

MANTLE

**An open-source
platform for
biosurveillance &
One Health
research**



EcoHealth Alliance

EXECUTIVE SUMMARY

Mantle is a free and open-source project, developed by EcoHealth Alliance under the Apache License 2.0, with all code hosted on GitHub which fosters collaborative development. Data that are made publicly available by contributors will be open access (free to any other Mantle user). **Mantle will be free to everyone.**

Infectious diseases cause great harm to human and animal health across the globe. Infectious disease emergence has destabilized economies, harmed biodiversity, and reduced the capacity of public health systems to respond to infectious disease outbreaks. The 2014-2015 Ebola epidemic revealed that the international public health community lacks the sheer capacity to respond to outbreaks when they first occur. Biosurveillance systems need to better detect infectious disease outbreaks and communicate surveillance information to minimize harm to the global economy, animals, and people.

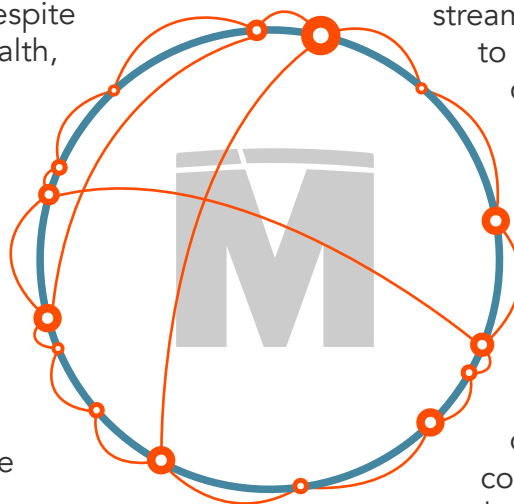
The One Health approach suggests that humans, animals, and the environment are closely tied together. Human interaction with wildlife and the environment contributes to increased risk for human, plant, and animal infectious disease outbreaks. Since human, animal, and ecosystem health are linked, interdisciplinary and holistic approaches are needed to prevent future infectious disease outbreaks. Despite the movement towards One Health, the software currently available to manage, analyze, and communicate the vast amount of One Health data is grossly inadequate. One Health data are continually growing in size and complexity, and new technologies must be developed to address the magnitude of the problem. Furthermore, the desire of single entities to control and leverage information for greater personal and organizational wealth and power directly opposes the goals of biosurveillance, One Health, and science. Open access and open source software are needed to address these complex One Health

problems, and to improve data accessibility, interoperability, and information communication.

Mantle is an open-source web platform designed for the storage, sharing, and visualization of One Health biosurveillance data and is designed to meet the needs of a wide variety of users. One Health scientists in the field or the lab will be able to upload datasets in multiple formats to Mantle's intelligent database, where they will be stored for easy download and analysis. Mantle users will be able to use fine-grained access controls to protect and share their uploaded datasets, and examine datasets in views appropriate to their content (e.g., tables, maps, and charts). Mantle's flexible storage layer will also display spatial datasets from different data sources alongside one another, and save and export combined datasets.

Mantle users will be able to belong to organizations and groups, and individual datasets will be grouped together into larger projects, all with group-level access permissions. These features enable scientists to collaborate across geographic, institutional, and disciplinary boundaries to accomplish large-scale data collection efforts that would not be otherwise possible. Mantle also includes a number of open-access datasets from partners, governmental organizations, and biosurveillance data streams that are available for users to combine with their own data or content.

Mantle will provide free access to high fidelity infectious disease data, which will enable scientists, health practitioners, and policymakers to tackle the world's biggest One Health challenges. Mantle will enable proactive response to infectious disease outbreaks as data are continuously uploaded, analyzed, and contextualized. Open access health data will help One Health research advance and Mantle will fill a critical gap in emerging infectious disease research, knowledge, and prevention.



Inadequacy of Current Biosurveillance Data Management Software

Our success in averting future public health crises, like those posed by emerging infectious diseases (EIDs) like highly pathogenic Avian Influenza, MERS, and Ebola, will depend largely on the One Health community's ability to accurately and rapidly collect, process, and analyze large amounts of information on potential infectious disease outbreaks. Existing software for managing the collection, analysis, and communication of One Health Data are inadequate for this discipline-spanning, inherently collaborative endeavor.

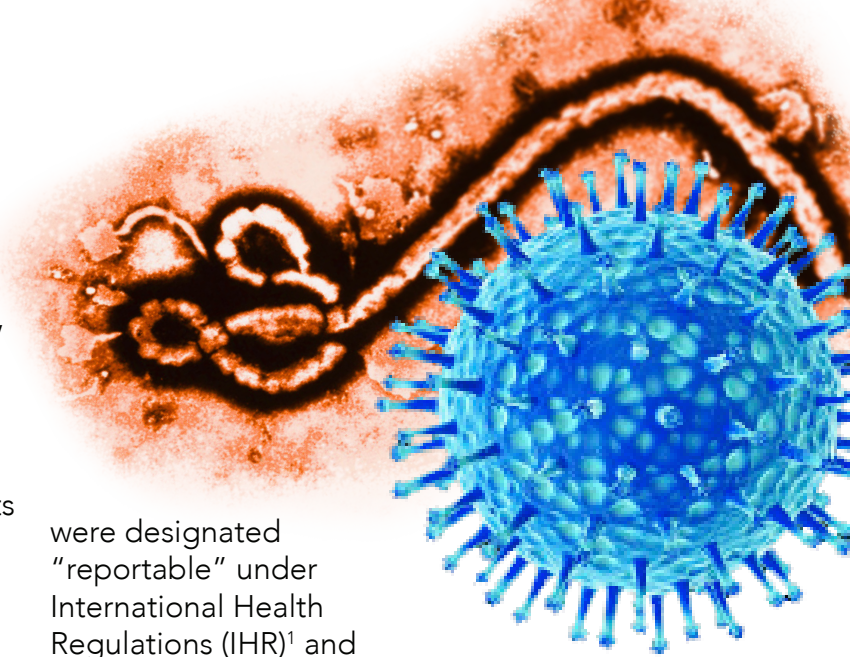
Thousands of independent datasets exist for a wide range of infectious disease phenomena, and can be used to inform traditional biosurveillance systems or develop novel methods^{1,2}. In practice, these datasets are generally large, complex, and located in disparate locations and formats, and these characteristics limits the effective scope of biosurveillance systems. Novel methods for combining, analyzing, and visualizing big and complex data can overcome these difficulties, and hold great promise for finding new solutions to complex global problems. For instance, heuristics can be applied to inconsistent data for infectious disease datasets in models that estimate the spatial distribution of risk given the available data². Even with relatively poor quality data, information from these models can improve the allocation of international development resources to improve environmental health, determine where biosurveillance gaps exist, and identify where public health and medical resources should be staged to more rapidly extinguish infectious disease outbreaks.

However, even with modeling methods flexible enough to work with a wide variety of datasets, the relevant datasets are generally kept private, inaccessible to scientists and policymakers, and in disparate formats. Consequently, a single implementation of the data combination and analysis methods is impossible or overly burdensome.

The field of biosurveillance exemplifies these problems. Traditional disease surveillance operated under a model where certain diseases

were designated "reportable" under International Health Regulations (IHR)¹ and other non-reportable diseases were largely ignored. Until 2005, the IHR included cholera, plague, yellow fever, and smallpox (before it was eradicated). These biosurveillance methods were not designed to detect EIDs or answer questions related to One Health. Event-based surveillance systems have begun to monitor other data feeds to respond to potential disease threat indicators, whether of known or unknown cause. Syndromic surveillance monitors streams of data like employee absenteeism and sales of prescription and over-the-counter medication to identify signals suggesting a public health event¹. These signals have not yet predicted an outbreak and are currently most effective as an early warning system to supplement the other types of existing biosurveillance systems¹.

Despite the hype, Twitter data, social media, and web search term data have not outperformed traditional disease surveillance systems in accuracy or speed³. While these data sources may temporarily fill gaps where public health systems are inadequate, nothing can replace hard scientific evidence collected in the field. USAID and EcoHealth Alliance's partner organizations have access to vast datasets of infectious disease and development related data, including passive and active surveillance on IHR diseases, novel pathogens, and animal and human emerging infectious diseases. These data are almost entirely spatial in nature: they describe (implicitly, if not explicitly) phenomena occurring across the Earth's surface. When these One Health data are combined, they can help identify biosurveillance information gaps, assist with resource allocation



during disease outbreaks, and accurately identify drivers of emerging infectious diseases. A system to host disparate datasets and make them interoperable would enable many new applications in One Health research and could dramatically improve public health via the One Health Analytical Approach.

No Existing Product is a Complete Solution

Existing tools only partly address the need to collate and combine infectious disease data. USAID's GeoCenter is a storehouse of spatial datasets hosted by ESRI. It hosts and can export spatial datasets, and takes advantage of their spatial nature by displaying each on an interactive map with zoom and pan data examination. However, it lacks open access and interoperability, is incompatible with tabular data, and requires an ArcGIS subscription to sign in and download datasets, which is prohibitively expensive for many users.

Another example, Dryad, is an open-source archive of datasets. Dryad hosts detailed metadata for each dataset, provides DOI numbers for datasets it hosts so that it can host datasets associated with published papers, and provides an interface to search both Dryad-hosted datasets and datasets indexed from KNC and TreeBASE, two other dataset archives. All three systems are part of DataONE, and the NSF-funded collaboration is working toward better data practices in science.

Dryad differs from USAID's GeoCenter in a few ways. It is not designed for a specific data type; it is intended for users to register and upload their own datasets. Both treat datasets as single, autonomous units, with dataset-level metadata but no sub-dataset-level metadata. In contrast, EcoHealth Alliance's Mantle aims to operationalize users' datasets and will uniquely allow the combining and filtering of existing datasets, generating new possibilities for existing surveillance data feeds not provided by Dryad or GeoCenter.

Currently, most One Health analyses are carried out on local machines, and on static datasets, like those hosted on GeoCenter and Dryad. Analyses might be carried out in GIS software, whether closed-source like ESRI's ArcGIS or open-source

like QGIS and GRASS. Programming languages like R and Python also allow in-depth analyses to be conducted on spatial data, using custom statistical procedures as warranted. These tools form a crucial part of the tool chain, and the ability to examine and explore datasets with them will be important for the foreseeable future. Mantle will allow users to download data for analysis in these software applications while providing value-added products and services for combined datasets.

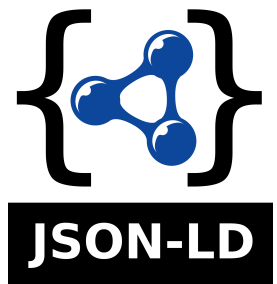
Mantle is the Solution

Mantle will be an open-source, cloud-compatible platform for storing, studying, and sharing data on infectious diseases across plants, animals, and humans. It will meet the needs of three groups of users: scientists, policymakers, and the general public. Mantle will be a superior technology that will address the aforementioned gaps in One Health software.

For scientists, Mantle will make datasets portable and connected. Scientists will be able to upload datasets to the Mantle website or collect data from the field using a mobile app. Users in Mantle will be able to easily make datasets entirely private, publicly accessible, or shared with specific users or groups.

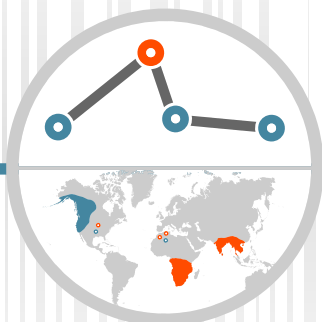
Mantle will handle tabular data, and other widely used spatial data formats. It will visualize and explore data in useful ways, and allow data to be downloaded as the originally uploaded file or in a customizable format for use in analytical software.

Mantle will store metadata—information about a dataset and its contents—using development standards for linked data (e.g., JSON-LD and WCSV, part of the overarching Resource Description Framework). Tapping into the emerging semantic web enables richer interactions with datasets, streamlining many common data tasks. Mantle will natively understand a number of data types common to One Health data, including spatial and temporal elements, taxonomic names, and case counts, and associates these with published ontologies. Mantle will also work seamlessly with any numeric, categorical, and textual data.

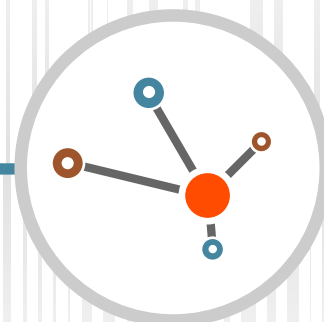




STORE



STUDY



SHARE

Mantle will be pre-populated with a variety of open-access datasets for common One Health data types, including land cover and land use change, human population, weather and climate, species ranges, and land use types. User datasets will be visualized alongside these datasets, and they will be exported alongside user datasets for use in analyses. In addition, any user-uploaded datasets marked as open-access can be visualized and exported in the same manner.

Policymakers and decision makers will be able to view real-time visualizations of Mantle data feeds in dashboards. Researchers will be able to upload datasets representing the output of models built in other analytical software, which can be shared with policymakers, who can also view and interact with the output of custom-built modeling modules to view timely and meaningful summaries of public health data feeds. Potential use cases for the general public include browsing day-to-day textual and syndromic surveillance information, viewing the predictions of a one-time study, and monitoring the latest calculated epidemic curve in an outbreak or ongoing epidemic.

Mantle will facilitate crosscutting collaborations between disciplines and institutions. Users will be able to create, manage, and join organizations

and groups. Groups of users will access and collaborate on collections of datasets, grouped manually or by specified properties. For instance, users interested in Ranavirus can view and contribute to the Global Ranavirus Reporting System, a collaborative effort by scientists worldwide to aggregate observed cases of Ranavirus across species and locations (a Mantle prototype).

Mantle is being developed as an open source project on GitHub. Users will be able to run their own Mantle servers, which will be able to interface with other Mantle instances using semantic web standards. Mantle's flexibility makes it suited to a spectrum of tasks related to infectious disease data, while it is simultaneously tailored to commonalities and frequent problems in One Health.

¹Morse, S. S. (2012). Public health surveillance and infectious disease detection. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, 10(1), 6–16. <http://doi.org/10.1089/bsp.2011.0088>

²Hay, S. I., George, D. B., Moyes, C. L., & Brownstein, J. S. (2013). Big data opportunities for global infectious disease surveillance. *PLoS Medicine*, 10(4), e1001413. <http://doi.org/10.1371/journal.pmed.1001413>

³Olson, D. R., Konty, K. J., Paladini, M., Viboud, C., & Simonsen, L. (2013). Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS Comput Biol*, 9(10), e1003256.



For more information, contact:

Andrew G. Huff, Ph.D., M.S.

1.212.380.4497 or huff@ecohealthalliance.org



EcoHealth Alliance