that amoebic gill disease (AGD) will cause major problems on salmon farms in Scotland and eventually spread to Norway in the advent of favourable warm conditions sometime in the next three to five summer seasons (e.g., the summers of 2012-2014), however, AGD has already re-emerged in southern regions of Norway in December 2012, the forecast was therefore accurate, but unexpectedly early. Emerging Issues currently arise at a rate of one to two issues every two months and are all updated on a regular basis when new information becomes available. Our research focus for future analysis tools is to provide tools that are effective in producing meaningful outputs, are easy to use, and promote collaboration within the user community. We are investigating forecasting tools, predictive modeling and cluster analysis, geo-spatial analysis, Delphi analysis and network analysis. The challenges we face with the tools are similar to the problems encountered with our original wiki i.e., we will need to transition the research into a user-friendly interface, load the tools to the website as early as possible, and refine them according to feedback from users.

Content and Analysis Quality

An often-encountered concern about our main development site, AquaticHealth.net, mostly by practicing/publishing aquatic animal health scientists, is the quality of the information it collects. Scientists and aquatic animal health experts in government roles argue that to be of any value, disease outbreak information must be either an evidenced-based, peer-reviewed scientific publication or officially endorsed information released by government or an immediate disease notification from the OIE.

It is not the intention or function of our online intelligence gathering and analysis tools to be a filter that are selective about the quality of information that enters the sites, the website's information collection capabilities function to collect all the available electronic information coming into the Internet, from tweets, to YouTube videos, to scientific journal articles, to OIE immediate disease notifications, to ProMed reports etc. Our tools intelligently, with the help of humans, sort the information, add-value to it and archives it so it can be retrieved later when an intelligence analyst is attempting to determine trends in the data, such as an emerging disease trend over a two year period. For example, the information can be used in a similar way to a journalist investigating a story. An emerging disease trend may start by alerting the analyst to a disease outbreak problem of unknown etiology, however, the initial information may have a sketchy, unverified source such as a YouTube video file and a tweet uploaded to the Internet by an interested bystander. Nevertheless, information like this may be crucial in confirming the first event in a series of disease events leading to a new and concerning emerging disease problem.

Our online tools have the capability to crowdsource quality judgments through the 4-star ranking system of reports and user comments attesting to the quality of the information that may eventually assist in later analysis. However, even if information is of poor quality, analysts can provide early disease trend information to aquatic health decision makers so that preparations for responsive action can be made. The intelligence analyst can interrogate the veracity of the information and verify the validity of the source during the course of investigating a disease trend or other issue and provide comment and/or rank accordingly. More often than not, disease outbreaks in the aquatic environment for example, will spread quickly to their natural limits with devastating impacts and consequences. Expensive emergency responses are then centered on containment, disposal and decontamination, rather than directing valuable and finite resources toward prevention. Aquatic health managers are keenly aware that consideration of timely information and early actions that are responsive and preventative are the primary and most effective considerations when managing

biosecurity in the aquatic environment, and that waiting for high-quality reports to be published, such as scientific papers or official government reports, is an ineffective approach to responsive aquatic health management.

Another criticism we have encountered from aquatic animal health experts is the issue of peer review, especially of our forecasts and analysis featuring in the Emerging Issues section of the website. We do have peer review, but it isn't the standard scholarly peer review process that publishing scientists are familiar with—i.e., the traditional process of sending out articles to a narrowly defined field of, usually anonymous expert reviewers. All our information is open-source in the public domain and viewable by anyone on the Internet at any time. Registered users can, either anonymously or not, comment or criticize anything without approval from a moderator in an open and transparent manner. Non-registered users can do the same, but someone within a trusted group in the network approves the comment first. All review comments are available for scrutiny and any registered user can make alterations, discuss improvements and access previous versions of the analysis to track and debate changes. We are mindful of creating an environment that does not shut down the debate and actively encourages an open discussion about all content. The open peer review process we have adopted is also a post-publication review process and is arguably more suitable for our purposes as we rely on open-intelligence analyses being collaborative, ongoing and updated as necessary, and directly targeted to the engaged demographic using our online tools.

There is potential for politically biased reporting to the site. Lobbyists for environmental groups have used the site to post information to all users via submission of disease news articles, however, the site's user community, while interested in browsing through all content in the context of reading the range of information available to them, are generally not the direct target for this kind of lobbying. As a result it has proven to be an ineffective information dissemination avenue for them that they no longer pursue. There are further ways in which this lobbying pressure can be mitigated.

Although the openness and democratic nature of the site allows the possibility of pushing an agenda, it also allows for effective crowdsourced agenda control. If a user starts promoting misleading information, other users can quickly comment on, or even edit that information to explain to everyone else that the information is contentious. We have had some users that joined the site to publish all articles from their own website; fortunately, the community quickly corrected this and unpublished all of the irrelevant reports. The overall effect was that all relevant reports from that site were published and brought to people's attention and all irrelevant reports remained in the raw data scan and industry news feeds. The system isn't foolproof, but no system is, and there is no reason to think that this system is especially prone to being gamed. We believe this is because a website of this nature is of little general interest outside the community of users. There is little opportunity to promote causes, increase sales or solicit money from the registered user community.

Conclusion

Our online intelligence gathering and analysis tools are an example of what can be achieved in today's technologically democratized world. Many of the key elements of our websites are outsourced. In their short history, our online applications have been used to capture emerging disease information, analyze and track disease trends, map diseases, organize data, perform basic predictive modeling, contribute to future health planning, provide biosecurity alerts, build biosecurity risk profiles and support responsive decision–making relating to imports and exports.

The power of raw data feeding into discussion and analysis in real–time is immense.

Our sites offer the capacity to intelligently process raw information in real–time with the added function of unlimited application. Our mission is to create openness in animal and plant health. We will work to establish a system of transparency, public participation, collaboration and trust. Openness will strengthen animal and plant health and promote efficiency and effectiveness in biosecurity. We hope to establish an effective open-intelligence community over time to build

critical mass for supporting the implementation of positive and responsive action (adapted from Obama 2009).

References

Brabham DC (2008) Crowdsourcing as a model for problem solving. Convergence: Int J Res New Med Tech 14:75

Floridi L (2009) Web 2.0 vs. the semantic web: A philosophical assessment. Episteme 6:25

Lyon A, Nunn M, Grossel G, Burgman M (2011) Comparison of web–based biosecurity intelligence systems: BioCaster, EpiSPIDER, and HealthMap. Transboundary and Emerging Diseases. Doi: 10.1111/j.1865-1682.2011.01 258.x

Lyon A, Nunn M, Grossel G, Burgman M (2012) Using internet intelligence to manage biosecurity risks: a case study for aquatic animal health. Diversity and Distributions.

Madoff LC (2004) ProMED-mail: An Early Warning System for Emerging Diseases. Clin Infect Dis 39:227-32

Nielsen M (2011) Reinventing Discovery: The New Era of Networked Science. Princeton University Press.

Obama B (2009) Subject: Transparency and Open Government, 23901. Federal Register /Vol. 74, No. 97 /Thursday, May 21, 2009 /Notices Memorandum for the Heads of Executive Departments and Agencies [Internet] [cited 2012 Mar 20]. Available from: http:// www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/

One Health (2011) 1st International One Health Congress [Internet] [cited 2012 Mar 17]; Available from: http://www.onehealth2011.com/