



# Mantle: Fusion of Global Biosurveillance Big Data, CBA 6, PI Andrew Huff, EcoHealth Alliance

**Objective:** To develop a multilingual, open source, and open access software application that combines disparate biosurveillance and infectious disease data from multiple foreign governments and international academic sources.

**Description of Effort:**

- Create centralized, secure, Internet based system, to collect, combine, and share disparate biosurveillance data sources, in multiple languages, and formats
- Obfuscating human health and agricultural data, from research or surveillance streams or field based research, to maintain regulatory compliance and privacy when necessary
- Provide mobile apps for One Health data collection in the field in multiple languages and data formats for integration directly to BSVE
- Create an application to automatically clean, sort, and combine biosurveillance data, in multiple languages, with minimal user input and effort

**Benefits of proposed technology:** Mantle will provide the technology to collect and combine biosurveillance data from multiple governments, in multiple languages, and multiple data formats that will result in a measurable impact on the bio-event timeline. No such technology exists.

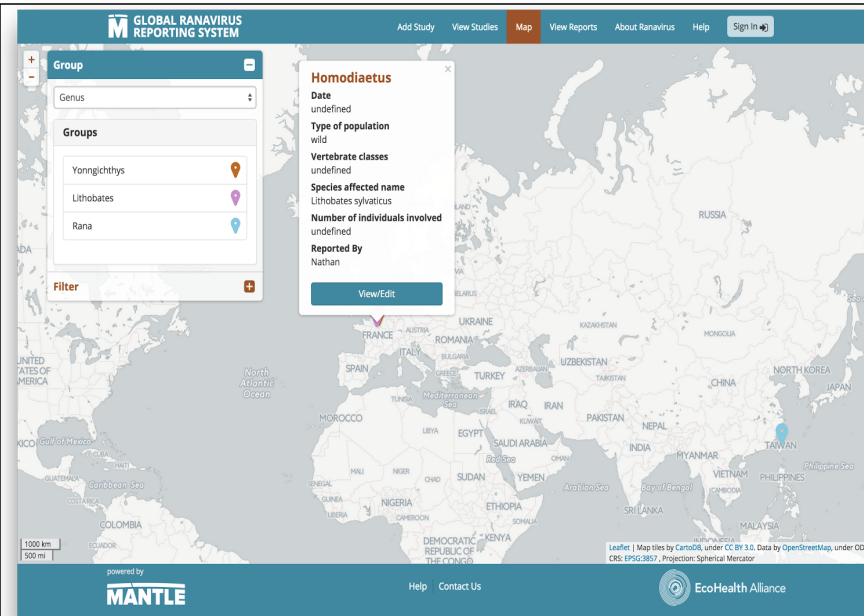
**Challenges:**

- Marketing technology internationally
- Assuring users of their ability to protect their data
- Malicious or inaccurate data injected into system

**Maturity of technology:** TRL 3 & MRL 3

**Technology goal:** TRL 8 & MRL 5

**Topic:** CBA-06



**Figure 1.** Users will be able to select biosurveillance data sets by a combination of location, area, or disease type.

**Major goals:**

(Base Year) Establish universal biosurveillance data standards. Develop an open-source and open access (free to all users) cloud-based global One Health biosurveillance data fusion app in multiple languages compatible with multiple data formats  
(Year 1) Develop a friendly and efficient user interface for international users to provide and interact with global biosurveillance data and provide data via API to BSVE developers and analysts

**Cost:** \$2,296,336.59

**Period of Performance:** 2 years

**Contact Information:** Andrew Huff ph# 1.612.743.1265  
huff@ecohealthalliance.org

## PROJECT SUMMARY

The goal of Mantle is to create an Internet app that enables non-technical scientists and governments to easily and efficiently apply metadata to biosurveillance data. Mantle is a free and open-source research and software development project, developed under the Apache License 2.0, dockerized, using the BSVE's SDK. Mantle will be an open-source web platform designed for the curation, integration, and sharing of global public health related data. By incorporating previously developed metadata standards and public health ontologies, and automating their application, Mantle will be designed to meet the data needs of a wide variety of domestic and international public health users. Public health professionals, in the office, field, or the lab, will be able to upload a wide variety of unstructured and structured datasets to Mantle in a variety of commonly used formats and languages. Additionally, Mantle will have an API that can be used to ingest and combine existing and continuously collected biosurveillance data in near real time.

Mantle's users can belong to governments, private organizations, and individuals, and individual datasets can be grouped together into larger projects, all with group-level access permissions. These features enable public health professionals to collaborate across geographic, institutional, and disciplinary boundaries to accomplish large-scale data collection efforts not otherwise possible. Mantle will also include a number of open-access datasets from EcoHealth Alliance's partners, and biosurveillance data, that with Mantle, will be openly available for users to combine with their own data or content.

Users of Mantle will be able to set fine-grained sharing and privacy controls on uploaded datasets to share or protect their data, and industry best practices will be employed to protect all data uploaded to Mantle. Once users create and sign into their user account, Mantle users will be able to examine publicly available and obfuscated datasets (to protect privacy) in a number of views appropriate to their content, including tables, maps, and charts. Additionally, Mantle will display datasets from different data sources alongside one another and save and export combined datasets. Users with export privileges will be able to download data in a number of formats for use with external software (e.g., .xlsx, .csv, .txt, .shp, .shx, .dbf, etc.).

Mantle will uniquely provide free access to high fidelity infectious disease data, which will help enable scientists, public health practitioners, and policymakers to tackle the world's biggest infectious disease threats. Furthermore, Mantle will enable identification and faster response to infectious disease threats as data can be continuously uploaded, validated, and contextualized via Mantle's API, rather than waiting for data to be collected and integrated after infectious disease threats are identified via traditional biosurveillance mechanisms (by significantly reducing data communication, data cleaning, and language translation time). Open access health data and open source biosurveillance software will help infectious disease and biosurveillance research advance, and Mantle will fill a critical gap in emerging infectious disease knowledge and infectious disease preparedness. Mantle will generalize across scientific fields as more big data ontologies are created, and Mantle will be able to be used broadly.

## **Broader Impact**

Mantle is currently aimed at biosurveillance and health data. Many components will have potential use beyond biosurveillance, so Mantle will be developed in a generalizable, reusable, and scalable manner. As with any data integration platform, data security must be addressed. As data become more portable, accessible and integrated, systems must be hardened against malicious attacks. Therefore sensitive data must be safeguarded, including personally identifying information, and security will be incorporated into the design of Mantle from the outset.

Databases that incorporate public data must also protect against the injection of false data. With proper security measures in place, Mantle will be useful for broad scale ecological and land use data, health data, and human behavior and demographics data. Furthermore, Mantle could serve as a novel way to integrate ecological and social data to improve understanding of how human and natural systems interact to change health outcomes and affect disease emergence.

Mantle is an innovative fusion of software engineering, data science, and public health research. For about the past decade, data portability and availability has been pursued by scientists and mandated by governments, but has not fundamentally improved. Previous studies have found that structural semantic heterogeneity are significant obstacles to overcome when combining data and when assigning metadata and these problems are exacerbated by the lack of formal training in data integration of most scientists in biomedical fields and across academia. Mantle directly addresses these technical problems and human deficiencies by automating the metadata application processes where possible and guiding users elsewhere, using machine-learning algorithms trained on existing data and crowd sourced dataset annotations. By automating these processes, we hope that Mantle will be used outside of public health and biomedical research, as it directly addresses a problem that is common to throughout scientific disciplines. A couple of examples of non-public health use-cases is 1) the use of databases as virtual laboratories in astronomy, where an astronomer can make and record a large number of virtual observations; and 2) the post hoc combination of data from differing wearable medical devices for greater analytical capability battlefield strength.

We hope that Mantle will succeed in overcoming current metadata practices by integrating Mantle with an API for data upload and download. This means that other BSVE developers can extend the system, perhaps directly uploading datasets from mobile devices or importing directly into an analysis application. We will develop secure mechanisms to obfuscate sensitive data. This will make Mantle compliant with regulations for sharing health data, broadening its set of use-cases. Furthermore, Mantle's metadata assignment features will exist in a user-focused, community-based platform, enabling scientists with domain knowledge and no expertise to contribute and to describe their data in flexible ways that make it interoperable with other similar datasets.