The Wildlife Disease News Digest and Global Wildlife Disease News Map

F. Joshua Dein[1,2], *fjdein@usgs.gov*, Megan K. Hines[1,3], and Christine M. Marsh[1,3]

1. Wildlife Disease Information Node - National Biological Information Infrastructure (NBII)
2. U.S. Geological Survey - National Wildlife Health Center
3. University of Wisconsin - Nelson Institute for Environmental Studies

With the increasing attention to the global surveillance of emerging infectious diseases, a number of different systems have been developed to extract useful information from open-source media reports and other unstructured text sources (HealthMap[1] Biocaster[2] GPHIN[3] etc.). This is most helpful to gain an awareness of events in areas without a substantial public health reporting infrastructure, as well as providing a more "real-time" recognition of events that may highlight the need for further investigation. [5] As it has been estimated that about 70 percent of recent emerging diseases have a wildlife component, it is important that report based monitoring systems focus on these species as well. To meet this need, the NBII Wildlife Disease Information Node (WDIN) has developed a paired set of Web based resources, the Wildlife Disease News Digest and the Global Wildlife Disease News Map, which provide a frequent textual and geospatial global snapshot of wildlife health events. To promote the one-health concept, this information is suitable for use in other reporting systems that focus on human and ecosystem health.

Methods

Wildlife Disease News Digest

To create the Wildlife Disease News Digest wdin.blogspot.com, over 125 RSS (Really Simple Syndication) Feeds, produced by national and international media sources, are used to harvest the news from the Internet. These feeds, which are electronic streams of links to Web content, are collected and organized into a formatted list using Google Docs docs.google.com. Yahoo Pipes pipes.yahoo.com, an open source tool that aggregates and manipulates Web content, filters this list using a custom set of over 100 keywords to eliminate irrelevant content, remove duplicates, and limit the output by publication date, as shown in **Figure 1**.

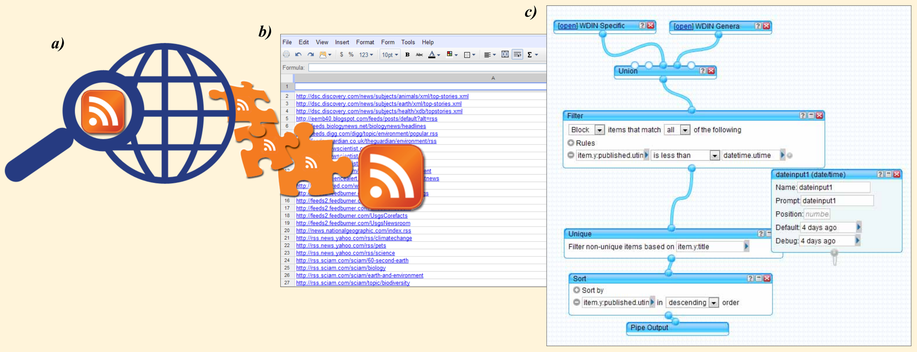


Figure 1: *Graphical example of the flow of information from RSS feeds: (a) which produced 900 possible references organized through Google Docs, (b) and filtered by Yahoo Pipes, (c) resulting in 300 needing review.*

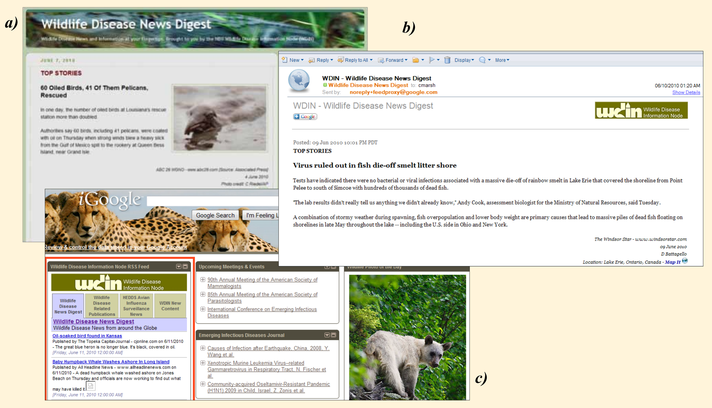
Stories appearing on the resulting list are then reviewed by WDIN staff and chosen for inclusion into the Digest by referring to a published News Selection Policy, wildlifedisease.nbii.gov/documents/Digest_News_Selection_Policy.pdf. The primary purpose of the policy is to ensure the scope of the Digest continues to focus on wildlife diseases as well as wildlife morbidity/mortality events, especially as they relate to human and ecosystem health.

Stories selected are arranged in a headlined newspaper-type format using the Blogger Web application www.blogger.com. The News Digest is usually published 5 days per week and appears at wdin.blogspot.com. This format consists of three to four main stories per day, which include the original first few paragraphs with a link to the original article source, as well as links to other related, non-headlined articles. Readers can also subscribe to receive each edition by e-mail, or through a RSS feed which can be interpreted by a "feed-reader," a stand-alone application, or the feed can also be built into other Web tools www.feedreader.com; reader.google.com. See **Figure 2** for an example of the different formats the Digest content can be viewed in.

Methods (cont.)

In addition to news articles published in the popular press, relevant articles from peer-reviewed journals are also selected and included using the same News Selection Policy. Using a similar procedure for processing news stories, publications are gathered, filtered, and managed. In this case about 70 journal publisher RSS feeds are used. Many publishers still do not offer RSS feeds, so additional manual searching and selecting is required. Some subscription services used for this process are ISI Web of Knowledge and Science.gov.

Figure 2: *The Wildlife Disease News Digest in its Web format (a), and how it appears as an e-mail message (b), and in the iGoogle Web application (c).*



The selected articles are cataloged using the open source citation manager, Connotea, www.connotea.org. Through this application, an article's available metadata populates a Connotea Web record form. Other information, such as keywords, are manually entered, and the application stores and manages the terms, allowing it to suggest keywords for future entries. Once cataloged, the articles are also made available through the Digest in the weekly segment, Wildlife Disease Journal Digest. (**Figure 3**).

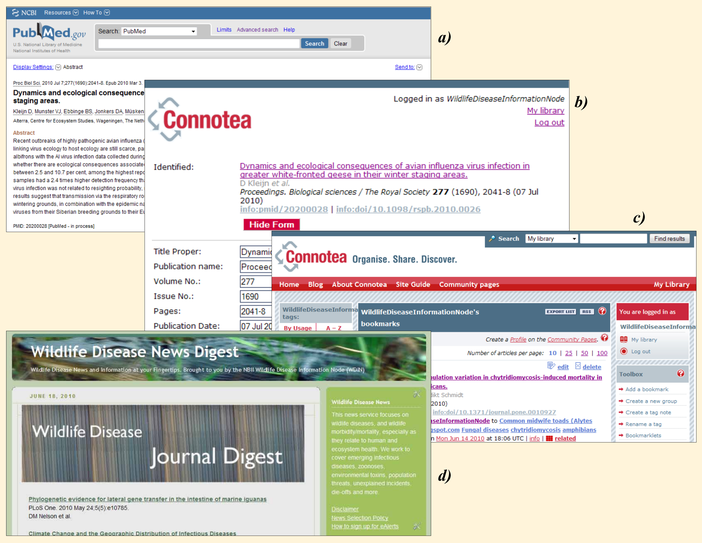


Figure 3: *Illustration of Connotea process. Journal article (a) has metadata captured (b) into a record and added to a designated WDIN journal citation collection (c), which is then featured in the Digest segment, the Wildlife Disease Journal Digest (d).*

Methods (cont.)

Global Wildlife Disease News Map

To create a geospatial depiction of Digest articles, a subset that represent a "wildlife health incident" – disease spread or detection (e.g., "Bird tests positive for West Nile virus in St. Clair County; first of this year") are incorporated into the Global Wildlife Disease News Map. Metadata for the articles are cataloged into a Dublin-Core based Microsoft SQL Server Database. The article's title, description, date published, source, and publisher are input along with information that gives it a geographic location. Geocoded news items include a full hierarchy of geographic information: Place name (if applicable), County (if U.S.), Administrative Unit, Country, and Continent. Place name includes coordinates for a specific place geocoded by the Geonames database, geonames.org, a metasource type gazetteer that combines a number of available place name databases and provides information and Web services through a creative commons license. County, Administrative Unit, Country, and Continent information are mapped at a centroid point.

In addition, a series of keywords or tags are applied to the resource depending on the article's contents. These terms are standardized and allow users to filter the map contents by topic and wildlife/human/livestock relationship, disease, country, species involved, and date.

Information from the database is transferred to create an interactive map wildlifedisease.nbii.gov/wdinNewsDigestMap.jsp using the Google Maps Application Programming Interface. This process takes the geocoded data about an article and represents it as a marker on the map. The basic framework of the map consists of HTML and JavaScript in order to provide the basic map, functionalities (like zooming and panning), and background layers (map, hybrid, and satellite). To load the pushpins into the map, an XML format is used to direct the display of the news articles. This XML file is dynamically (on-demand) generated when it is requested by the Map, and includes the last 45 days of geographically relevant stories.

When a user clicks a displayed pushpin, basic information about those news items is displayed (item title, link, short abstract, geographic detail, source, and publication date). News items are grouped by geographic information, so that multiple stories may show up within a single pushpin if they share the same geographic detail. (**Figure 4**) Links are also provided that direct the user to background disease information available on the WDIN Web portal found at wildlifedisease.nbii.gov. (**Figure 5**).

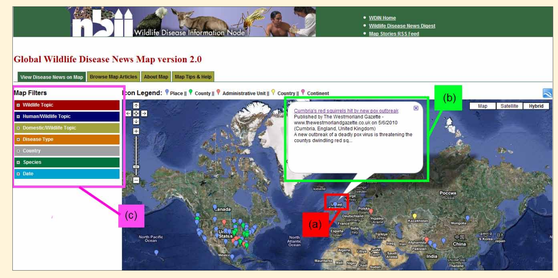


Figure 4: *The Global Wildlife Disease News Map showing the markers (a), information available from the marker (b), and filtering tools (c).*

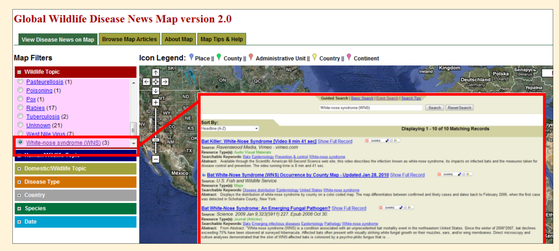


Figure 5: *Illustration of linkage between Global Wildlife Disease News Map (a) and WDIN Web portal (b).*

The data from the News Map is incorporated into a GeoRSS feed (an RSS feed with geospatial information), which can be used by other reporting systems. An example of the result of this process used by HealthMap healthmap.org is shown in **Figure 6**.

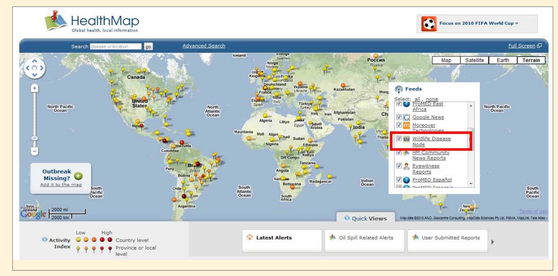


Figure 6: *Illustration of how the Global Wildlife Disease News Map GeoRSS feed is integrated into the HealthMap Global Disease Alert Map.*

Discussion:

The analysis of electronic media products offers an opportunity to expand the options for global disease surveillance techniques and strategies.[4] As described here, the Wildlife Disease News Digest and companion Global Wildlife Disease News Map extend this capability by focusing on disease events relating to wildlife. As a high percentage of emerging and zoonotic diseases have a wildlife component, these tools offer an important addition to Internet content derived surveillance. Although these products have been developed using open-source applications and incorporate automatic filtering mechanisms, manual review and selection of articles is still needed to ensure a consistent and focused resource. By using a similar process to screen and catalog relevant peer-reviewed literature, an effective linkage between current events and previous research can be made. Providing RSS data streams that can be used by other Web based surveillance systems concentrated on human or ecosystem health, we can contribute to the integration of information which is essential to advancing the vision of the One Health concept.

1. Freifeld, CC et al. 2007. HealthMap: Global infectious disease monitoring through automated classification and visualization of Internet media reports. JAMIA 15(2): M2544.
http://www.healthmap.org

2. Collier, N et.al. 2008. BioCaster: detecting public health rumors with a Web-based text mining system. Bioinformatics.24(24):2940-1.
http://biocaster.nii.ac.jp

3. Mykhalovskiy E and L Weir. 2006. The Global Public Health Intelligence Network and early warning outbreak detection: a Canadian contribution to global public health. Can J Public Health. 2006 Jan-Feb;97(1):42-4.
http://www.phac-aspc.gc.ca/gphin/index-eng.php

4. Brownstein JS, CC Freifeld , and LC Madoff. 2009. Influenza A (H1N1) virus, 2009--online monitoring. N Engl J Med. (21):2156.

5. Keller, Mikaela et al. 2009. Use of Unstructured Event-Based Reports for Global Infectious Disease Surveillance. Emerg Infect Dis. 15(4): 689-695

wdin Wildlife Disease Information Node

THE NELSON INSTITUTE FOR ENVIRONMENTAL STUDIES University of Wisconsin-Madison TOGETHER. FOR THE PLANET.

USGS science for a changing world