

## APPLICATION FOR FEDERAL ASSISTANCE

## SF 424 (R&amp;R)

3. DATE RECEIVED BY STATE		State Application Identifier
1. TYPE OF SUBMISSION*		4.a. Federal Identifier
<input type="radio"/> Pre-application <input type="radio"/> Application <input checked="" type="radio"/> Changed/Corrected Application		b. Agency Routing Number
2. DATE SUBMITTED 2015-10-06	Application Identifier	c. Previous Grants.gov Tracking Number GRANT12011542
5. APPLICANT INFORMATION <span style="float: right;">Organizational DUNS*: 0770900660000</span>		
Legal Name*: EcoHealth Alliance, Inc. Department: Division: Street1*: 460 West 34th Street Street2*: 17th Floor City*: New York County: State*: NY: New York Province: Country*: USA: UNITED STATES ZIP / Postal Code*: 10001-2317		
Person to be contacted on matters involving this application Prefix: Dr.      First Name*: Andrew      Middle Name:      Last Name*: Huff      Suffix: Position/Title: Associate Vice President Street1*: 460 West 34th Street Street2*: 17th Floor City*: New York County: State*: NY: New York Province: Country*: USA: UNITED STATES ZIP / Postal Code*: 10001-2317 Phone Number*: 1.212.380.4497      Fax Number: +1.212.380.4465      Email: huff@ecohealthalliance.org		
6. EMPLOYER IDENTIFICATION NUMBER (EIN) or (TIN)*		311726494
7. TYPE OF APPLICANT*		M: Nonprofit with 501C3 IRS Status (Other than Institution of Higher Education)
Other (Specify): <b>Small Business Organization Type</b> <input type="radio"/> Women Owned <input type="radio"/> Socially and Economically Disadvantaged		
8. TYPE OF APPLICATION*		If Revision, mark appropriate box(es).
<input checked="" type="radio"/> New <input type="radio"/> Resubmission <input type="radio"/> Renewal <input type="radio"/> Continuation <input type="radio"/> Revision		<input type="radio"/> A. Increase Award <input type="radio"/> B. Decrease Award <input type="radio"/> C. Increase Duration <input type="radio"/> D. Decrease Duration <input type="radio"/> E. Other (specify):
Is this application being submitted to other agencies?* <input type="radio"/> Yes <input checked="" type="radio"/> No      What other Agencies?		
9. NAME OF FEDERAL AGENCY* National Institutes of Health		10. CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER TITLE:
11. DESCRIPTIVE TITLE OF APPLICANT'S PROJECT* Mantle: A Unifying Software for Disparate Big Data		
12. PROPOSED PROJECT Start Date*      Ending Date* 11/01/2015      10/31/2018		13. CONGRESSIONAL DISTRICTS OF APPLICANT NY-010

**14. PROJECT DIRECTOR/PRINCIPAL INVESTIGATOR CONTACT INFORMATION**

Prefix: Dr. First Name\*: Andrew Middle Name: Last Name\*: Huff Suffix:

Position/Title: Associate Vice President

Organization Name\*: EcoHealth Alliance, Inc.

Department:

Division:

Street1\*: 460 West 34th Street

Street2: 17th Floor

City\*: New York

County:

State\*: NY: New York

Province:

Country\*: USA: UNITED STATES

ZIP / Postal Code\*: 10001-2317

Phone Number\*: 1.212.380.4497 Fax Number: Email\*: huff@ecohealthalliance.org

**15. ESTIMATED PROJECT FUNDING**

a. Total Federal Funds Requested\* \$1,220,359.25

b. Total Non-Federal Funds\* \$0.00

c. Total Federal & Non-Federal Funds\* \$1,220,359.25

d. Estimated Program Income\* \$0.00

**16. IS APPLICATION SUBJECT TO REVIEW BY STATE EXECUTIVE ORDER 12372 PROCESS?\***

a. YES ☐ THIS PREAPPLICATION/APPLICATION WAS MADE AVAILABLE TO THE STATE EXECUTIVE ORDER 12372 PROCESS FOR REVIEW ON:

DATE:

b. NO ☐ PROGRAM IS NOT COVERED BY E.O. 12372; OR

☒ PROGRAM HAS NOT BEEN SELECTED BY STATE FOR REVIEW

**17. By signing this application, I certify (1) to the statements contained in the list of certifications\* and (2) that the statements herein are true, complete and accurate to the best of my knowledge. I also provide the required assurances \* and agree to comply with any resulting terms if I accept an award. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties. (U.S. Code, Title 18, Section 1001)**

☒ I agree\*

\* The list of certifications and assurances, or an Internet site where you may obtain this list, is contained in the announcement or agency specific instructions.

**18. SFLL or OTHER EXPLANATORY DOCUMENTATION**

File Name:

**19. AUTHORIZED REPRESENTATIVE**

Prefix: Mr. First Name\*: Aleksei Middle Name: Last Name\*: Chmura Suffix:

Position/Title\*: Authorized Organizational Representative

Organization Name\*: EcoHealth Alliance, Inc.

Department:

Division:

Street1\*: 460 West 34th Street

Street2: 17th Floor

City\*: New York

County:

State\*: NY: New York

Province:

Country\*: USA: UNITED STATES

ZIP / Postal Code\*: 10001-2317

Phone Number\*: +1.212.380.4473 Fax Number: +1.212.380.4465 Email\*: chmura@ecohealthalliance.org

**Signature of Authorized Representative\***

Aleksei Chmura

**Date Signed\***

10/06/2015

**20. PRE-APPLICATION** File Name:**21. COVER LETTER ATTACHMENT** File Name: 1244-Cover Letter.pdf

## 424 R&R and PHS-398 Specific Table Of Contents

### Page Numbers

SF 424 R&R Cover Page	1
Table of Contents	3
Performance Sites	4
Research & Related Other Project Information	5
Project Summary/Abstract(Description)	6
Project Narrative	7
Facilities & Other Resources	8
Research & Related Senior/Key Person	10
Research & Related Budget Year - 1	30
Research & Related Budget Year - 2	33
Research & Related Budget Year - 3	36
Budget Justification	39
Research & Related Cumulative Budget	43
PHS398 Cover Page Supplement	44
PHS 398 Research Plan	46
Specific Aims	47
Research Strategy	48
Human Subjects Section	59
Protection of Human Subjects	59
Bibliography & References Cited	60
Consortium/Contractual	63
Letters Of Support	64
Resource Sharing Plans	67

**Project/Performance Site Location(s)****Project/Performance Site Primary Location**

☐ I am submitting an application as an individual, and not on behalf of a company, state, local or tribal government, academia, or other type of organization.

Organization Name: EcoHealth Alliance, Inc.  
Duns Number: 0770900660000  
Street1\*: 460 West 34th Street  
Street2: 17th Floor  
City\*: New York  
County:  
State\*: NY: New York  
Province:  
Country\*: USA: UNITED STATES  
Zip / Postal Code\*: 10001-2317  
Project/Performance Site Congressional District\*: NY-010

---

File Name

**Additional Location(s)**



## RESEARCH &amp; RELATED Other Project Information

<b>1. Are Human Subjects Involved?*</b> <input type="radio"/> Yes <input checked="" type="radio"/> No	
1.a. If YES to Human Subjects Is the Project Exempt from Federal regulations? <input type="radio"/> Yes <input type="radio"/> No If YES, check appropriate exemption number:   — 1 — 2 — 3 — 4 — 5 — 6 If NO, is the IRB review Pending? <input type="radio"/> Yes <input type="radio"/> No IRB Approval Date: Human Subject Assurance Number	
<b>2. Are Vertebrate Animals Used?*</b> <input type="radio"/> Yes <input checked="" type="radio"/> No	
2.a. If YES to Vertebrate Animals Is the IACUC review Pending? <input type="radio"/> Yes <input type="radio"/> No IACUC Approval Date: Animal Welfare Assurance Number	
<b>3. Is proprietary/privileged information included in the application?*</b> <input checked="" type="radio"/> Yes <input type="radio"/> No	
<b>4.a. Does this project have an actual or potential impact - positive or negative - on the environment?*</b> <input checked="" type="radio"/> Yes <input type="radio"/> No	
4.b. If yes, please explain:   Promotes global conservation through biosurveillance.	
4.c. If this project has an actual or potential impact on the environment, has an exemption been authorized or an environmental assessment (EA) or environmental impact statement (EIS) been performed? <input type="radio"/> Yes <input type="radio"/> No	
4.d. If yes, please explain:	
<b>5. Is the research performance site designated, or eligible to be designated, as a historic place?*</b> <input type="radio"/> Yes <input checked="" type="radio"/> No	
5.a. If yes, please explain:	
<b>6. Does this project involve activities outside the United States or partnership with international collaborators?*</b> <input type="radio"/> Yes <input checked="" type="radio"/> No	
6.a. If yes, identify countries:	
6.b. Optional Explanation:	
<b>7. Project Summary/Abstract*</b>	Filename 1240-ProjectSummary.pdf
<b>8. Project Narrative*</b>	1241-Project Narrative.pdf
<b>9. Bibliography &amp; References Cited</b>	1242-References Cited.pdf
<b>10. Facilities &amp; Other Resources</b>	1243-Facilities and Other Resources.pdf
<b>11. Equipment</b>	

## PROJECT SUMMARY

The goal of Mantle is to create a software platform that enables non-data scientists to easily and efficiently apply metadata to their own datasets. Mantle is a free and open-source research and software development project, developed under the Apache License 2.0, with all code hosted on GitHub. Mantle will be an open-source web platform designed for the curation, integration, and sharing of 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 public health users. Public health researchers, 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.

Mantle's users can belong to organizations and teams, and individual datasets can 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 not otherwise possible. Mantle will also include a number of open-access datasets from EcoHealth Alliance's partners, and biosurveillance data, that are 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, practitioners, and policymakers to tackle the world's biggest infectious disease threats. Furthermore, Mantle will enable 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. 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 will be able to be used broadly.

## **PROJECT NARRATIVE**

Mantle is a software platform that enables non-data scientists to easily and efficiently combine disparate big data to increase the speed and efficiency of public health research. Mantle does this by reducing the complexity of cleaning and merging big data through the automation of existing methods to combine big data. Mantle will enable public health scientists to answer highly complex public health questions more rapidly.

## **FACILITIES AND OTHER RESOURCES**

**EcoHealth Alliance, New York, USA** (Andrew Huff, Peter Daszak, Brock Arnold, Toph Allen) EcoHealth Alliance is a 40-year old scientific research NGO that specializes in multidisciplinary research and surveillance of the spread of zoonotic emerging diseases. EcoHealth Alliance is based in New York City with 10,000 square feet of office space including a meeting room. The scientific and technology staff (15 scientists and 9 engineers] and associated interns are supported by a core administrative staff of 11 people that are available for work on this project and is funded through core funds.

EcoHealth Alliance is equipped with fiber optic Internet access and video conferencing facilities to facilitate easy communication between collaborators. EcoHealth Alliance employees have around the clock access to servers, VPNs, encryption software, IT support, and all necessary software including Git and Github (Hosted software revision/audit service), Sublime and Vim text editors, Vagrant and Oracle Virtualbox virtual machines, Google Apps (Hosted email and collaboration web based software), Ansible (Server provisioning software framework), Python, NodeJS, and R programming languages, Meteor (Javascript framework), Bash shell scripts, Jenkins (Continuous Integration server), Microsoft Office and Adobe CS6 running on both Apple Mac OS X, Ubuntu linux, and Windows Operating Systems. Additionally we have a dedicated quad-core Linux server, and another dedicated dual quad-core Mac Pro Server - each with 4TB hard drives. Either server individually or in combination may be used for intensive computational modeling and/or database processing by all the grantees. Access to the cloud and supercomputing services (Amazon) is provided by core funding to EHA.

### **The Clango Group**

Clango has an office in Minneapolis, MN. Clango has their own virtual private networks and phone systems, and share financial and administrative personnel. They also have on-site office managers, lunchrooms, private meeting rooms, and advanced conference facilities including large screen projection systems and whole-room Polycom video conferencing systems. The proposed work will be performed at the Arlington and Minneapolis sites. Clango has a mixed environment of personal and shared computing platforms. Employees average two computers per person (desktop, laptop, and/or personal home owned system), with each computer typically equipped with multiple multi-core processors, a high-performance graphics card, dual monitors, and 8GB or more of main memory. These personal systems run a mix of Windows, Mac OS X, and Linux operating systems. Shared resources include compilation and testing farms and workstations running a variety of alternative operating systems for testing purposes.

### **International Society for Disease Surveillance (ISDS)**

ISDS has an office in Brighton, MA. ISDS has their own virtual private networks and phone systems, and share financial and administrative personnel. They also have on-site office managers, lunchrooms, private meeting rooms, and advanced conference facilities including large screen projection systems and whole-room Polycom video conferencing systems. The proposed work will be performed at the Arlington and Minneapolis sites. ISDS has a mixed environment of personal and shared computing platforms. Employees average two computers

per person (desktop, laptop, and/or personal home owned system), with each computer typically equipped with multiple multi-core processors, a high-performance graphics card, dual monitors, and 8GB or more of main memory. These personal systems run a mix of Windows, Mac OS X, and Linux operating systems. Shared resources include compilation and testing farms and workstations running a variety of alternative operating systems for testing purposes.

## RESEARCH &amp; RELATED Senior/Key Person Profile (Expanded)

PROFILE - Project Director/Principal Investigator				
Prefix: Dr.	First Name*: Andrew	Middle Name	Last Name*: Huff	Suffix:
Position/Title*:	Associate Vice President			
Organization Name*:	EcoHealth Alliance, Inc.			
Department:				
Division:				
Street1*:	460 West 34th Street			
Street2:	17th Floor			
City*:	New York			
County:				
State*:	NY: New York			
Province:				
Country*:	USA: UNITED STATES			
Zip / Postal Code*:	10001-2317			
Phone Number*:	1.212.380.4497	Fax Number:	E-Mail*: huff@ecohealthalliance.org	
Credential, e.g., agency login: ANDREWHUFF				
Project Role*: PD/PI		Other Project Role Category:		
Degree Type:		Degree Year:		
Attach Biographical Sketch*:		File Name		
Attach Current & Pending Support:		1235-AndrewHuff.pdf		

PROFILE - Senior/Key Person				
Prefix: Dr.	First Name*: Peter	Middle Name	Last Name*: Daszak	Suffix:
Position/Title*:	President			
Organization Name*:	EcoHealth Alliance			
Department:				
Division:				
Street1*:	460 West 34th Street			
Street2:				
City*:	New York			
County:				
State*:	NY: New York			
Province:				
Country*:	USA: UNITED STATES			
Zip / Postal Code*:	10001-2320			
Phone Number*:	1.212.380.4473	Fax Number:	E-Mail*: daszak@ecohealthalliance.org	
Credential, e.g., agency login: daszak				
Project Role*: Post Doctoral Scholar		Other Project Role Category:		
Degree Type:		Degree Year:		
Attach Biographical Sketch*:		File Name		
		1236-PeterDaszak.pdf		
Attach Current & Pending Support:				

PROFILE - Senior/Key Person				
Prefix: Mr.	First Name*: Brock	Middle Name	Last Name*: Arnold	Suffix:
Position/Title*:	Senior Software Developer			
Organization Name*:	EcoHealth Alliance			
Department:				
Division:				
Street1*:	460 West 34th Street			
Street2:				
City*:	New York			
County:				
State*:	NY: New York			
Province:				
Country*:	USA: UNITED STATES			
Zip / Postal Code*:	10001-2320			
Phone Number*:	1.518.521.8579	Fax Number:	E-Mail*: arnold@ecohealthalliance.org	
Credential, e.g., agency login:				
Project Role*: Other Professional		Other Project Role Category: Senior Software Developer		
Degree Type:		Degree Year:		
Attach Biographical Sketch*:		File Name		
		1237-BrockArnold.pdf		
Attach Current & Pending Support:				

PROFILE - Senior/Key Person				
Prefix: Dr.	First Name*: Noam	Middle Name	Last Name*: Ross	Suffix:
Position/Title*:				
Organization Name*: EcoHealth Alliance				
Department:				
Division:				
Street1*: 460 West 34th Street				
Street2:				
City*: New York				
County:				
State*: NY: New York				
Province:				
Country*: USA: UNITED STATES				
Zip / Postal Code*: 10001-2320				
Phone Number*: 212.380.4460		Fax Number:		
E-Mail*: ross@ecohealthalliance.org				
Credential, e.g., agency login:				
Project Role*: Post Doctoral Scholar			Other Project Role Category:	
Degree Type:			Degree Year:	
Attach Biographical Sketch*:			File Name	
			1238-NoamRoss.pdf	
Attach Current & Pending Support:				

PROFILE - Senior/Key Person				
Prefix:	First Name*: Laura	Middle Name	Last Name*: Streichert	Suffix:
Position/Title*:				
Executive Director				
Organization Name*: International Society for Disease Surveillance				
Department:				
Division:				
Street1*: 26 Lincoln St				
Street2:				
City*: Brighton				
County:				
State*: MA: Massachusetts				
Province:				
Country*: USA: UNITED STATES				
Zip / Postal Code*: 02135-1458				
Phone Number*: 6177797227		Fax Number:		
E-Mail*: lstreichert@syndromic.org				
Credential, e.g., agency login: lcstreichert				
Project Role*: Post Doctoral Scholar			Other Project Role Category:	
Degree Type:			Degree Year:	
Attach Biographical Sketch*:			File Name	
			1239-LauraStreichert.pdf	
Attach Current & Pending Support:				



**BIOGRAPHICAL SKETCH**

Provide the following information for the Senior/key personnel and other significant contributors.  
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: **Andrew G. Huff**

eRA COMMONS USER NAME (credential, e.g., agency login): **ANDREWHUFF**

POSITION TITLE: **Associate Vice President of Data & Technology / Scientist**

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
University of Minnesota	B.A.	05/2010	Psychology, Clinical Research
University of Minnesota	M.S.	08/2011	Engineering, Security Technologies & GIS
University of Minnesota	Ph.D.	01/2014	Public Health, Environmental Health

**A. Personal Statement**

As the Vice President of Data & Technology at EcoHealth Alliance, I am working to develop novel methods of biosurveillance, data analytics and visualization for disease detection, and unique methods to identify disease emergence. My research centers on the interaction of human disease reporting, machine learning of passive and active surveillance data streams, and the interaction of both engineered and natural systems. Additionally, I have been an active participant in government and private sector committees that work to establish effective policies for food systems and multiple aspects of public health. My diverse experiences in big data, public health, and software development make me uniquely positioned to supervise and lead the U01 BD2K grant.

While working at the University of Minnesota, I researched the human, environmental, and engineered aspects of global food systems and patented novel technologies to collect and fuse data from multiple disparate sources to determine which food systems are at risk, map global supply chains in near real-time, and to rapidly identify contaminated food products and supply sources (software development, big data integration, public health):

Huff, A. G., Kennedy, S. P., Kircher, A. L., & Hoffman, J. T. (2014). *U.S. Patent Application 14/212,749*. Available at: <https://www.google.com/patents/CA2846818A1>

At Sandia National Laboratories as a Senior Member of the Technical Staff, I lead interdisciplinary teams in developing novel methods of biosurveillance (big data, public health, software engineering), public health capacity building in foreign governments (public health, modeling), and modeled the effects of pandemics on interdependent infrastructure systems (big data, modeling, high performance computing):

Huff, A. G., Beyeler, W. E., Kelley, N. S., & McNitt, J. A. (2015). How resilient is the United States' food system to pandemics?. *Journal of Environmental Studies and Sciences*, 5(3), 337-347.

**B. Positions and Honors Academic**

2014-Present	Associate Vice President Data & Technology, EcoHealth Alliance
2014-Present	Adjunct Faculty, Columbia University
2013-2014	Senior Member of the Technical Staff, Sandia National Laboratories
2012	Teaching Assistant, University of Minnesota, School of Public Health
2012	Teaching Assistant, University of Minnesota, College of Science and Engineering
2011-2013	Research Fellow, University of Minnesota, Food Protection and Defense Institute

2008-2010 Research Assistant, University of Minnesota, Center for Interest Measurement Research  
 2008-2010 Judicial Extern, State of Minnesota, Office of Administrative Hearings, Administrative Court

### Other Experience and Professional Memberships

2006-2009 Program Assistant/Manager, United States Department of Veterans Affairs  
 2002-2010 United States Army & Minnesota Army National Guard, Infantryman

### C. Contribution to Science

I have always considered myself to be an interdisciplinary scientist, and over time I have become a strong and out spoken leader. By combining my military experience in war fighting, my training as an engineer, and my scientific experience I have the ability to effectively lead complex projects. I am not sure what my “greatest contribution to science” is, but the text that follows illustrates how I approach problems methodically and scientifically. Initially, investigating and developing solutions to improve security risk analysis and management intrigued me. After serving our country, I noticed that many of the risk-based decisions that the Department of Homeland Security were making seemed arbitrary and without any scientific basis. In my first attempt to make risk analysis more objective, I engineered a spatial modeling solution that analyzed big data from the National Neighborhood Crime Study, with a home built supercomputer, that could predict violent crime spatially in the U.S. with very high accuracy (tested by withholding data and Monte Carlo simulation):

- **Huff, A. G.** (2011). *Youth bulges, education, property crime, and income disparity: Utilizing geographic information systems to predict violence within the United States* (Master thesis). University of Minnesota, Minneapolis, MN.

After completing this work, I decided to continue improving the Risk Analysis methods employed by the government. Based on my excellent performance as a master student, I was appointed as a full salaried Research Fellow while earning my doctorate. As a Research Fellow, I was provided with all of the data used to evaluate national security risks in the United States’ food system. These big data were collected by subject matter experts and entered into software developed by DHS. This software, named FASCAT, also calculated risk and criticality scores for the government to allocate security mitigation resources. I suspected that there could be bias from SMEs during data collection so I created and implemented an observational study employing social science methods to identify potential biases:

- **Huff, A. G.**, Hodges, J. Kircher, A., & Kennedy, S. (2014). State officials’ perceptions of Food and Agriculture Sector Criticality Assessment Tool (FASCAT), food-system risk, and food defense funding. *Journal of Homeland Security and Emergency Management*, 0, 1-16.
- **Huff, A. G.**, Kircher, A., Hoffman, J., & Kennedy, S. P. (2013). The development and use of the Food and Agriculture Systems Criticality Assessment Tool (FASCAT), *Food Protection Trends*, 33, 218-223.

From these studies I found that there was likely biases introduced by SMEs during the data collection process. After these studies were collected, I analyzed the big data collected from the software to determine if any biases or problems with the software could be identified statistically.

- **Huff, A. G.**, Hodges, J. S., Kennedy, S. P., & Kircher, A. (2015). Evaluation of the Food and Agriculture Sector Criticality Assessment Tool (FASCAT) and the Collected Data. *Risk Analysis*.

After conducting this comprehensive analysis of the FASCAT software, the research indicated that a new software system should be created to evaluate risk, since the software and the SMEs were introducing biases into the final risk scores. From these studies, I learned how I could engineer a system to remove these biases and improve the accuracy of the risk estimates. However, one of the major challenges of creating a bias free risk analysis system was that these analyses always required humans to enter data that would describe the supply chain for thousands of food products – an arduous and impossible task. From analyzing thousands of food production systems in the world I learned that the vast majority of companies used *enterprise resource planning* software to track the purchase and sale of food products and ingredients. I then had the idea of extracting the time and location in sales records for each company’s record of purchases (food ingredients) and sales (processed food products). I then created software to automatically “link” each company’s purchases

and sales to reveal what the nations food supply looked like. The software created visualizes the relationships between food companies and their food products up and down the supply chain, from “farm to fork”, globally. Lastly, algorithms were devised and applied to data to quantify risk objectively. This resulted in a utility patent that is under evaluation and a significant change to how data is collected and analyzed in supply chains:

- **Huff, A. G.**, Kennedy, S. P., Kircher, A. L., & Hoffman, J. T. (2014). *U.S. Patent Application 14/212,749. Available at:* <https://www.google.com/patents/CA2846818A1>

## D. Research Support

### Ongoing Related

- **2015-2017**      **Defense Threats Reduction Agency (DTRA)**      **Principle Investigator**  
**Global Rapid Identification of Threats**      **HDTRA1-15-C-0041**  
 The purpose of this project is to forecast emerging infectious diseases by combining near real time big data from natural language processing, field collected infectious disease surveillance data, and environmental data.
- **2014-2019**      **USAID**      **Key Member**  
**PREDICT 2 – Modeling and Analytics**      **GHN- A-00-09-00010-00**  
 The purpose of this project is to build and test models, from field-collected data (humans and animals) to predict the emergence of viruses with pandemic potential spatially.

### Completed Related

- **2014-2015**      **Defense Threats Reduction Agency (DTRA)**      **Principle Investigator**  
**Global Rapid Identification of Threats**      **HDTRA1-13-C-0029**  
 The goal of this project is to develop software to diagnose emerging infectious diseases with bag data using natural language processing and novel algorithms.
- **2014-2015**      **Department of Agriculture (USDA)**      **Principle Investigator**  
**Mantle: Global Ranavirus Reporting System**      **14-JV-11261953-052**  
 The goal of this project was to develop a software system containing a user friendly graphic user interface, to merge, join, and share disparate tabular infectious disease datasets for amphibian diseases.
- **2010-2013**      **Department of Homeland Security**      **Key Member**  
**Criticality Spatial Analysis**      **2010-ST-061-FD0001**  
 In this project, a complex software system was developed to ingest and fuse supply chain big data from multiple companies automatically, to identify and display risks within and across the supply chain.

### Ongoing Unrelated

- **2014-2019**      **USAID**      **Key Member**  
**PREDICT 2 – Liaison to Jordan and Sudan**      **AID-O-AA-A-14-00102**  
 I am responsible for overseeing zoonotic infectious disease surveillance in Jordan and Sudan as it relates to emerging pandemic threats (e.g., MERS virus).

### Completed Unrelated

- **2013-2014**      **Department of Homeland Security**      **Principle Investigator**  
**Supply Chains, bioterrorism, and pandemics**      **Sandia National Laboratories**  
 Details are classified. Software development, big data fusion, and modeling.
- **2013-2014**      **Department of Veterans Affairs**      **Key Member**  
**Pandemic Influenza HPC modeling**      **Sandia National Laboratories**  
 Details are Official Use Only. Details are classified. Software development, big data fusion, and modeling.
- **2013-2014**      **Defense Threats Reduction Agency (DTRA)**      **Principle Investigator**  
**Animal Movement Models and Bioterrorism**      **Sandia National Laboratories**  
 Details are classified. Software development, big data fusion, and modeling.

Program Director/Principal Investigator (Last, First, Middle): Daszak, P.

**BIOGRAPHICAL SKETCH**

Provide the following information for the Senior/key personnel and other significant contributors in the order listed on Form Page 2.

Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Peter Daszak		POSITION TITLE President & Chief Scientist	
eRA COMMONS USER NAME: daszak			
EDUCATION/TRAINING			
INSTITUTION AND LOCATION	DEGREE	MM/YY	FIELD OF STUDY
Bangor University (UK)	BSc. (hons)	07/86	Zoology
University of East London (UK)	Ph.D	03/93	Infectious Diseases

**A. Personal Statement**

I am President and Chief Scientist of EcoHealth Alliance, a US-based organization that conducts research and outreach programs on global health, conservation and international development. My 20+ years of research has focused on understanding the causes of disease emergence in wildlife, livestock and people, particularly viral zoonotic diseases. My work has included identifying the bat origin of SARS, the causes of Nipah and Hendra virus emergence, producing the first ever global emerging disease 'hotspots' map, identifying the first case of a species extinction due to disease, coining the term 'pathogen pollution', and discovering the disease chytridiomycosis as the cause global amphibian declines. I am a member of the IOM's Forum on Microbial Threats, the NRC Committee to Advise the USGCRP, the Supervisory Board of the One Health Platform, the One Health Commission Council of Advisors, and the DHS-funded CEEZAD External Advisory Board. I have served on the IOM Committee on global surveillance for emerging zoonoses, the NRC committee on the future of veterinary research, the International Standing Advisory Board of the Australian Biosecurity CRC; and have advised the Director for Medical Preparedness Policy on the White House National Security Staff and the Director of the OST on global health issues. I am the EHA institutional lead for USAID-EPT-PREDICT and PREDICT-2, and serve on the Editorial Board of *Conservation Biology*, *One Health*, and *Transactions of the Royal Society of Tropical Medicine & Hygiene*, and am Editor-in-Chief of *Ecohealth*. I have authored over 300 scientific papers on emerging diseases.

**B. Positions and Honors****Positions and Employment**

1993-8 Senior Faculty Research Scientist, Kingston University  
 1998 Guest Researcher, Centers for Disease Control and Prevention (CDC)  
 1999-2001 Faculty Research Scientist, University of Georgia  
 2001- Adjunct Faculty, Tufts Univ. Sch. Veterinary Med.; Univ. Georgia; Columbia Univ.  
 2001-9 Executive Director, Consortium for Conservation Medicine, EcoHealth Alliance, New York  
 2009- President & Chief Scientist, EcoHealth Alliance New York.

**Other Experience and Professional Membership**

Keynote speaker Merieux Foundation Conference on Emerging paramyxoviruses, France (2000); UN Millenium Ecosystem Assessment: Lead Author, human infectious diseases (2006); NIH: ad hoc member, ZRG1 IDM-G 90 study section: Virology, Biodefense & Emerg. Diseases (2003-5); Editorial Board, *Conservation Biology* (Blackwell); Founding Co-Editor *EcoHealth* (Springer) (2004-10); NAS – Committee Member, Future Needs in Veterinary Research (2004-5); DIVERSITAS (UNESCO-ICSU): Member of Scientific Committee (2004-11; Treasurer 2007-11); NIAID: Steering Committee, workshop on virus-host shifts & emergence of new pathogens (2005); Australian Biosecurity Cooperative Research Center: International Standing Advisory Committee (2005-10); NIH: ad hoc member, ZRG1 IRAP-Q study section (infectious diseases, epidemiology) (2005-7); International EcoHealth Association: Founding board of directors, Treasurer (2006-11); CDC: ad hoc member, ZCD1 SGI, 09PAR07-231, R36 Research Dissertation Awards (2007); European CDC: Keynote speaker, future infectious disease threats (2008); NAS-IOM Committee Member, Global capacity for EID surveillance (2008-9); Scientific Advisory Board, NIAID Center of Excellence, avian influenza (CRISAR), UCLA (2008-9); Reviewer IOM report on Infectious Disease Movements in a Borderless World (2009); NIAID: Steering Committee, workshop on viruses from bats (2009); NAS-IOM Participant, workshop on H1N1, Committee on Emerging Microbial Threats (2009); NIH: ZRG1 IRAP-Q Review panel ARRA Challenge grants (2009); Organizing Committee, 1<sup>st</sup> International One Health Symposium, Australia

Program Director/Principal Investigator (Last, First, Middle): Daszak, P.

(2010); Member, Council of Advisors One Health Commission (2010-); Editor-in-Chief, *EcoHealth* (2010-); Scientific Advisory Board, Oxford Univ. Clinical Research Unit, Vietnam (2010-); Member of IOM Forum on Microbial Threats (2010-); External Advisory Board, CEEZAD (Center of Excellence for Emerging & Zoonotic Animal Diseases) – Dept. Homeland Security, Science & Technology Center of Excellence, Kansas State Univ. (2010-); Steering Committee, NIAID Workshop on Arboviruses (2011); Organizer IOM Forum on Microbial Threats briefing on MERS-CoV (2013); Chair, Scientific Steering Committee, Future Earth ecoHEALTH project (2014-); Editorial Board, *Transactions of the Royal Society of Tropical Medicine and Hygiene* (2014-); Member NRC Advisory Committee to advise the US Global Change Research Program (USGCRP) (2014-); Supervisory Board, One Health Platform (2015-); Senior Fellow, Center for Development Research (ZEFc), University of Bonn (2015-); NSF/NIH Ecology & Evolution of Infectious Diseases review panel (2015); Ed. Board *One Health* (2015-).

### Honors

Meritorious service award, CDC (1999); CSIRO silver medal for collaborative research (2000); Honored by the naming of a new species of centipede, *Cryptops daszaki* (*J Nat Hist* 2002; 36: 76-106) (2002); ISI Fast-breaking paper (2002); CBS 60 Minutes documentary on Nipah virus research; 6<sup>th</sup> Annual Lecturer, Medicine & Humanities, Texas A&M (2003); Editor's choice, *Science* (2006); Zayed International Prize for the Environment (2<sup>nd</sup>) (2006); Finalist, Director's Pioneer Award (2007); Discovery Channel documentary on Nipah virus research, Bangladesh (2008); Presidential Lecturer, University of Montana (2008); Elected member of the Cosmos Club 2012; Honored by the naming of a new species of parasite, *Isospora daszaki* (*Parasitol. Res.* 2013; 111:1463-1466) (2012); Awarded the Hsu-Li Distinguished Lectureship in Epidemiology (2013); Robert Leader Endowed Lecture in Food Safety, Michigan State Univ. (2015).

### C. Contribution to Science (Note: \* = Corresponding Author)

**Studies of wildlife ecology to understand emerging zoonoses.** It's been known for a long time that many emerging diseases are zoonotic with wildlife origins (e.g. HIV, Hantavirus pulmonary syndrome, West Nile virus). In the 1990s, new collaborations among ecologists studying wildlife population dynamics and medical scientists began to show that understanding disease dynamics in wildlife can allow better forecasting of disease risk in people, and ultimately help combat emerging pathogens. I was one of the leaders in this field, with a paper highlighting similarities between emerging diseases in people and wildlife that we published in *Science* (Daszak *et al.* 2000). I have continued this work with review papers on the linkages among biodiversity and health (Keesing *et al.*, 2010) and environmental change and health. Over the past 20 years, I have applied this approach to understand how the ecology of West Nile virus in birds can explain risk to people (e.g. Kilpatrick *et al.*, 2006), and how zoonotic diseases such as Nipah virus will likely alter their ecology in the future (e.g. Daszak *et al.*, 2013), as well as a range of studies to identify the ecological underpinnings of disease emergence.

- Daszak P, Cunningham AA, Hyatt AD (2000). Emerging infectious diseases of wildlife - threats to biodiversity and human health. **Science** 287: 443-449
- Keesing F, Belden LK, Daszak P, Dobson A, Harvell CD, Holt RD, Hudson P, Jolles A, Jones KE, Mitchell CE, Myers SS, Bogich T & Ostfeld RS. (2010). Impacts of biodiversity on the emergence and transmission of infectious diseases. **Nature** 468:647-652.
- Kilpatrick, A.M., Kramer, L.D., Jones, M.J., Marra, P.P. and Daszak, P. (2006). Host heterogeneity dominates West Nile virus transmission. **Proc Roy Soc B** 273: 2327-2333.
- Daszak P, Zambrana-Torellio C, Bogich TL, Fernandez M, Epstein JH, Murray KA, Hamilton H (2013). Interdisciplinary approaches to understanding disease emergence: The past, present and future drivers of Nipah virus emergence. **PNAS** 110: 3681-3688

**Discovery that diseases can cause extinction of species.** Biodiversity loss is one of the key threats to our planet, and one of the grand challenges cited by the National Academies of Science. Before the 1990s, it was known that diseases can cause outbreaks and sometimes smallscale population declines in wildlife. However, the role of diseases in largescale declines or even extinctions was controversial and widely disputed. Working with a veterinary colleague I was able to identify the first ever definitively-proven case of extinction of a species by an infectious agent (in *Partula* snails – Cunningham & Daszak, 1998). As one of the key members of a series of collaborations among veterinarians and ecologists, I identified a new fungal disease causing global declines and extinctions in amphibians (Berger *et al.*, 1998). I won the 2000 CSIRO Collaborative Research Medal for this work, and it has now led to a major focus of research in wildlife disease ecology which we summed up in a paper in *Science* in 2006 (Mendelson *et al.*, 2006).

Program Director/Principal Investigator (Last, First, Middle): **Daszak, P.**

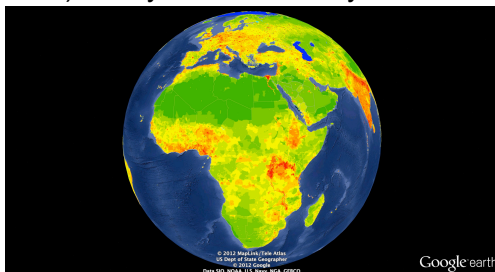
- Cunningham, A.A. & Daszak, P. 1998. Extinction of a species of land snail due to infection with a microsporidian parasite. **Conservation Biology** 12: 1139-1141.
- Berger L, Speare R, Daszak P, Green DE, Cunningham AA, Goggin CL, Slocombe R, Ragan MA, Hyatt AD, McDonald KR, Hines HB, Lips KR, Marantelli G, Parkes H (1998). Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. **PNAS** 95: 9031-9036.
- Mendelson JR, Lips KR, Gagliardo RW, Rabb GB, Collins JP, Diffendorfer JE, Daszak P et al. (2006). Confronting amphibian declines and extinctions. **Science** 313: 48.

**Research on the bat origins of SARS and Nipah virus.** The majority of EIDs are zoonotic in origin, with the majority of these originating in wildlife. Therefore, to prevent future disease emergence, it is critical that we better understand the wildlife origins of recent important zoonoses such as SARS and Nipah virus. As PI on three R01s my work has shown that SARS-CoV evolved from SARS-like CoVs in bats (Li *et al.*, 2005) and, by virus isolation, that giant fruit bats are the reservoir of Nipah virus (Rahman *et al.*, 2010). Collaborating with virologists in China, we have now isolated and characterized SL-CoVs from bats that use the same human host cell receptor (ACE-2) as SARS-CoV (Ge *et al.*, 2013). This work provides critical reagents and resources that now led to advances in viral understanding and ultimately will contribute to vaccine development by other groups. Our work has also identified why Nipah virus emerged in Malaysia (Pulliam *et al.*, 2012) and in Bangladesh, and the likely causes of Hendra virus emergence in Australia.

- Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, Zhang J, McEachern J, Field H, Daszak P, Eaton BT, Zhang S & Wang L-F (2005). Bats are natural reservoirs of SARS-like coronaviruses. **Science** 310: 676-679.
- Rahman SA, Hassan SS, Olival KJ, Mohamed M, Chang L-Y, Hassan L, Saad NM, Shohaimi SA, Mamat ZC, Naim MS, Epstein JH, Suri AS, Field HE, Daszak P & HERG (2010). Characterization of Nipah virus from naturally infected *Pteropus vampyrus* bats, Malaysia. **EID** 16: 1990-1993
- Ge X-Y, Li J-L, Yang X-L, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, Zhang Y-J, Luo C-M, Tan B, Wang N, Zhu Y, Crameri G, Zhang S-Y, Wang L-F, Daszak P\*, Shi Z-L\* (Co-Corresponding Authors) (2013). Isolation and characterization of a bat SARS-like Coronavirus that uses the ACE2 receptor. **Nature** 503: 535-538.
- Pulliam JRC, Epstein JH, Dushoff J, Rahman SA, Bunning M, HERG, Jamaluddin AA, Hyatt AD, Field HE, Dobson AP & Daszak P\* (Corresponding Author) and the Henipavirus Ecology Research Group (HERG). (2012). Agricultural intensification, priming for persistence, and the emergence of Nipah virus: a lethal bat-borne zoonosis. **J Roy Soc Interface** 9:89-101

#### **Analyzing drivers of emerging diseases to produce predictive EID 'hotspots' maps and models.**

Emerging infectious diseases are one of the biggest threats to global health. However, their emergence is sporadic, complex, and seemingly unpredictable. In the early 2000s I started to use ecological analytical approaches to see if there are patterns in disease emergence, and if these are predictable. By collating a database of all known prior EID events, identifying their point origins, and correcting for reporting biases, I published the first ever predictive 'hotspots' maps of where disease emergence is most likely (Jones *et al.*, 2008; Fig 1.). I have continued this approach, publishing predictive maps of H5N1 spread (Kilpatrick *et al.*, 2006), analyses of the likely dimensions of unknown viral diversity in mammals (Anthony *et al.*, 2013), maps



and predictive models of pandemic spread and a series of other papers analyzing EID risk (reviewed in Morse *et al.*, 2012).

Fig. 1. Heatmap showing the most likely sites of future disease emergence (for zoonotic diseases of wildlife origin). Red = highest risk, Green = lowest risk. The high risk regions have a combination of high human population density and high wildlife biodiversity – both correlate significantly with EID presence, corrected for reporting bias (from Jones *et al.*, 2008).

- Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, and Daszak P\*. (2008). Global trends in emerging infectious diseases. **Nature** 451:990-993
- Kilpatrick AM, Chmura AA, Gibbons DW, Fleischer RC, Marra PP & Daszak P (2006). Predicting the global spread of H5N1 avian influenza. **PNAS** 103: 19368-19373.

Program Director/Principal Investigator (Last, First, Middle): Daszak, P.

- Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrel CM, Solovyov A, Ojeda-Flores R, Arrigo NC, Islam A, Ali Khan S, Hosseini P, Bogich TL, Olival KJ, Sanchez-Leon MD, Karesh W, Goldstein T, Luby SP, Morse SS, Mazet JAK, Daszak P\*, Lipkin WI. (2013). A strategy to estimate unknown viral diversity in mammals. **MBio** 4(5): e00598-13.
- Morse SS, Mazet JAK, Woolhouse M, Parrish CR, Carroll D, Karesh WB, Zambrana-Torrel C, Lipkin WI, Daszak P\* (2012). Prediction and prevention of the next pandemic zoonosis. **Lancet** 380:1956-1965.

**Developing the field of 'One Health' and 'EcoHealth' in research and policy.** Throughout my career I have worked collaboratively across medical, veterinary and ecological disciplines. During the last couple of decades, I've tried to solidify this approach by becoming very active in the fields of 'One Health' and 'EcoHealth'. I've written a series of book chapters and reviews on these issues, serve on national and international committees for both fields, am an editor of the journal *One Health* and Editor-in-Chief of the journal *EcoHealth*, and am the CEO of the EcoHealth Alliance. I have begun to work with economic modelers to assess how efficient One Health approaches could be in global efforts to deal with emerging diseases (e.g. Pike *et al.*, 2014; Castillo-Chavez *et al.*, 2015). My aim is to use these analyses, and wealth of other published work to influence health and environmental policy so that the One Health approach is adopted widely for the benefit of public health and conservation. To help solidify this, I've been very active on the boards of, or working with national and intergovernmental agencies for health (IOM Forum on Microbial Threats; International EcoHealth Society; One Health Commission; One Health Platform) and Conservation (IUCN, CBD, DIVERSITAS, Future Earth). I have also published extensively on the policy implications of this work (e.g. Smith *et al.*, 2009; Rodriguez *et al.*, 2007).

- Pike J, Bogich TL, Elwood, SE, Finnoff DC, Daszak P\* (2014). Economic optimization of a global strategy to address the pandemic threat. (2014). **PNAS**.111:18519-18523.
- Castillo-Chavez C, Curtiss R, Daszak P\*, Levin SA, Patterson-Lomba O, Perrings C, Poste G, Towers S. (2015). Beyond Ebola: lessons to mitigate future pandemics. **Lancet Global Health** 3: e354-355.
- Smith KF, Behrens M, Schloegel LM, Marano N, Burgiel S, Daszak P\* (2009). Reducing the risks of the wildlife trade. **Science** 324:594-595.
- Rodríguez JP, Taber AB, Daszak P, Sukumar R, Valladares-Padua C, Padua S, Aguirre LF, Medellín R, Acosta M, Aguirre AA, Bonacic C, Bordino P, Bruschini J, Buchori D, González S, Mathew T, Mendez M, Mugijca L, Pacheco LF, Dobson AP, Pearl M (2007). Policy Forum: The globalization of conservation: A view from the South. **Science** 317: 755-756.

## D. Research Support

### Ongoing Research Support

USAID EPT PREDICT-2 Mazet (PI) 10/01/14 – 09/30/19  
 Conducting surveillance for novel pathogens in wildlife, livestock and people; characterizing human risk behavior; modeling risk of novel disease emergence; identifying mitigation strategies  
 Amount: \$35 Million subcontract from a \$100 Million award  
 Role: PI on Subcontract

1R01AI110964 Daszak (PI) 06/01/14 – 05/31/19  
 NIAID: Understanding the Risk of Bat Coronavirus Emergence  
 Bat ecological, human risk behavioral and virological studies to understand the risk of bat coronavirus emergence  
 Role: PI

NSF DEB 1414374 Perrings (PI) 10/15/14 - 10/14/17  
 NSF-NIH-USDA EEID, joint UK BBSRC BB/M008894/1  
 US-UK Collab: Risks of Animal and Plant Infectious Diseases through Trade (RAPID Trade)  
 Role: Co-Investigator

NSF Daszak (PI) 07/01/10-06/30/15  
 EcoHealthNet - a Research Coordination Network  
 Funding for student exchange and workshops to fuse veterinary science, ecology and human medical sciences  
 Role: PI 1R01GM100471 Perrings (PI) 09/15/11-06/30/15



Program Director/Principal Investigator (Last, First, Middle): Daszak, P.

HDTRA1 Huff (PI) 04/15/15 - 04/14/17  
 Office of Naval Research, Defense Threat Reduction Agency  
 Rapid identification of undiagnosed EID Events  
 Role: Co-Investigator

**Completed Research Support**

USAID EPT PREDICT-1 Mazet (PI) 10/01/09 – 09/30/14  
 Modeling hotspots for disease emergence and conducting surveillance in wildlife in hotspots for new emerging zoonoses  
 Amount: \$18 million subcontract on a \$75 million award  
 Role: PI on Subcontract

2 R01TW005869 Daszak (PI) 09/01/08 – 08/31/13  
 NIH Ecology of Infectious Diseases (Fogarty International Center)  
 The Ecology, Emergence and Pandemic Potential of Nipah virus in Bangladesh  
 To conduct mathematical modeling and fieldwork to understand the dynamics of Nipah virus in Bangladesh  
 Role: PI

NIAID Non-Biodefense Emerging Infectious Diseases  
 Risk of viral emergence from bats.  
 To model hotspots for bat viral diversity, identify & characterize new bat viruses & understand their pathology  
 Role: PI

NSF BCS 0826779 Daszak (PI) 10/01/08 – 03/31/12  
 NSF Human and Social Dynamics  
 AOC - HSD – Collaborative Research: Human-related factors affecting emerging infectious diseases  
 To analyze how socio-economic and environmental drivers predict risk of EIDs  
 Role: PI on lead proposal

R01TW005869 - supplemental Daszak (PI) 09/01/08 – 08/31/11  
 NIH EID (Fogarty International Center)  
 Supplemental funding: Predicting the risk of global H5N1 spread  
 This project will involve mathematical modeling and fieldwork in Bangladesh and China to understand risk of H5N1 spread.  
 Role: PI

NSF EF-062239 Kilpatrick (PI) 09/01/06 - 08/30/11  
 NSF/NIH: Ecology & Evolution of Infectious Diseases  
 Predicting spatial variation in West Nile virus transmission  
 Study interaction among WNV vector, reservoir host populations across an urban-to-rural gradient.  
 Role: Co-PI

R01 TW05869 Daszak (PI) 08/01/02 - 05/31/07  
 NIH/Fogarty International Center  
 Anthropogenic change & emerging zoonotic paramyxoviruses  
 To identify the cause of emergence of Nipah and Hendra viruses in Malaysia and Australia.  
 Role: PI

HSD 0525216 Daszak (PI) 10/15/05 - 10/14/06  
 National Science Foundation: Human and Social Dynamics  
 Collaborative Research: Socio-Economic and Environmental Drivers of Emerging Diseases  
 To analyze patterns of disease emergence globally and produce a broad risk assessment.  
 Role: PI



**BIOGRAPHICAL SKETCH**

Provide the following information for the Senior/key personnel and other significant contributors.  
Follow this format for each person. DO NOT EXCEED FIVE PAGES.

NAME: Brock Arnold

eRA COMMONS USER NAME (credential, e.g., agency login):

POSITION TITLE: Senior Software Developer

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Clarkson University	B.S.	05/2003	Computer Engineering

**A. Personal Statement**

I am a creative, experienced software developer, delivering innovative new software, adding functionality to existing software, and improving development processes at EcoHealth Alliance. In the past I have delivered these same results to multinational and SME SaaS companies, including a 10 year tenure with IBM, followed by software startup experience at two different companies. I have extensive experience in front and back end planning, design, and implementation, as well as user interface design. I have initiated software improvement and new application development projects to address undiscovered needs and carried the projects to completion. I approach each project as developer, designer, and user, and am excited to bring my expertise to the problem of managing scientific datasets and applying metadata.

Core technical competencies include: Java, C/ObjectiveC/C++, JavaScript, AngularJS, JQuery, iOS Development, OpenGL, RESTful principles, Mapbox/Leaflet mapping, Ionic Framework, and relational database design

**B. Positions and Honors****Positions and Employment**

2002	IBM Extreme Blue, Member of IBM's premier internship program
2003-2013	International Business Machines, Staff Software Engineer
2013-2014	Activate Networks, Inc, Senior Software Engineer
2014-2015	Cargotel/Primotus, Senior Software Engineer
2015-Current	EcoHealth Alliance, Senior Software Developer

**Other Experience and Professional Memberships**

2006-2009	Faculty Advisory Board Member, Worcester Polytechnic Institute, Faculty Diversity Advisory
2008-2013	IBM New England Diversity Council Board Member
2015	Member, American Indian Science and Engineering Society

## **C. Contribution to Science**

2015      Technical Lead, Global Rapid Identification Tool System development team. Global infectious disease outbreak warning software, designed to monitor online news articles, using natural language processing and machine learning techniques, for mentions of infectious disease threats. Developed in partnership with ProMED, the International Society for Infectious Disease, and Kitware.

## **D. Research Support**

As a Senior Software Engineer at Activate Networks, Inc., I architected, designed, and deployed SAAS applications that helped organizations across the country understand their social networks - utilizing technologies like MongoDB, Bootstrap, Handlebars, JQuery, Grunt, Node.js, AWS, Java, as well as numerous other open source frameworks. This work is based on the social-science research by recognized leaders in network science, including Professor Nicholas Christakis, MD, PhD, MPH, of Yale University (formerly, of Harvard University); Professor James Fowler, PhD, of UC San Diego; and Professor Rob Cross, DBA, of the University of Virginia.

As a Senior Software Engineer at a Baltimore, Maryland based stealth startup, I helped architect and plan the release of a new enterprise scale user mobile workflow platform that uses AngularJS, Jade templates, Less, Grunt, Gulp, Leaflet/Mapbox, Bootstrap, and a RESTful API provided by a Scala backend. I also was instrumental in designing and implementing a new hybrid mobile application using the Ionic Framework and Cordova plugins for forms based user generated content submission.

As a Staff Software Engineer working on IBM's SmartCloud SaaS platform, I discovered a flaw that would overwhelm our entire platform, losing critical customer data, during the initial onboarding of a large enterprise client's system just 2 days before the client was to go live. Thinking outside the box, and within a very small window of time, I architected and delivered a custom monitoring and throttling mechanism that addressed this critical flaw. The client successfully onboarded and provisioned over 200,000 users successfully and on time.

I founded Longhouse Interactive LLC, and Sweet Zo's, Inc, both boutique mobile app development companies. A successful concrete ready-mix operation in Northern NY is currently the launch customer of a custom truck tracking and routing application from Sweet Zo's. This application enables rapid turnaround at the central batch plant to enable quick, efficient, and accurate delivery of material to clients while reducing driver error and improving customer satisfaction.

As a personal research project, I wrote a custom real time stock trading application against the Interactive Brokers Java API, and initiated a paper trading simulation. The simulation tracked all last trade information for all members of the S&P 500 and executed trades based on signals generated by a custom algorithm. The trading algorithm was generated and refined by utilizing a genetic algorithm using JGAP (Java Genetic Algorithms Package) which provided basic genetic mechanisms that applied evolutionary principles to the problem set.

**BIOGRAPHICAL SKETCH**

Provide the following information for the Senior/key personnel and other significant contributors.  
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Noam Ross

eRA COMMONS USER NAME (credential, e.g., agency login):

POSITION TITLE: Disease Ecologist

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Brown University	B.S.	05/2006	Environmental Science
University of California-Davis	Ph.D.	09/2015	Ecology

**A. Personal Statement**

The Mantle project's goal is to create an open-source, open-access platform for the collection and sharing of biosurveillance data to improve the detection and study of emerging infectious diseases. My background in computational disease ecology and open-source development and training allows me to support this project in several ways. My recent dissertation research consisted of developing and analyzing models for emerging fungal diseases in both plants and wildlife, and importantly, the creation of methods and open-source tools for simulating, fitting, and performing optimization using such models. As a contributor and review board member of the ROpenSci project, I develop and review open-source software packages for scientists that interface with a diverse set of online data sources and services. This gives me experience with the complexity of the task of working with varied data standards and formats, standardizing and testing tools, and best practices in open-source collaborative development. My experience working with field scientists on computational skills will allow me to support the project in creating user-friendly interfaces and training modules. I am a workshop instructor for the Software and Data Carpentry foundations, and the founder of the Davis R Users' group, a statistical computing training network with more than 400 members. This experience will allow me to support this project in designing tools to the needs and technical skills of a broad variety of field and lab researchers. In summary, the Mantle project is a natural fit for my research, development, and teaching experience.

**B. Positions and Honors****Positions and Employment**

2015- Disease Ecologist, EcoHealth Alliance, New York, NY  
 2010-2015 Graduate Researcher, University of California-Davis  
 2007-2009 Senior Analyst, Corporate Environmental Strategy and Governance, GreenOrder, New York, NY  
 2006-2007 Analyst, Corporate Environmental Strategy and Governance, GreenOrder, New York, NY  
 2006 Contract Researcher: Energy Efficient Products Initiative, Wal-Mart, Providence, RI

**Other Experience and Professional Memberships**

2015- Instructor, Software Carpentry Foundation  
 2015- Instructor, Data Carpentry Foundation

2015-	Review Board Member, ROpenSci
2014-	Contributor, ROpenSci
2013-	Member, Ecological Society of America
2012-2015	Founder and Organizer, Davis R Users' Group
2012-2013	Member, NSF IGERT.org advisory board

#### Awards and Fellowships

2012	Don Dahlsten Memorial Grant, California Forest Pest Council
2012	NSF IGERT Bridge Fellowship
2010	NSF IGERT Traineeship in Rapid Environmental Change
2010	UC Davis Graduate Ecology Fellowship

### **C. Contribution to Science**

My recent dissertation addressed modeling emerging fungal disease epidemics using a framework traditionally used for parasites of stable populations. While the mathematical basis of these models for populations at or approximately at equilibrium is well established, their dynamic properties are less well known due to analytical intractability, and this they are little-used in emerging diseases and epidemics. My work showed how and where these models diverged from other, traditional models in their dynamical properties, and identified statistical patterns that could be used to identify where these models are appropriate. I developed numerical tools for their simulation, modeling and control.

1. **Noam Ross** (in prep), Fungal Disease and Stage Structure: Modeling Short and Long-Term Dynamics
2. **Noam Ross**, James N. Sanchirico and Alan Hastings (in prep) Optimal control for Individual-Based Models of Disease
3. **Noam Ross** (in prep), Modeling dynamics of dispersion in macroparasite epidemics with the Conway-Maxwell-Poisson distribution
4. cmp: a package for fitting under- and over-dispersed data with the Conway-Maxwell-Poisson distribution. [github.com/noamross/cmp](https://github.com/noamross/cmp)

Another forthcoming manuscript describes a new mathematical method, and open-source code, to calculate establishment and extinction probabilities for small populations with continuous traits such as size structure or disease load.

5. Sebastian Schreiber and **Noam Ross** (in prep) Individual-Based Integral Projection Models: The Role of Size-Structure on Extinction Risk And Establishment Success

Through the Software Carpentry (SWC) and Data Carpentry foundations I have led training workshops in computational skills and data science for field and lab researchers in a variety of disciplines. I have also been a contributor to SWC's curricula and its open-source platform for developing, testing, and disseminating materials and best pedagogical practices for these topics via a large community of contributors.

6. Gabriel Devenyi and Christina Koch (eds), 63 authors including **Noam Ross**. Software Carpentry: The Unix Shell. Software Carpentry Foundation, Version 5.3, May 2015, 10.5281/zenodo.27355

#### Other Publications:

7. Carl Boettiger\*, **Noam Ross**\*, Alan Hastings (2013) Early Warning Signals: The Charted And Uncharted Territories. Theoretical Ecology <http://dx.doi.org/10.1007/s12080-013-0192-6> (**\*Co-equal authors**)
8. Kate Fuller, David Kling, Kaelin Kroetz, **Noam Ross**, and James N. Sanchirico (2013) Economics and Ecology of Open-Access Fisheries. In: Shogren, J.F., (ed.) Encyclopedia of Energy, Natural Resource, and Environmental Economics, Vol. 2 Encyclopedia of Energy, Natural Resource, and Environmental Economics p.39-49. Amsterdam: Elsevier. <http://dx.doi.org/10.1016/B978-0-12-375067-9.00114-5>

#### **D. Research Support**

W911NF-13-1-0305 Hastings (PI) 9/1/13-8/31/16  
Army Research Office Mathematical Sciences Core Program  
Dynamics at Intermediate Time Scales and Management of Ecological Populations  
Role: Supported Graduate Student

EF-0622770 Rizzo (PI) 8/23/06-8/31/11  
NSF Ecology of Infectious Disease Program  
Collaborative Research: Sudden Oak Death: Feedback Between a Generalist Pathogen, Hosts, and  
Heterogeneous Environments at Multiple Spatial and Temporal Scales  
Role: Supported Graduate Student

**BIOGRAPHICAL SKETCH**

NAME: Streichert, Laura C.

eRA COMMONS USER NAME (credential, e.g., agency login): lcstreichert

POSITION TITLE: Executive Director

**EDUCATION/TRAINING**

INSTITUTION AND LOCATION	DEGREE	Completion Date MM/YYYY	FIELD OF STUDY
Princeton University	AB	1981	Biology
Stanford University	PhD	1991	Neuroscience
University of Washington	MPH	2005	Health Services
University of Washington	Certificate	1998	Admin. & Management
University of Oregon	postdoc	1994	Developmental Neuroscience
University of Washington	Sr. Fellow	1997	Neuroscience

**A. Personal Statement**

Laura Streichert, PhD, MPH, brings subject matter expertise and perspective to this project as an experienced executive leader, neuroscientist, biomedical researcher, and public health specialist. As Executive Director of the International Society for Disease Surveillance (ISDS), Laura works across disciplines and sectors researchers and surveillance practitioners at the local, state, national, and global levels, and other stakeholders worldwide to forge new collaborative ventures to advance the science and practice of disease surveillance. Laura additionally brings knowledge and experience in system-level approaches; qualitative and quantitative research methodologies; public health informatics; program management and evaluation; real-time (i.e., syndromic) surveillance; One Health Surveillance; knowledge management; building Communities of Practice; surveillance policy; and assessment to identify public health priorities, practices, and workforce development needs. Most recently, as PI on the DTRA-funded *Analytic Solutions for Challenges in Real-Time Biosurveillance* project, Dr. Streichert is leading efforts to convene public health practitioners who have identified analytic problems with the statisticians, computer scientists, and other developers to create solutions. To date, she has orchestrated the development of two consultancies on 1) asyndromic cluster detection; and 2) evaluating disease-forecasting models for public health utility. The ability to engage end-users from the onset of the development process helps to ensure the practical utility of any solution developments. This is particularly relevant to this EcoHealth Alliance project where Dr. Streichert will provide expertise and tap into the collective expertise of the ISDS membership to provide community input to the project. The paper below details the process and outcomes of ISDS's approach:

- a. Faigen Z, Deyneka L, Ising A, Neill D, Conway M, Fairchild G, Gunn J, Swenson D, Painter I, Johnson L, Kiley C, **Streichert LC**, Burkom H. (2015). Cross-Disciplinary Consultancy to Bridge Public Health Technical Needs and Solution Development Expertise: Asyndromic Surveillance Use Case. Public Health Reports (in press).

## B. Positions and Honors

1991 – 1994 Postdoctoral Fellow, University of Oregon, Institute for Neuroscience  
 1995 – 1998 Senior Research Fellow, University of Washington, Department of Biological Structure  
 1998 – 1999 Science Education Program Manager, HutchLab, Fred Hutchinson Cancer Research Center  
 1999 – 2001 Outreach Programs Manager, Northwest Association for Biomedical Research  
 2001 – 2003 Development Officer, Seattle Biomedical Research Institute  
 2004 Instructor, Northwest Center for Public Health Practice Summer Institute  
 2005 – 2006 Manager of Center Operations, Univ. of Washington Exploratory Center for Obesity Research  
 2007 – 2008 Assistant Director, University of Washington Center for Obesity Research  
 2008 – 2010 Public Health Consultant, Streichert Strategic Consulting  
 2010 – 2010 Program Director, Harvard School of Public Health Center for Public Health Preparedness  
 2011 – Executive Director, International Society for Disease Surveillance (ISDS)

## Other Experience and Professional Memberships

2011 – Joint Public Health Informatics Taskforce  
 2011 – Council of State and Territorial Epidemiologists (CSTE), affiliate member  
 2009 – Newton Medical Reserve Corps  
 1981 – 1984 U.S. Peace Corps, Nepal

## C. Contribution to Science

1. My early work in basic science addressed unanswered questions related to the structural and functional changes that neurons and synaptic connections undergo during development or following injury of the nervous system. The results demonstrated changes in the anatomical, cellular, and electrophysiological features of the developing nervous system in a variety of animal models. The NIH training I received as a PhD student and postdoctoral fellow in neuroscience helped to build my foundational skills in research design, data analysis, and quantitative methods.
  - a. **Streichert, L.C.**, Birnbach, C.D. & Reh, T.A. (1999). A diffusible factor from normal retinal cells promotes rod photoreceptor survival in an *in vitro* model of retinitis pigmentosa. *J Neurobiol.* 39: 475-490.
  - b. **Streichert, L.C.**, Pierce, J.T., Nelson, J.A. & Weeks, J.C. (1997). Steroid hormones act directly to trigger segment-specific programmed cell death of identified motoneurons *in vitro*. *Dev Biol.* 183: 95-107.
  - c. **Streichert, L.C.** & Sargent, P.B. (1992). The role of acetylcholinesterase in denervation supersensitivity in the frog cardiac ganglion. *J Physiol.* 445: 249-260.
2. Innovative strategies for obesity prevention require the application of a social-ecological model that understands and addresses the root causes of the problem. At the NIH-funded Exploratory Center for Obesity Research (ECOR) at the University of Washington, and subsequently as an independent contractor with a public health agency, I worked to bridge disciplines and sectors to examine environmental determinants of obesity from the cellular to the community levels. This project also utilized my ability to weave common threads across disciplines and scientific levels in a *Lab to Leadership* approach. For example, I was part of a team that connected basic research on the hypothalamic control of glucose metabolism in developing mice to possible mechanisms for the effects of high sugar diets in children or looking at the relationship between fast food locations and socioeconomic status using GIS approaches. The results and conceptual underpinnings of this translational research were applied to the activities of the Public Health—Seattle & King County Obesity Prevention Initiative and the creation of a Community Action Plan that has been used to guide policy making and community-based interventions in Pierce County, Washington. The outcomes of this research are detailed, in part in the following publications:

- a. Hurvitz PM, Moudon AV, Rehm CD, **Streichert LC**, Drewnowski A. (2009). Arterial roads and area socioeconomic status are predictors of fast food restaurant density in King County, WA. *Int J Behav Nutr Phys Act*. Jul 24;6:46.
  - b. Podrabsky M, **Streichert LC**, Levinger D, Johnson DB. (2007). Campus-community-school partnerships to evaluate a multi-component nutrition intervention. *Public Health Reports*. Jul-Aug;122(4):566-9.
  - c. **Streichert LC**. (2009). *Pierce County Community Action Plan for Active Living and Healthy Eating*. Tacoma-Pierce County Health Department. <http://www.tpchd.org/files/library/f88dec11fd52279e.pdf>
  - d. **Streichert LC**, Johnson DB, Drewnowski, A. (2008). Reframing Obesity Prevention. *Northwest Public Health*. 25 (1): 6-7. (Guest editor of special issue: *Preventing Obesity: Moving Beyond Individual Responsibility*). <http://www.nwpublichealth.org/archives/s2008>
3. As Executive Director at ISDS I have initiated and led a number of projects to advance the science and practice of disease surveillance through assessments and evaluations, standards development, knowledge management, community building, workforce development, and advocacy, particularly in the area of syndromic surveillance. My broad-based experience and expertise has enabled me to create unique action-oriented partnerships to develop tools and approaches to data collection, management, analysis, visualization, inter-jurisdictional sharing, and application to decision-making. I have been an active contributor to the groundwork leading to the launching of the National Syndromic Surveillance Program. In addition, my leadership has helped to expand ISDS's efforts in the realm of integrating animal, human, and environmental surveillance through the advancement of One Health strategies. Recent work in creating functional requirements for analytic solutions to particular public health challenges is aimed to identify and address barriers to real-time surveillance of pressing public health challenges. Working with the surveillance community, I have helped to identify and disseminate information on topics of priority of the surveillance community, including the need for sustainable approaches to the surveillance enterprise:
- a. Mirza N, Reynolds TL, Coletta M, Suda K, Soyiri I, Markle A, Leopold H, Lenert L, Samoff E, Siniscalchi A, Streichert LC. (2013). Steps to a Sustainable Public Health Surveillance Enterprise. *Online J Public Health Inform*. 2013;5(2). Available at: <http://ojphi.org/ojs/index.php/ojphi/article/view/4703>

### **Complete List of Published Work in MyBibliography:**

<http://www.ncbi.nlm.nih.gov/sites/myncbi/1r5K3ZvZkma5l/bibliographay/48710397/public/?sort=date&direction=ascending>

## **D. Research Support**

### **Ongoing Research Support**

#### **Defense Threat Reduction Agency (DTRA)**

HDTRA1-15-C-0004      Streichert (PI)      01/06/15-01/5/17

#### *Analytic Solutions for Challenges in Real-Time Biosurveillance*

The goal of this study is to advance analytic capabilities in real-time biosurveillance (BSV) by expediting next-generation solutions to currently intractable problems through focused consultancies that join problem owners from civilian and military public health agencies with solution developers in academia, industry, and government.

Role: PI



### **Skoll Global Threats Fund (SGTF)**

#14-02503                      Streichert (PI)                      11/1/14-12/31/15

#### ***Advancing One Health Surveillance Strategies***

The purpose of this project is to strengthen alliances between animal and public health surveillance professionals for the identification and development of practical One Health Surveillance approaches to challenges in disease detection and response. In parallel, given that most of the emerging infectious diseases with an animal source originate from resource constrained settings, the project will provide a platform to share experiences on the use of mobile technologies in these contexts and discuss their application for capturing, transmitting, and analyzing health information in a timely and efficient fashion.

Role: PI

### **Completed Research Support**

#### **National Association of County and City Health Officials (NACCHO)**

#2014-090801                      Streichert (PI)                      10/1/14-6/30/15

#### ***Surveillance practice and technical assistance needs among local health departments in the US.***

ISDS, in collaboration with NACCHO coordinated, executed, and analyzed a rigorous nation-wide assessment of current disease surveillance practices and technical assistance needs of local public health departments within the U.S.

Role: P

## RESEARCH &amp; RELATED BUDGET - SECTION A &amp; B, Budget Period 1

ORGANIZATIONAL DUNS\*: 0770900660000

Budget Type\*: ☒ Project ☐ Subaward/Consortium

Enter name of Organization: EcoHealth Alliance, Inc.

Start Date\*: 11-01-2015

End Date\*: 10-31-2016

Budget Period: 1

**A. Senior/Key Person**

Prefix	First Name*	Middle Name	Last Name*	Suffix	Project Role*	Base Salary (\$)	Calendar Months	Academic Months	Summer Months	Requested Salary (\$)*	Fringe Benefits (\$)*	Funds Requested (\$)*
1 . Dr.	Andrew		Huff		PD/PI	140,000.00	3.00			35,000.00	12,285.00	47,285.00
2 . Dr.	Peter		Dazsak		Senior Research Scientist	183,300.00	0.02			4,230.00	1,484.00	5,714.00
3 . Mr.	Brock		Arnold		Senior Software Developer	110,000.00	3.00			27,498.00	9,651.00	37,149.00
4 . Mr.	Christopher		Allen		Data Scientist	80,000.00	4.50			30,768.00	10,799.57	41,567.57
<b>Total Funds Requested for all Senior Key Persons in the attached file</b>												
<b>Additional Senior Key Persons:</b>		File Name:									<b>Total Senior/Key Person</b>	<b>131,715.57</b>

**B. Other Personnel**

Number of Personnel*	Project Role*	Calendar Months	Academic Months	Summer Months	Requested Salary (\$)*	Fringe Benefits*	Funds Requested (\$)*
	Post Doctoral Associates						
	Graduate Students						
	Undergraduate Students						
1	Secretarial/Clerical	3.00			10,000.00	3,509.86	13,509.86
2	Key EHA Staff as necessary	3.00			26,500.00	9,301.22	35,801.22
3	Total Number Other Personnel					Total Other Personnel	49,311.08
Total Salary, Wages and Fringe Benefits (A+B)							181,026.65

RESEARCH &amp; RELATED Budget {A-B} (Funds Requested)

**RESEARCH & RELATED BUDGET - SECTION C, D, & E, Budget Period 1****ORGANIZATIONAL DUNS\*:** 0770900660000**Budget Type\*:** ☒ Project ☐ Subaward/Consortium**Organization:** EcoHealth Alliance, Inc.**Start Date\*:** 11-01-2015**End Date\*:** 10-31-2016**Budget Period:** 1

<b>C. Equipment Description</b>	
List items and dollar amount for each item exceeding \$5,000	
<b>Equipment Item</b>	<b>Funds Requested (\$)*</b>
<b>Total funds requested for all equipment listed in the attached file</b>	
<b>Total Equipment</b>	
<b>Additional Equipment:</b> File Name:	

D. Travel	Funds Requested (\$)*
1. Domestic Travel Costs ( Incl. Canada, Mexico, and U.S. Possessions)	16,284.00
2. Foreign Travel Costs	0.00
<b>Total Travel Cost</b>	<b>16,284.00</b>

E. Participant/Trainee Support Costs	Funds Requested (\$)*
1. Tuition/Fees/Health Insurance	
2. Stipends	
3. Travel	
4. Subsistence	
5. Other:	
<b>Number of Participants/Trainees</b>	<b>Total Participant Trainee Support Costs</b>

RESEARCH &amp; RELATED Budget (C-E) (Funds Requested)

**RESEARCH & RELATED BUDGET - SECTIONS F-K, Budget Period 1****ORGANIZATIONAL DUNS\*:** 0770900660000**Budget Type\*:** ☒ Project ☐ Subaward/Consortium**Organization:** EcoHealth Alliance, Inc.**Start Date\*:** 11-01-2015**End Date\*:** 10-31-2016**Budget Period:** 1

<b>F. Other Direct Costs</b>	<b>Funds Requested (\$)*</b>
1. Materials and Supplies	6,045.00
2. Publication Costs	4,500.00
3. Consultant Services	
4. ADP/Computer Services	1,000.00
5. Subawards/Consortium/Contractual Costs	50,000.00
6. Equipment or Facility Rental/User Fees	
7. Alterations and Renovations	
8. Student Tuition	38,592.00
9. Meeting Costs	500.00
10. Recruiting Costs	700.00
<b>Total Other Direct Costs</b>	<b>101,337.00</b>

<b>G. Direct Costs</b>	<b>Funds Requested (\$)*</b>
<b>Total Direct Costs (A thru F)</b>	<b>298,647.65</b>

<b>H. Indirect Costs</b>			
<b>Indirect Cost Type</b>	<b>Indirect Cost Rate (%)</b>	<b>Indirect Cost Base (\$)</b>	<b>Funds Requested (\$)*</b>
1. EcoHealth Alliance F&A Rate	44.10	260,057.06	114,685.16
<b>Total Indirect Costs</b>			<b>114,685.16</b>
<b>Cognizant Federal Agency</b>			
(Agency Name, POC Name, and POC Phone Number)			

<b>I. Total Direct and Indirect Costs</b>	<b>Funds Requested (\$)*</b>
<b>Total Direct and Indirect Institutional Costs (G + H)</b>	<b>413,332.81</b>

<b>J. Fee</b>	<b>Funds Requested (\$)*</b>

<b>K. Budget Justification*</b>	File Name: 1234-Budget Justification.pdf (Only attach one file.)
---------------------------------	---

RESEARCH &amp; RELATED Budget (F-K) (Funds Requested)

## RESEARCH &amp; RELATED BUDGET - SECTION A &amp; B, Budget Period 2

ORGANIZATIONAL DUNS\*: 0770900660000

Budget Type\*: ☒ Project ☐ Subaward/Consortium

Enter name of Organization: EcoHealth Alliance, Inc.

Start Date\*: 11-01-2016

End Date\*: 10-31-2017

Budget Period: 2

**A. Senior/Key Person**

Prefix	First Name*	Middle Name	Last Name*	Suffix	Project Role*	Base Salary (\$)	Calendar Months	Academic Months	Summer Months	Requested Salary (\$)*	Fringe Benefits (\$)*	Funds Requested (\$)*
1 . Dr.	Andrew		Huff		PD/PI	140,000.00	3.00			35,000.00	12,285.00	47,285.00
2 . Dr.	Peter		Dazsak		Senior Research Scientist	183,300.00	0.02			4,230.00	1,484.00	5,714.00
3 . Mr.	Brock		Arnold		Senior Software Developer	110,000.00	3.00			27,498.00	9,651.00	37,149.00
4 . Mr.	Christopher		Allen		Data Scientist	80,000.00	4.50			30,768.00	10,799.57	41,567.57
<b>Total Funds Requested for all Senior Key Persons in the attached file</b>												
<b>Additional Senior Key Persons:</b>		File Name:									<b>Total Senior/Key Person</b>	<b>131,715.57</b>

**B. Other Personnel**

Number of Personnel*	Project Role*	Calendar Months	Academic Months	Summer Months	Requested Salary (\$)*	Fringe Benefits*	Funds Requested (\$)*
	Post Doctoral Associates						
	Graduate Students						
	Undergraduate Students						
1	Secretarial/Clerical	3.00			10,000.00	3,509.86	13,509.86
2	Key EHA Staff as necessary	3.00			26,500.00	9,301.22	35,801.22
3	Total Number Other Personnel					Total Other Personnel	49,311.08
Total Salary, Wages and Fringe Benefits (A+B)							181,026.65

RESEARCH &amp; RELATED Budget {A-B} (Funds Requested)

**RESEARCH & RELATED BUDGET - SECTION C, D, & E, Budget Period 2****ORGANIZATIONAL DUNS\*:** 0770900660000**Budget Type\*:** ☒ Project ☐ Subaward/Consortium**Organization:** EcoHealth Alliance, Inc.**Start Date\*:** 11-01-2016**End Date\*:** 10-31-2017**Budget Period:** 2**C. Equipment Description**

List items and dollar amount for each item exceeding \$5,000

<b>Equipment Item</b>	<b>Funds Requested (\$)*</b>
<b>Total funds requested for all equipment listed in the attached file</b>	
<b>Total Equipment</b>	
<b>Additional Equipment:</b> File Name:	

<b>D. Travel</b>	<b>Funds Requested (\$)*</b>
1. Domestic Travel Costs ( Incl. Canada, Mexico, and U.S. Possessions)	5,572.50
2. Foreign Travel Costs	13,414.50
<b>Total Travel Cost</b>	<b>18,987.00</b>

<b>E. Participant/Trainee Support Costs</b>	<b>Funds Requested (\$)*</b>
1. Tuition/Fees/Health Insurance	
2. Stipends	
3. Travel	
4. Subsistence	
5. Other:	
<b>Number of Participants/Trainees</b>	<b>Total Participant Trainee Support Costs</b>

RESEARCH &amp; RELATED Budget (C-E) (Funds Requested)

**RESEARCH & RELATED BUDGET - SECTIONS F-K, Budget Period 2****ORGANIZATIONAL DUNS\*:** 0770900660000**Budget Type\*:** ☒ Project ☐ Subaward/Consortium**Organization:** EcoHealth Alliance, Inc.**Start Date\*:** 11-01-2016**End Date\*:** 10-31-2017**Budget Period:** 2

<b>F. Other Direct Costs</b>	<b>Funds Requested (\$)*</b>
1. Materials and Supplies	150.00
2. Publication Costs	4,500.00
3. Consultant Services	
4. ADP/Computer Services	1,000.00
5. Subawards/Consortium/Contractual Costs	50,000.00
6. Equipment or Facility Rental/User Fees	
7. Alterations and Renovations	
8. Student Tuition	38,592.00
9. Meeting Costs	500.00
<b>Total Other Direct Costs</b>	<b>94,742.00</b>

<b>G. Direct Costs</b>	<b>Funds Requested (\$)*</b>
<b>Total Direct Costs (A thru F)</b>	<b>294,755.65</b>

<b>H. Indirect Costs</b>			
<b>Indirect Cost Type</b>	<b>Indirect Cost Rate (%)</b>	<b>Indirect Cost Base (\$)</b>	<b>Funds Requested (\$)*</b>
1. EcoHealth Alliance F&A Rate	44.10	256,163.65	112,968.17
<b>Total Indirect Costs</b>			<b>112,968.17</b>
<b>Cognizant Federal Agency</b>			
(Agency Name, POC Name, and POC Phone Number)			

<b>I. Total Direct and Indirect Costs</b>	<b>Funds Requested (\$)*</b>
<b>Total Direct and Indirect Institutional Costs (G + H)</b>	<b>407,723.82</b>

<b>J. Fee</b>	<b>Funds Requested (\$)*</b>

<b>K. Budget Justification*</b>	<b>File Name: 1234-Budget Justification.pdf</b>
	(Only attach one file.)

RESEARCH &amp; RELATED Budget {F-K} (Funds Requested)

## RESEARCH &amp; RELATED BUDGET - SECTION A &amp; B, Budget Period 3

ORGANIZATIONAL DUNS\*: 0770900660000

Budget Type\*: ☒ Project ☐ Subaward/Consortium

Enter name of Organization: EcoHealth Alliance, Inc.

Start Date\*: 11-01-2017

End Date\*: 10-31-2018

Budget Period: 3

## A. Senior/Key Person

Prefix	First Name*	Middle Name	Last Name*	Suffix	Project Role*	Base Salary (\$)	Calendar Months	Academic Months	Summer Months	Requested Salary (\$)*	Fringe Benefits (\$)*	Funds Requested (\$)*
1 . Dr.	Andrew		Huff		PD/PI	140,000.00	3.00			35,000.00	12,285.00	47,285.00
2 . Dr.	Peter		Dazsak		Senior Research Scientist	183,300.00	0.02			4,230.00	1,484.00	5,714.00
3 . Mr.	Brock		Arnold		Senior Software Developer	110,000.00	3.00			27,498.00	9,651.00	37,149.00
4 . Mr.	Christopher		Allen		Data Scientist	80,000.00	4.50			30,768.00	10,799.57	41,567.57
<b>Total Funds Requested for all Senior Key Persons in the attached file</b>												
<b>Additional Senior Key Persons:</b>		File Name:									<b>Total Senior/Key Person</b>	<b>131,715.57</b>

## B. Other Personnel

Number of Personnel*	Project Role*	Calendar Months	Academic Months	Summer Months	Requested Salary (\$)*	Fringe Benefits*	Funds Requested (\$)*
	Post Doctoral Associates						
	Graduate Students						
	Undergraduate Students						
1	Secretarial/Clerical	3.00			10,000.00	3,509.86	13,509.86
2	Key EHA Staff as necessary	3.00			26,500.00	9,301.22	35,801.22
3	Total Number Other Personnel					Total Other Personnel	49,311.08
Total Salary, Wages and Fringe Benefits (A+B)							181,026.65

RESEARCH &amp; RELATED Budget {A-B} (Funds Requested)



**RESEARCH & RELATED BUDGET - SECTION C, D, & E, Budget Period 3****ORGANIZATIONAL DUNS\*:** 0770900660000**Budget Type\*:** ☒ Project ☐ Subaward/Consortium**Organization:** EcoHealth Alliance, Inc.**Start Date\*:** 11-01-2017**End Date\*:** 10-31-2018**Budget Period:** 3**C. Equipment Description**

List items and dollar amount for each item exceeding \$5,000

Equipment Item	Funds Requested (\$)*
<b>Total funds requested for all equipment listed in the attached file</b>	
<b>Total Equipment</b>	
<b>Additional Equipment:</b> File Name:	

**D. Travel****Funds Requested (\$)\***

1. Domestic Travel Costs ( Incl. Canada, Mexico, and U.S. Possessions)	13,143.00
2. Foreign Travel Costs	0.00
<b>Total Travel Cost</b>	<b>13,143.00</b>

**E. Participant/Trainee Support Costs****Funds Requested (\$)\***

1. Tuition/Fees/Health Insurance	
2. Stipends	
3. Travel	
4. Subsistence	
5. Other:	
<b>Number of Participants/Trainees</b>	<b>Total Participant Trainee Support Costs</b>

RESEARCH &amp; RELATED Budget (C-E) (Funds Requested)

**RESEARCH & RELATED BUDGET - SECTIONS F-K, Budget Period 3****ORGANIZATIONAL DUNS\*:** 0770900660000**Budget Type\*:** ☒ Project ☐ Subaward/Consortium**Organization:** EcoHealth Alliance, Inc.**Start Date\*:** 11-01-2017**End Date\*:** 10-31-2018**Budget Period:** 3

<b>F. Other Direct Costs</b>	<b>Funds Requested (\$)*</b>
1. Materials and Supplies	150.00
2. Publication Costs	4,500.00
3. Consultant Services	
4. ADP/Computer Services	1,000.00
5. Subawards/Consortium/Contractual Costs	50,000.00
6. Equipment or Facility Rental/User Fees	
7. Alterations and Renovations	
8 . Student Tuition	38,592.00
9 . Meeting Costs	500.00
<b>Total Other Direct Costs</b>	<b>94,742.00</b>

<b>G. Direct Costs</b>	<b>Funds Requested (\$)*</b>
<b>Total Direct Costs (A thru F)</b>	<b>288,911.65</b>

<b>H. Indirect Costs</b>			
<b>Indirect Cost Type</b>	<b>Indirect Cost Rate (%)</b>	<b>Indirect Cost Base (\$)</b>	<b>Funds Requested (\$)*</b>
1 . EcoHealth Alliance F&A Rate	44.10	250,319.65	110,390.97
<b>Total Indirect Costs</b>			<b>110,390.97</b>
<b>Cognizant Federal Agency</b>			
(Agency Name, POC Name, and POC Phone Number)			

<b>I. Total Direct and Indirect Costs</b>	<b>Funds Requested (\$)*</b>
<b>Total Direct and Indirect Institutional Costs (G + H)</b>	<b>399,302.62</b>

<b>J. Fee</b>	<b>Funds Requested (\$)*</b>

<b>K. Budget Justification*</b>	File Name: 1234-Budget Justification.pdf
	(Only attach one file.)

RESEARCH &amp; RELATED Budget {F-K} (Funds Requested)

## **BUDGET JUSTIFICATION**

### **A. Senior/Key Personnel:**

PI, Dr. Andrew Huff, will commit 3 months (520 hours) per year each year of this budget. He will be primarily responsible for the management of Mantle's research and development. Dr. Huff will oversee all aspects of the project, and manage communication and collaboration with subawardees. He will also contribute to data analysis and manuscript writing.

Senior Research Scientist, Dr. Peter Daszak will commit 1 day every other month (8 hours x 6 days= 48 hours) to this budget. He will also contribute to research and development, overseeing the project, general management, and will contribute to manuscript writing.

Senior Software Developer, Brock Arnold will commit 3 months (520 hours) each year to this budget. Mr. Arnold will serve as Senior Software Developer on the project and will manage the developer and data science staff. Brock will contribute to the coding of the Mantle app and manage Mantle's code base.

Data Scientist, Toph Allen will commit approximately 4.5 months (800 hours) per year each year of this budget. Mr. Allen will be primarily responsible for research and development on Mantle's metadata capabilities. Mr. Allen will also test and train models and machine learning algorithms to be incorporated into Mantle. He will also contribute to data analysis and manuscript writing.

EHA key personnel (engineers and scientists) will be used as needed throughout the project in small percentages compared to the overall allotment of staff time. Supplementing the development of Mantle with other key staff members ensures that Mantle will be successful when obstacles to research and development are inevitably encountered.

### **B. Other Personnel:**

Ms. Robyn Schreiber will commit 3 months (520 hours) each year to this budget. Ms. Schreiber will serve a Secretarial/Clerical role on the project and will manage administrative duties including scheduling, recording and preparing reports, facilitating effective communication and being a point of contact for inquiries.

Key EHA Staff as necessary will commit time amounting to approximately 3 months (520 hours) each year to this budget. Staff time will be utilized in the area of their specific expertise. Key EHA Staff includes contribution from expert disease modeler Dr. Noam Ross and contribution from an EHA software developer.

Fringe benefits for all years are calculated for EcoHealth Alliance's federally approved rate of 35.1% of salary amount.

### **C. Equipment:**

No equipment costing more than \$5,000 will be purchased.

### **D. Travel:**

### **Domestic Travel**

\$16,284.00 is requested for Year 1 domestic travel to attend/present at two conferences and one meeting with NIH stakeholders. The PD/PI and 2 personnel will attend each conference and meeting.

Travel expenses for the first conference, the Joint Statistical Meetings 2016 Conference in Chicago, Illinois, are calculated *per person* for 3 personnel as follows: 5 night, 6 day trip \$71 for meals and incidentals (note: 75% of per diem rate is applied to the arrival and departure days, as per federal guidelines) [  $(\$71 \times 4 \text{ days}) + (\$71 \times 75\% \times 2 \text{ days}) = \$390.50$  ]; \$192 for hotel costs [  $\$192 \times 5 = \$960$  ]; \$500 for round-trip airfare; and \$200 for local transportation to/from the airport to the hotel and to/from the hotel to the conference.

Travel expenses for the second conference, the International Society for Disease Surveillance Conference in Denver, Colorado, are calculated *per person* for 3 personnel follows: 3 night, 4 day trip \$66 for meals and incidentals [  $(\$66 \times 2 \text{ days}) + (\$66 \times 75\% \times 2 \text{ days}) = \$231$  ]; \$163 for hotel costs [  $\$163 \times 3 = \$489$  ]; \$500 for round-trip airfare; and \$300 for local transportation to/from the airport to the hotel and to/from the hotel to the conference.

\$5,572.50 is requested annually (Year 1, Year 2 and Year 3) as travel expenses for the PD/PI and 2 personnel to travel to the BD2K consortium, held within the United States at a location to be determined by NIH staff. Expenses are calculated, using the District of Columbia per diem rates, *per person* as follows: as follows: 4 night, 5 day trip (note: per diems are based off of the District of Columbia) \$71 for meals and incidentals [  $(\$71 \times 3 \text{ days}) + (\$71 \times 75\% \times 2 \text{ days}) = \$319.50$  ]; \$222 for hotel costs [  $\$222 \times 4 = \$888$  ]; \$500 for round-trip airfare; and \$150 for local transportation to/from the airport to the hotel and to/from the hotel to the consortium.

\$8,299.50 is requested additionally for Year 3 domestic travel to attend/present at one conference and one launch event.

Travel expenses for the conference, the International Workshop on Big Data in Bioinformatics and Healthcare Informatics, are based off per diem rates for the District of Columbia and are calculated *per person* for 3 people (PD/PI and 2 personnel) as follows: 3 night, 4 day trip \$71 for meals and incidentals [  $(\$71 \times 2 \text{ days}) + (\$71 \times 75\% \times 2 \text{ days}) = \$248.50$  ]; \$222 for hotel costs [  $\$222 \times 3 = \$666$  ]; \$500 for round-trip airfare; and \$200 for local transportation to/from the airport to the hotel and to/from the hotel to the conference.

Travel expenses for the product (Mantle) launch in New York, New York are calculated *per person* for 2 personnel (remote personnel traveling to New York) as follows: 2 night, 3 day trip \$71 for meals and incidentals [  $(\$71 \times 1 \text{ day}) + (\$71 \times 75\% \times 2 \text{ days}) = \$177.50$  ]; \$268 for hotel costs [  $\$268 \times 2 = \$536$  ]; \$500 for round-trip airfare; and \$150 for local transportation to/from the airport to the hotel and to/from the hotel to the launch event.

### ***International Travel***

\$13,414.50 is requested for Year 2 international travel of the PD/PI and 2 personnel to attend/present at the International Society for Infectious Diseases Conference in London, England. Travel expenses for the ISID are calculated *per person* as follows: 4 night, 5 day trip \$175 for meals and incidentals [  $(\$175 \times 3 \text{ day}) + (\$175 \times 75\% \times 2 \text{ days}) = \$787.50$  ]; \$346 for hotel costs [  $\$346 \times 2 = \$1384$  ]; \$2,000 for round-trip airfare; and \$300 for local transportation to/from the airport to the hotel and to/from the hotel to the conference.

### **E. Participant/Trainee Support Costs:**

There are no participant/trainee support costs.

### **F. Other Direct Costs**

#### ***Materials/Supplies***

We request \$6,045.00 in Year 1 for equipment to facilitate product (program and app) development including 1 Nexus 9 tablet (\$399), 1 iPad Air 2 (\$599), 1 128 GB iPhone 6 (\$399), 1 MacBook Pro (\$1,499), 1 Thunderbolt Display (\$999), 5 books or reference materials calculated at \$30 each (\$150) and consumables for enhanced work productivity (\$2,000). Consumables will include office supplies such as ink, paper, post-its, pens, USB drives, and occasional lunches.

#### ***Publication Fees***

We request funding for two publications each year. Publication fees are calculated based on average publication fee for PLoS (Public Library of Science) journals: 2 publications x \$2,250 per publication = \$4,500.

#### ***Computer Services***

We request \$160 each year to cover an annual cost for domain name and security certificates. We request \$840 each year to cover the cost of Google Apps Services and data hosting. Cloud Application expenses (google apps with unlimited storage) are calculated as follows: \$10 per user, per month X 12 months X 7 users = \$840.

#### ***Subawards/Consortium/Contractual Costs***

We request contractual support for two organizations: the International Society for Disease Surveillance and Clango at \$25,000 each/year for Year 1, Year 2 and Year 3.

#### ***Master of Science / Master of Arts Student Tuition***

We request funding for tuition for two research students pursuing a related master's degree at an accredited university. Tuition expenses are calculated *per student* as follows: \$1608 per credit X 12 credits each year = \$19,296.

#### ***Meeting Costs***

We request \$500 each year for meeting costs. This funding will support two all-day meetings with the partners at the EcoHealth Alliance office in NY. It covers printing, copying, additional support for meeting materials, conference lines, and other supplies to enhance partner collaboration.

#### *Recruiting*

We request recruiting funds for the Year 1 to hire additional software development help as necessary. Recruiting expenses are calculated as follows: 2 job listings x \$350 per 30-day listing = \$700.

#### **H. Indirect Costs**

We request the EcoHealth Alliance federally approved indirect cost rate of 44.1% on all applicable direct costs. Indirect is taken only on the first \$25,000 for each consortium/contractual agreement in each year. As there are two (one to International Society for Disease Surveillance and the other to Clango), a total of \$50,000 ( $\$25,000 \times 2$ ) is requested as indirect costs on consortium/contractual/subaward agreements. This is not included as part of direct cost calculations.

**RESEARCH & RELATED BUDGET - Cumulative Budget**

	Totals (\$)	
Section A, Senior/Key Person		395,146.71
Section B, Other Personnel		147,933.24
Total Number Other Personnel	9	
Total Salary, Wages and Fringe Benefits (A+B)		543,079.95
Section C, Equipment		
Section D, Travel		48,414.00
1. Domestic	34,999.50	
2. Foreign	13,414.50	
Section E, Participant/Trainee Support Costs		
1. Tuition/Fees/Health Insurance		
2. Stipends		
3. Travel		
4. Subsistence		
5. Other		
6. Number of Participants/Trainees		
Section F, Other Direct Costs		290,821.00
1. Materials and Supplies	6,345.00	
2. Publication Costs	13,500.00	
3. Consultant Services		
4. ADP/Computer Services	3,000.00	
5. Subawards/Consortium/Contractual Costs	150,000.00	
6. Equipment or Facility Rental/User Fees		
7. Alterations and Renovations		
8. Other 1	115,776.00	
9. Other 2	1,500.00	
10. Other 3	700.00	
Section G, Direct Costs (A thru F)		882,314.95
Section H, Indirect Costs		338,044.30
Section I, Total Direct and Indirect Costs (G + H)		1,220,359.25
Section J, Fee		

## PHS 398 Cover Page Supplement

OMB Number: 0925-0001

## 1. Project Director / Principal Investigator (PD/PI)

Prefix: Dr.  
First Name\*: Andrew  
Middle Name:  
Last Name\*: Huff  
Suffix:

## 2. Human Subjects

Clinical Trial? ☒ No ☐ Yes  
Agency-Defined Phase III Clinical Trial?\* ☐ No ☐ Yes

## 3. Permission Statement\*

If this application does not result in an award, is the Government permitted to disclose the title of your proposed project, and the name, address, telephone number and e-mail address of the official signing for the applicant organization, to organizations that may be interested in contacting you for further information (e.g., possible collaborations, investment)?

☒ Yes ☐ No

## 4. Program Income\*

Is program income anticipated during the periods for which the grant support is requested? ☐ Yes ☒ No

If you checked "yes" above (indicating that program income is anticipated), then use the format below to reflect the amount and source(s). Otherwise, leave this section blank.

Budget Period*	Anticipated Amount (\$)*	Source(s)*
.....	.....	.....
.....	.....	.....
.....	.....	.....
.....	.....	.....
.....	.....	.....



## PHS 398 Cover Page Supplement

### 5. Human Embryonic Stem Cells

Does the proposed project involve human embryonic stem cells?\* ☒ No ☐ Yes

If the proposed project involves human embryonic stem cells, list below the registration number of the specific cell line(s) from the following list: [http://grants.nih.gov/stem\\_cells/registry/current.htm](http://grants.nih.gov/stem_cells/registry/current.htm). Or, if a specific stem cell line cannot be referenced at this time, please check the box indicating that one from the registry will be used:

Cell Line(s): ☐ Specific stem cell line cannot be referenced at this time. One from the registry will be used.

### 6. Inventions and Patents (For renewal applications only)

Inventions and Patents\*: ☐ Yes ☐ No

If the answer is "Yes" then please answer the following:

Previously Reported\*: ☐ Yes ☐ No

### 7. Change of Investigator / Change of Institution Questions

☐ Change of principal investigator / program director

Name of former principal investigator / program director:

Prefix:

First Name\*:

Middle Name:

Last Name\*:

Suffix:

☐ Change of Grantee Institution

Name of former institution\*:

## PHS 398 Research Plan

Please attach applicable sections of the research plan, below.

OMB Number: 0925-0001

1. Introduction to Application (for RESUBMISSION or REVISION only)	
2. Specific Aims	1245-Specific Aims.pdf
3. Research Strategy*	1246-Research Strategy.pdf
4. Progress Report Publication List	
Human Subjects Sections	
5. Protection of Human Subjects	1247-Human Subjects Protection.pdf
6. Inclusion of Women and Minorities	
7. Inclusion of Children	
Other Research Plan Sections	
8. Vertebrate Animals	
9. Select Agent Research	
10. Multiple PD/PI Leadership Plan	
11. Consortium/Contractual Arrangements	1248-Consortium and Contractual Arrangements.pdf
12. Letters of Support	1249-Letters of Support.pdf
13. Resource Sharing Plan(s)	1250-Resource Sharing Plan.pdf
Appendix (if applicable)	
14. Appendix	

## SPECIFIC AIMS

The ability to prevent and respond to future public health crises, like those posed by emerging infectious diseases (e.g., pandemic flu, Ebola, and MERS) depends largely on our ability to rapidly and efficiently collect, combine, and analyze vast amounts of disparate data. In practice, cleaning, joining, and analyzing data is greatly limited by the nature of the datasets, which are generally vast, have varying structures, and are in disparate formats. Datasets are generally kept private and are inaccessible to scientists and policymakers and are in disparate formats such that a single implementation of the data combination is impossible or overly burdensome. By creating a system to host and integrate disparate datasets and make them interoperable we would enable many new applications in health research and biosurveillance, allowing, for example, heuristics to be applied to the data to generate new knowledge and improve the allocation of resources to combat infectious disease threats worldwide. This is why software that can collect, combine, and integrate varying data structures and formats is needed. **The proposed work** addresses these integration issues and we will create a software platform called **Mantle**.

**Specific Aim 1: Mantle will use existing ontologies, controlled vocabularies, and create interactive tools for applying metadata.** We will: **(a)** establish universal data standards for health data and metadata in multiple languages using Bing translator API (back-end); **(b)** incorporate existing open source and open access ontologies and vocabularies created by CEDAR in our cloud-based software that combines and integrates disparate data sources in a variety of formats (back-end); and, **(c)** obfuscate human health data, from health researchers, biosurveillance streams, or field based research, to maintain regulatory compliance and privacy when necessary (e.g., HIPAA, SOX; IRB; IACUC; back-end).

**Specific Aim 2: Mantle will use crowdsourcing and machine learning to refine, improve and automate metadata assignment.** We will: **(a)** use machine learning algorithms to suggest ontologies that might fit the contents of datasets (front and back-end); **(b)** use crowdsourcing, via services like Mechanical Turk and through analyzing users' ontology selections, to improve upon the machine learning algorithms (back-end); and, **(c)** crowdsource refinements to existing health ontologies to improve Mantle's performance and select appropriate matches (front and back-end).

**Specific Aim 3: Mantle will incentivize users to share, combine, and export cleaned and joined data.** We will host and create an online community which incentivizes users to use Mantle by alleviating major pain points barriers in data curation. We will: **(a)** develop a friendly and efficient user interface for researchers, policy makers, and the general public (including corporate entities) to interact with health and biosurveillance data (front end); **(b)** enable advanced users to download shared data combined, with other open access data layers that are relevant to health research and automatically apply metadata (front-end); and, **(c)** create a system to combine disparate data sources (APIs or uploaded data sets) and a RESTful API for others to ingest or use Mantle data (back-end).

## RESEARCH STRATEGY

### A. SIGNIFICANCE

Data integration could alleviate many of the problems facing the biosurveillance and health fields. The process of data integration takes semantically incompatible data, from disparate sources, and in different file types, and merges them into one widely understandable format with metadata to make these data discoverable and available to scientists broadly (1). A successful data integration system must solve multiple problems (1-4) and Mantle will address these key unresolved issues:

**Issue 1. Dataset availability:** Many datasets are owned by entities who keep them private for reasons related to intellectual property and research concerns, or for security or privacy reasons (5-7). However, even public datasets often lack discoverability, are not stored in a centralized location, and do not have a searchable index of datasets due to the time and difficulty formatting data to the required standards (8). Mantle will incentivize researchers to make datasets available and will link public data in a centralized and searchable database.

**Issue 2. Structural heterogeneity:** Data are stored in different software formats and structures. Tabular data may be stored in Excel spreadsheets or CSV (or other character-delimited text files), or in a variety of relational database systems. Spatial data may be stored in a tabular format, or may be in a large number of spatial formats or semi structured formats used by various GIS systems and their pre and post processing software (9). Mantle will combine disparate forms of data and will provide homogenous data to users.

**Issue 3. Semantic incompatibility:** Data, even in the same format, are incompatible in a number of ways. Numerical (continuous or ordinal) variables, like cases of a disease, or temperature, are generally aggregated (summed, averaged) by some period of time and/or some spatial area. The level of aggregation is not (and should not be) standardized, but there is no standard way to refer to the level of organization. Nominal values refer to the same semantic entity (i.e., a particular species) are differently represented in individual dataset schemas and are irreconcilable without mediation. Semantic integration is a challenging and complex problem (4, 10). Broadly speaking, semantic integration involves *mapping* the language in which individual datasets are expressed, the *local schemas*, with a *global schema* (1-4). Two prevailing methods, *global-centric* or *global-as-view* and *source-centric* or *local-as-view*, approach the problem slightly differently, and trade advantages and disadvantages in ease of querying items in the global database, schema flexibility, and others. Mantle will use *global-centric* or *global-as-view*, and *source-centric* or *local-as-view*, to combine previously incompatible datasets.

**Issue 4. Ontologies:** Ontologies are essential to data integration. They provide a structured, logical definition of a domain's concepts and their relationships (11). In data integration, they serve as mediators, mapping the relationships between heterogeneous representations of concepts in individual datasets (3, 11-14). Ontologies themselves are not standardized. However, methods exist for aligning and evolving ontologies, including the work of Stanford's

CEDAR group, bioontology.org, and Protégé (15-17). Additionally, the specification of ontologies is codified in a set of web standards centered around linked data, including Resource Documentation Framework (RDF; 18), Web Ontology Language (OWL; 19-20), and others. A new standard, recently approved, standardizes a metadata for tabular data within this framework (19). As technology coalesces around these standards for ontology specification, and ontologies are developed to codify scientific data collection (12), new types of software are possible, using online, standardized, curated libraries of ontologies to integrate disparate datasets and these ontologies make it possible to query vast amounts of data in a unified interface. Mantle will use machine learning and curated libraries to assign metadata.

**Barriers in scientific and institutional culture:** Solving the conceptual and technological challenge of matching datasets to ontologies will not itself address the problem of data sharing. Cultural and institutional barriers must be addressed, both programmatically (e.g., by grantors mandating data sharing and metadata annotation policies) and by providing tools and education to make the concept of metadata annotation accessible to scientists (21-22, 16). In one survey of 1329 scientists (23), only 26% were satisfied with tools for metadata preparation. 32% reported dissatisfaction, and 42% neither agreed nor disagreed, indicating “either [that] they truly are indifferent or they are unsure about what metadata means”. The authors speculated the latter is true—46% answered that they do not currently use metadata to describe their dataset. Some of the inertia in the adoption of data-sharing practices can be explained by the fact that large majorities of the scientists surveyed report being satisfied with their data collection, searching, cataloguing, and short-term storage processes. Conversely, it is promising that even larger numbers (78%) report that they would be willing to openly share some of their data in a central repository; fewer report that they would be willing to openly share all of their data.

**Identified problems:** Data owners have financial, political and other reasons for keeping data private (5). It is not possible to integrate these private datasets with public data, and even for dataset owners integration is still beset by the existing challenges. Also, scientists with domain-specific expertise lack formal training in computer science skills to conduct complex multi-dataset merges (14). Current software solutions do not bridge the skills or understanding gap between scientists and the data problems they must solve. (23-24)

**The metadata and data integration problem remains unsolved:** Existing software packages do not provide tools to apply metadata to both health and biosurveillance data. Some services allow scientists to upload and store datasets, but treat datasets as monolithic chunks of data. (25). Dryad, for example, is an open-source archive of datasets. These services generally host detailed metadata for each dataset, provide DOI numbers for datasets for publication purposes, and provide a search interface. However, they often only allow the annotation of dataset-level metadata, not allowing the use of ontologies for data integration (e.g., Dryad) or use metadata standards which do not conform to current linked data specifications (e.g., KNB’s Ecological Metadata Language; 26). Both examples given are part of DataONE (27), an NSF-funded collaboration working toward better data practices in science. While these formats provide

valuable services for data portability and sharing, they lack the data integration aspect that is crucial to furthering the objectives of biomedical science.

Previous attempts to create metadata systems encountered lack of awareness and acceptance of metadata standards (23-24). For example, Ecological Metadata Language (EML) is a metadata standard defined as a large XML schema encoding properties of ecological datasets of various types. In 2008, the Long-Term Ecological Research (LTER) program mandated a move to EML for all datasets (24). This move, however, has been notably slow. Scientists involved with its application were interviewed, and they found numerous points of friction (24). Sites were given the standard as an XML specification—250 pages long—and software provided for its application were often incompatible with previously existing systems, and how to reconcile it with other existing metadata systems and data structures was not always clear. Despite its complexity, other researchers found it too limited. Applying EML to existing datasets was a “mostly unfunded mandate”, and had no immediate payoff for scientists; despite voicing support for EML and metadata in general, “...when the rubber hits the road, an unfunded mandate to be altruistic (and simultaneously to lose one’s own tried-and-true local bricolage with data structures) does not prove highly attractive” (24).

Metadata are noted as a “product”, a monolithic structure, which does not account for the iterative and ad-hoc nature of metadata use in day-to-day research (24). Scientists use datasets for different purposes, and in doing so, will describe their structure and composition in discourse; this constitutes “metadata as process”, and is not a single, comprehensive structure. Attempts to impose concrete metadata structures on scientific data of a certain type have, then, met with at best tepid success and “almost-use” (24). For metadata annotation to become common practice among scientists, it must become approachable and offer a value proposition to scientists. Mantle will offer both an understandable, accessible way to annotate datasets, and add facilitate the process of collecting and managing data, so that scientists adopt them as part of their commonplace data workflows.

**Broader Impacts:** 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 (28-30) Therefore sensitive data must be safeguarded, including personally identifying information, and security will be incorporated into the design of Mantle from the outset. Databases which incorporate public data must also protect against the injection of false data (31). 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 integrated ecological and social data to improve understanding of how human and natural systems interact to change health outcomes and affect disease emergence.

## B. INNOVATION

This project is an innovative fusion of software engineering, data science, and public health research. The incorporation of existing vocabularies and ontologies developed by CEDAR, will help establish universal data standards universally and has not yet been attempted. This approach will allow us to better understand gaps in the creation, and assignment of ontologies, across languages. For about the past decade, data portability and availability has been pursued by scientists and mandated by governments, but has not fundamentally improved (16). Previous studies have found that structural semantic heterogeneity are significant obstacles to overcome when combining data and when assigning metadata (32, 33) and these problems are exacerbated by the lack of formal training in data integration of most scientists in biomedical fields and across academia (14). 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 crowdsourced dataset annotations. By automating these processes, we hope that Mantle will be used outside of biomedical research, as it directly addresses a problem that is common to throughout scientific disciplines. For example, many fields of science are becoming data intensive, and thus reliant on cyberinfrastructure. An example is the use of databases as virtual laboratories in astronomy, where an astronomer can make and record a large number of virtual observations (14).

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 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. These mechanisms will also enable the use of Mantle with data which cannot be shared for other reasons, and facilitate the *partial* sharing of such datasets, so that they can maximally contribute to other scientific endeavors and the public good. Furthermore, Mantle's metadata assignment features will exist in a user-focused, community-based platform. This will mean that disparate datasets, which are part of Mantle's system and have been assigned ontologies, can be merged and aggregated with unprecedented ease. Past interfaces have not been user-focused, and have relied on scientists with no expertise in metadata application to do the technical data management work. We are designing ways to enable 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.

### C. APPROACH

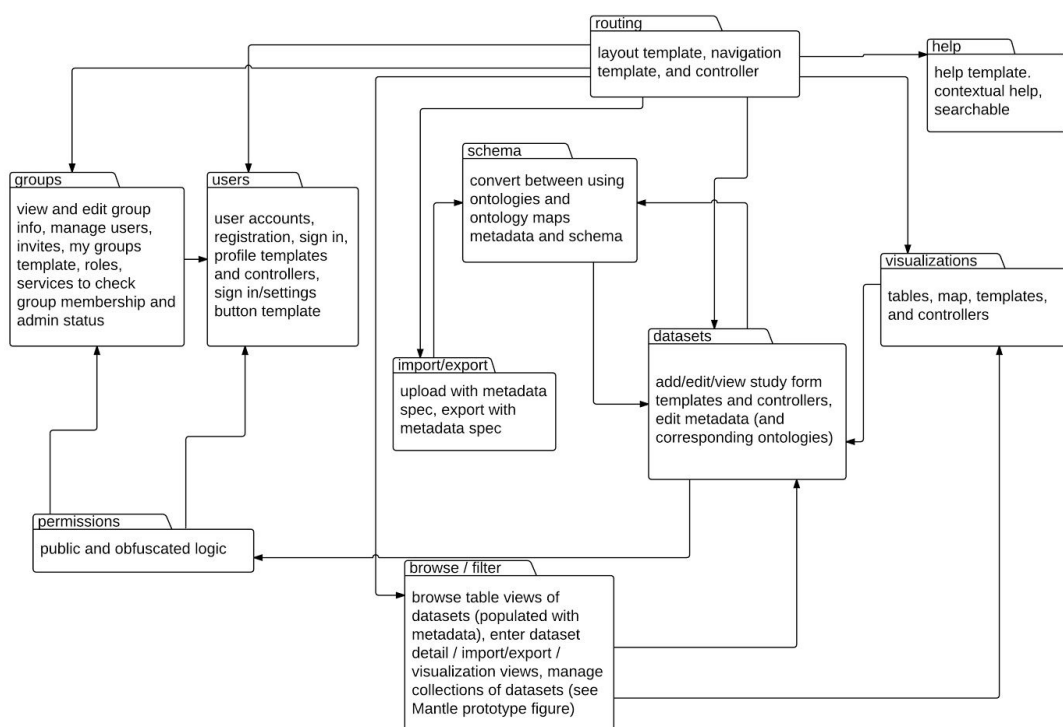
**Share, combine, and export cleaned and joined data:** Mantle will use a cloud-based federated database to handle the integration of disparate data storage types. The front-end graphic user interface will provide the portal for users to input, query, and download data. Mantle will use the Resource Description Framework (RDF), which is a standard data interchange model for web-based applications (18). Mantle will use XML syntax to define both the relationship between two entities and the two ends of any link to accommodate evolving data schemas. These Uniform Resources Identifiers (URIs) will allow Mantle to organize and

index data from sources (e.g., tabular data), based on best practices developed by the W3C working group (34). In the case of a CSV file, all of the columns, rows, and cells will be converted to an annotated tabular data model in Mantle. Mantle will be a Linked Data Platform Resource (LDPR) (35), which is an HTTP resource that can be modified and accessed using HTTP code and is managed through a LDP server (35). It will store and retrieve big data using Amazon's NoSQL service, Amazon's Relational Database Service, and MongoDB cluster hosted on ec2 instances. Data sharing *will occur via the* traditional server/client model in Amazon S3 and via an open source peer-to-peer model using Bittorrent protocol and open source compression algorithms to increase speed of transfer from a distributed network of repositories. The entire Mantle environment can be templated into a public Amazon Machine Image for others to use on demand. Amazon provides multiple tools to allow API calls from many different languages, and development environments, thereby enabling scientists from around the world to use Mantle. Overall, these solutions enable researchers to outsource information management and administration and enables them to focus on their primary research internationally.

To enable efficient querying across many linked datasets, Mantle will need to implement a triplestore (36) or leverage an existing implementation of one. Triplestores are designed specifically for the purpose of querying the subject–predicate–object expressions that form the basis of the RDF data model. SPARQL is the most prevalent query language for triplestores (37). Although complicated, it is a very powerful language in its capacity to use many types of data at once. For example, long, trailing queries with many, nested prepositions like the following can be expressed with relative economy: “find all the symptoms of diseases transmitted by bats that migrate to locations with a longitude greater 30 degrees that patient X traveled to in the last 3 months.” Mantle will have in interface that simplifies the process of building these types of complex queries for the user.

Mantle will become a Digital Object Identifier (DOI) provisioner so that hosted datasets can be cited and used in published works. DOIs may include permanent URLs that allow users to locate referenced data. We will likely use DataCite (38) to register DOIs for datasets that are uploaded to Mantle because they use an XML schema. Existing tools like Dryad currently use DataCite for this purpose (38). Data will be stored in the cloud hosted by Amazon Web Services on the GovCloud for enhanced security and privacy. See Figure 1 for a detailed illustration of Mantle's schematic.





**Figure 1.** Diagram of components of Mantle's functions, permissions, and data visualizations.

**Metadata:** Mantle must solve the data integration challenges including dataset availability, structural heterogeneity, and semantic heterogeneity. Mantle will facilitate the collection, storage, versioning, and sharing of data, as well as fitting into existing analytical toolchains, reducing friction in scientific workflows. We will allow scientists to upload datasets in a number of common formats (**Aim 1**). In particular, we will support tabular data in Excel, .csv, and other character-delimited files, spatial raster and vector data in formats supported by the Geospatial Data Abstraction Library (GDAL) software package, and textual data (**Aim 1**). Users will be able to export data tabular and spatial data in supported file formats, and always have access to the original file (**Aim 1**). Uploaded datasets will be converted to Mantle's native data types (WCSV, JSON-LD, and other RDF-compatible formats) (**Aim 1, 3**).

Users will be guided through an interactive process of matching their dataset with cloud-hosted ontologies (**Aim 1**). We will develop and implement machine learning algorithms to suggest ontologies that might fit the contents of the dataset (**Aim 2**). We will present possible ontology matches to users and allow them to select an appropriate one (**Aim 1, 2, 3**). Matches selected by Mantle's users will be used to update machine learning algorithms to improve future guesses (**Aim 2**), and applied to datasets as metadata (**Aim 1**).

Our ontology-matching engine will be developed in collaboration with CEDAR, a BD2K funded project and a leader in ontology applications. We will work with them to interface with their APIs

for searching cloud-hosted libraries of ontologies (37), assist with the curation and creation of biosurveillance-related ontologies in industry-standard formats, and call upon their experience in ontology searching and alignment (39) for the development and training of NLP feature extraction, information retrieval, and machine learning classifiers (**Aim 1, 2**).

Ontology searches will draw on a sequence of techniques. The corpus of curated ontology terms and their synonyms will be used as the basis of textual feature extraction using regular expressions and other text-matching methods. These features will be combined with features of ontologies, such as view counts (39), as predictors for an ensemble of machine learning classifiers, trained on previously-uploaded datasets and optimized using Mechanical Turk or other crowdsourcing methods (**Aim 1, 2**). These techniques will select the most likely ontological match for each entity in a dataset—specifically, data types and the values of nominal variables. Users will be able to easily view and select alternate matches, and these corrections will be used to update classifier training (**Aim 1, 2, 3**).

**Overcoming barriers in in scientific and institutional culture:** To achieve greater success, Mantle’s goal is to provide an environment for data management that researchers *want* to use. The merit of Mantle is that it will enable users to assign metadata and ontologies to their data seamlessly and enable them to combine with other publicly available data. This will save Mantle’s users significant amounts of time by not having to learn how to clean and structure data to combine with other disparate data.

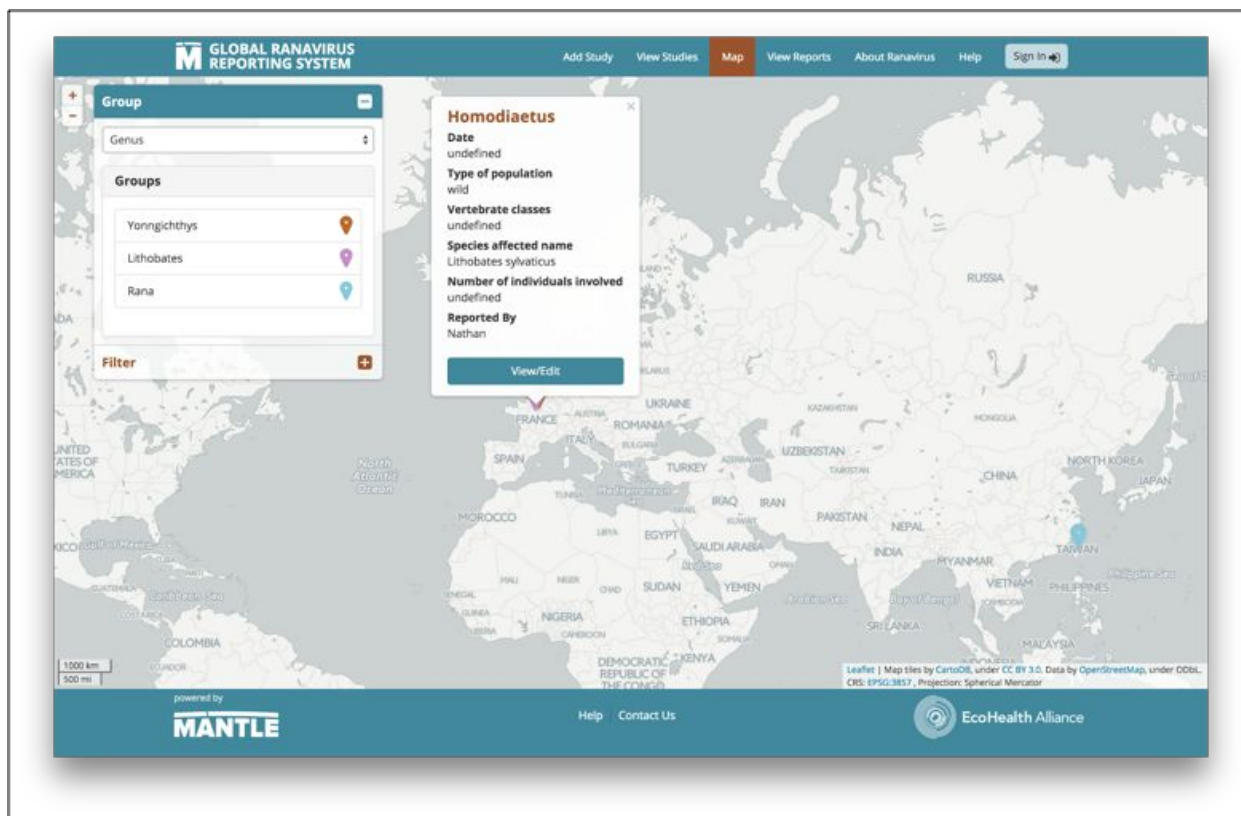
**Security and Privacy:** While outside the scope of this proposal, we understand that security is an integral part of systems like Mantle. Systems should enable the core security features of role-based access, passwords, and audit trails.

**Data integration:** As described above, datasets often suffer from various types of heterogeneity. For example, dataset heterogeneity, especially structural and semantic heterogeneity, differences in file structure, and how concepts are referenced. These are the primary barriers to better data sharing practices across science. Overcoming these barriers can facilitate and incite multidisciplinary research, which is becoming an increasingly important part of the study of human infectious disease, as human health is inextricably bound to ecological, environmental, and animal health factors (40-41). By facilitating the integration of heterogeneous, multidisciplinary data (e.g., infectious disease observations and samples collected in the field, laboratory tests, sociodemographic and economic data, environmental data, data on human and animal population and movement, various other spatiotemporal data types), Mantle will contribute to more effective public health research and disease forecasting.

Focusing on the field of biosurveillance and infectious disease research, we will explore various methods for applying metadata to datasets, in collaboration with leaders in the field of data integration and biosurveillance. Our project will explore cutting-edge methods to facilitate scientists applying metadata annotations, based on a curated selection of structured, web-standard ontologies, to datasets. Technologies for data integration by consistently applying metadata about ontologies will not solve the problem of data heterogeneity unless scientists use it. We are thus developing Mantle, a platform to facilitate the storage, sharing, and analysis of

data, with the capability to apply ontological metadata as part of the capture and upload processes. Mantle will, at each step of the way, use a combination of heuristics and machine-learning algorithms to present users with its best guesses for data type and metadata fields, allow users to correct its guesses, and use corrected data to update its guesses for future computations.

**Preliminary data & prototype development:** Last year, EcoHealth Alliance was contracted by the U.S. Forest Service to build an infectious disease surveillance system for amphibian diseases. The resultant project was a focused prototype of Mantle that implemented a basic subset the full project's planned functionality, allowing users to upload tabular datasets of amphibian disease, browse similar datasets, and export combined datasets in CSV format. When not logged in, visitors to the site can browse a list of publicly-shared datasets, view them on a map, and export records from merged datasets. When users sign up, they gain the ability to upload datasets to the system. The Mantle prototype only accepts data in one format, a CSV file structured around an ontology created specifically for the project. Users must provide metadata for datasets, including contact information for data owners and sharing preferences (public, private, or public with certain information hidden). This process is made easier by the user filling out a profile, which provides default values for a new dataset's ownership metadata. Logged-in users can browse their own private data, as well as publicly-shared data, and export a merged CSV file combining their datasets with other users' shared data. The prototype does not include a flexible upload or export engine, flexible ontologies for datasets, or machine-learning algorithms for intelligent metadata assignment.



**Figure 2.** A screenshot of a prototype of Mantle’s graphic user interface and web portal. This view shows the amphibian disease biosurveillance system, a “proof of concept” which allows users to upload disease reports in CSV format, merge tabular datasets using a specific, tailored ontology, and query data spatially (developed for the U.S. Forest Service).

#### D. TIMELINE & MANAGEMENT PLAN

EcoHealth Alliance will be the primary organization leading this project with Dr. Andrew Huff as the project lead. EcoHealth Alliance leads cutting-edge research into the critical connections between human and wildlife health and delicate ecosystems. Project staff includes software developers, data scientists, public health scientists, clinicians, diagnostic laboratory personnel, veterinarians, information and cyber security experts, and administrative support (Figure 3). Our management plan blends strong scientific expertise in global emerging infectious disease (EID) surveillance, using Agile software methodologies for rapid application development. The project will be managed by our team of data scientists and software developers at EcoHealth Alliance, in consultation with thought leaders in the field of biosurveillance (EHA, ISDS), infused with innovative technologies DIT Inc., developers of leading edge, high quality software, and secured by in the information and cyber security experts at Clango Inc (Figure 4).

	Mantle Timeline	Base Year				Option 1				Option 2			
		Fall	Winter	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter	Spring	Summer
Planning	Conduct Mantle kickoff meeting	x	x										
	Daily scrums and planning sessions	x	x	x	x	x	x	x	x	x	x	x	x
	Construct data collection plan	x	x										
	Conduct user advisory group meeting	x		x		x		x		x		x	
	Construct web hosting development plan	x	x										
	Construct research & development plan	x	x										
	Information & cyber security plan	x	x	x									
	Establish data & information management plan	x	x	x									
	Establish marketing and communication strategy	x	x	x									
Implementation	Develop application	x	x	x	x	x	x	x	x	x	x		
	Software testing			x	x	x	x	x	x	x	x	x	
	Field test application			x	x			x	x	x	x		
	Project management meetings	x		x		x		x		x		x	
	Software refinement and stress testing					x	x	x	x	x	x	x	x
	Host a Mantle launch event											x	
	Mantle training and workshops				x		x		x		x		x
	User focus groups and feedback			x	x	x	x	x	x	x	x	x	x
	Marketing and outreach			x	x	x	x	x	x	x	x	x	x
Evaluation and Deliverables	Prepare peer-reviewed journal articles				x			x	x			x	x
	Attend regional, national, and international conferences to present results		x			x	x			x	x		
	Financial reports	x	x	x	x	x	x	x	x	x	x	x	x
	Copyright (no patent required as open source open access application)				x				x				x
	Progress & technical reports		x				x				x		
	Property reports				x				x				x
	Analyze user traffic, user accounts, and data uploads									x	x	x	x
	Annual audit			x				x				x	

**Figure 3.** Planning, implementation, project evaluation, and deliverables for Mantle.

**Clango** will be responsible for the security aspects of Mantle. Clango has 15 years experience in identity and access management, anti-fraud solutions, governance, and advisory services worldwide. Clango has deployed numerous identity and access management (IAM) capabilities like user registration and lifecycle management; adaptive and federated authentication; privileged administration and access governance. Clango has worked in finance, healthcare, higher education, and across federal, state, and local governments. Clango assures that only authorized identities have access to the right data at the right time.

The **International Society for Disease Surveillance (ISDS)**, works to improve population health by advancing the science and practice of surveillance to support timely and effective prevention and response. The International Society for Disease Surveillance (ISDS) is a 501(c) 3 nonprofit organization founded in 2005 and dedicated to the improvement of population health by advancing the science and practice of disease surveillance. ISDS' membership represents professional and academic subject matter experts in the fields of public health

surveillance, clinical practice, health informatics, health policy, and other areas related to national and global health surveillance. ISDS works toward a vision of timely, effective, and coordinated disease prevention and response among a skilled public health workforce through programs that position us at the vanguard of the disease surveillance field.

	EcoHealth Alliance							ISDS	Clango				
AREAS OF EXPERTISE	Andrew Huff	Peter Daszak	Nathan Breit	Brock Arnold	Jonathan Goiley	Noam Ross	Frederico Rosario	Toph Allen	Daniel Sullivan	Laura Streichert	Anun Kothamath	Steven Hawkins	Kamille White
Emerging Infectious Disease Modeling													
Data Management	X	X	X	X		X	X	X		X	X	X	X
Informatics	X	X	X	X		X	X	X		X	X	X	X
Modeling and Simulation	X	X	X	X		X	X	X		X			
Visualization	X	X	X	X	X	X		X		X			
Software Development													
Applications Development	X		X	X	X	X	X	X	X		X	X	X
Functionality & Design	X		X	X	X	X	X	X			X	X	X
Mobile Applications			X	X	X			X			X	X	X
Scientific Development	X	X	X	X	X	X	X	X			X	X	X
Systems Development	X	X	X	X	X	X	X	X			X	X	X
Testing & Automation	X	X	X	X	X	X	X	X			X	X	X
Cyber & Information Security													
Application and Systems Development Security	X		X	X			X	X			X	X	X
Business Continuity and Disaster Recovery Planning	X						X				X	X	X
Cryptography			X								X	X	X
High Availability Systems	X						X	X			X	X	X
Identity and Access Management / Control			X								X	X	X
Laws, Investigation, and Ethics	X										X	X	X
Physical Security	X						X				X	X	X
Security Management Practices	X		X				X				X	X	X
Telecommunications and Networking Security			X				X				X	X	X
One Health													
Bacteriology		X								X			
Biology		X				X							
Biostatistics	X	X				X	X	X		X			
Biosurveillance	X	X		X		X	X	X		X			
Clinical Laboratory Science		X				X				X			
Environmental Health Science	X	X				X	X						
Epidemiology	X	X						X		X			
Field Surveillance		X						X		X			
Food Systems	X												
Health Systems Research	X	X						X		X			
Infectious Disease Ecology	X	X				X	X						
Medicine		X								X			
Parasitology		X								X			
Plant Pathology		X				X							
Policy	X	X								X			
Veterinary Medicine		X											
Virology		X								X			
Zoology / Wildlife		X											
Other													
Cartography	X					X	X						
Data Science	X		X	X		X	X	X			X	X	
Geographic Information Systems	X					X	X	X				X	
Governance		X								X	X	X	X
Implementation Science	X		X	X		X	X	X			X	X	
Legal and Regulatory	X	X								X	X	X	X
Linguistics / Ontology	X		X	X		X		X		X	X	X	X
Spatial Databases	X		X	X	X	X	X	X			X	X	
Statistics	X		X	X		X	X	X					
Social Science													
Anthropology													
Psychology	X												
Sociology								X					

**Figure 4. Areas of expertise and capacity of staff.**

## **PROTECTION OF HUMAN SUBJECTS**

Human data collected in other studies may be uploaded to Mantle. Mantle research may include secondary data analysis subject to IRB exemptions due to de-identified data status based on HIPAA criteria. Data that potentially contains personally identifiable information will be detected using natural language processing software and will not be allowed to be uploaded to Mantle without the proper deidentification.



## REFERENCES CITED

1. M. Lenzerini, Data integration: a theoretical perspective, in *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on the Principles of Database Systems*. 14 (2002).
2. A.P. Sheth, Changing focus on interoperability in information systems: From system, syntax, structure to semantics, in *Interoperating Geographic Information Systems* M. Goodchild, M. Egenhofer, R. Fegeas, C. Kottman, Ed. (Springer, Boston, Massachusetts, 1999) pp. 5–29.
3. I. F. Cruz, H. Xiao, The role of ontologies in data integration. *Engineering intelligent systems for electrical engineering and communications*. 13.4, 245 (2005)
4. A. Cali, D. Calvanese, G. De Giacomo, M. Lenzerini, Data integration under integrity constraints. *Information Systems*. 29(2), 147-163 (2004).
5. J. Polonetsky, O. Tene, Privacy and Big Data: Making Ends Meet, in *Big Data and Privacy: Making Ends Meet - Future of Privacy Forum and Stanford Law School The Center for Internet and Society*. pp. 1-6 (2013).
6. M. Birnhack, Big Data as A New Informational Privacy Paradigm, in *Big Data and Privacy: Making Ends Meet - Future of Privacy Forum and Stanford Law School The Center for Internet and Society*. pp. 7-10 (2013).
7. J. Brookman, G. S. Hans, Surveillance as a De Facto Privacy Harm, in *Big Data and Privacy: Making Ends Meet - Future of Privacy Forum and Stanford Law School The Center for Internet and Society*. pp. 11-14 (2013).
8. K. Braunschweig, J. Eberius, M. Thiele, W. Lehner, The state of open data, in *WWW2012, Lyon, France: ACM*. (2012).
9. Warmerdam, F. (2008). The geospatial data abstraction library. In *Open Source Approaches in Spatial Data Handling* (pp. 87-104). Springer Berlin Heidelberg.
10. M. Uschold, M. Gruninger, Ontologies and semantics for seamless connectivity, in *ACM SIGMod Record*. 33(4), 58-64 (2004).
11. J. S. Madin, S. Bowers, M. P. Schildhauer, M. B. Jones, Advancing ecological research with ontologies. *Trends Ecol Evol (Amst)*. 23(3), 159–68 (2008).
12. J. Madin, S. Bowers, M. Schildhauer, S. Krivov, D. Pennington, *et al.* An ontology for describing and synthesizing ecological observation data. *Ecological Informatics*. 2, 279-296 (2007).
13. N. Schuurman, A. Leszczynski, Ontology-Based Metadata. *Transactions in GIS*. 10(5), 709–26 (2006).
14. T. Arrison, S. Weidman, Ed. *Steps Toward Large-Scale Data Integration in the Sciences: Summary of a Workshop* (National Academies Press, Washington, DC, 2010).



15. P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, *et al.* BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res.* 39(Web Server issue), 541-5 (2011)
16. M.A. Musen, C. A. Bean, K-H. Cheung, M. Dumontier, K. A. Durante, *et al.* The center for expanded data annotation and retrieval. *J Am Med Inform Assoc.* (2015)
17. T. Tudorache, S. Falconer, N. F. Noy, C. Nyulas, T. B. Üstün TB, *et al.* Ontology Development for the Masses: Creating ICD-11 in WebProtégé, in *Knowledge Engineering and Management by the Masses* (Springer, Berlin Heidelberg, 2010) pp. 74–89.
18. K. S. Candan, H. Liu, R. Suvarna, Resource description framework: metadata and its applications. *ACM SIGKDD Explorations Newsletter.* 3(1), 6-9 (2001).
19. M. J. O'Connor, C. Halaschek-Wiener, M. A. Musen, Mapping Master: A Flexible Approach for Mapping Spreadsheets to OWL, in *The Semantic Web – ISWC 2010* P. F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, *et al.* Ed. (Springer, Berlin Heidelberg, 2010) pp. 194–208.
20. G. Antoniou, E. Franconi, F. V. Harmelen, Introduction to Semantic Web Ontology Languages, in *Reasoning web* N. Eisinger, J. Maluszynski, Ed. (Springer, Berlin Heidelberg, 2005).
21. B. Notay, Improved Metadata for tracking Research Output across the fields. *Euroscientist.* (2013)
22. P. Diviacco, P. Fox, C. Pshenichy, A. Leadbetter, *Collaborative Knowledge in Scientific Research Networks* (IGI Global, Hershey, Pennsylvania, 2015).
23. C. Tenopir, S. Allard, K. Douglass, A. U. Aydinoglu, L. Wu, *et al.* Data sharing by scientists: practices and perceptions. *PLoS ONE.* 6(6) (2011).
24. P. N. Edwards, M. S. Mayernik, A. L. Batcheller, G. C. Bowker, C. L. Borgman, Science friction: data, metadata, and collaboration. *Social Studies of Science.* 41(5), 667-90 (2011).
25. J. Greenberg, Theoretical Considerations of Lifecycle Modeling: An Analysis of the Dryad Repository Demonstrating Automatic Metadata Propagation, Inheritance, and Value System Adoption. *Cataloging & Classification Quarterly. Metadata and Open Access Repositories.* 47(3-4), 380-402 (2009).
26. E. H. Fegraus, S. Andelman, M. B. Jones, M. Schildhauer, Maximizing the value of ecological data with structured metadata: an introduction to Ecological Metadata Language (EML) and principles for metadata creation. *Bull. Ecol. Soc. Am.* 86(3), 158–168 (2005).

27. W. Michener, D. Vieglais, T. Vision, J. Kunze, P. Cruse, G. Janee, DataOne: Data Observation Network for Earth - Preserving Data and Enabling Innovation in the Biological and Environmental Sciences. *D-Lib Magazine*. 17(1/2), (2011).
28. C. Tankard, Big Data Security. *Network Security*. 2012(7), 5-8 (2012).
29. B. Brown, M. Chui, J. Manyika, Are you ready for the era of “big data”? *McKinsley Quarterly*. 4, 24-35 (2011).
30. J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, *et al.* The Mobile Data Challenge: Big Data for Mobile Computing Research. *Pervasive Computing*. (2012).
31. S. Rosset, E. Aharoni, H. Neuvirth, Novel Statistical Tools for Management of Public Databases Facilitate Community-Wide Replicability and Control of False Discovery. *Genetic epidemiology*. 38.5, 477-481 (2014).
32. L. Seligman, A. Rosenthal, A metadata resource to promote data integration, in *Proceedings of the IEEE Metadata Conference, Silver Spring*. (1996).
33. A. Rosenthal, E. Sciore, S. Renner, Toward unified metadata for the department of defense, in *IEEE Metadata Workshop*. (1997).
34. Semantic web best practices and deployment working group, *W3 Technology and Society domain: Semantic Web Activity*.  
<http://www.w3.org/2001/sw/BestPractices/#FreqLinks1>
35. A. Le Hors, A. Bertails, A. Sambra, A. Seaborne, A. Loizou, *et al.* Ed. Linked Data Platform 1.0 W3C Recommendation 26 February 2015. <http://www.w3.org/TR/ldp/>
36. LargeTripleStores, W3C. <http://www.w3.org/wiki/LargeTripleStores>
37. E. Prud'hommeaux, A. Seaborne, SPARQL Query Language for RDF W3C Recommendation 15 January 2008. <http://www.w3.org/TR/rdf-sparql-query/>
38. DataCite Metadata, *Dryad*. [http://wiki.datadryad.org/Old:DataCite\\_Metadata](http://wiki.datadryad.org/Old:DataCite_Metadata)
39. N. F. Noy, P. R. Alexander, R. Harpaz, P. L. Whetzel, R. W. Ferguson, *et al.* Getting Lucky in Ontology Search: A Data-Driven Evaluation Framework for Ontology Ranking, in *The Semantic Web – ISWC 2013* H. Alani, L. Kagal, A. Fokoue, P. Groth, C. Biemann, *et al.* Ed. (Springer, Berlin Heidelberg 2013) pp. 444–59.
40. Daszak, P., Cunningham, A. A., & Hyatt, A. D. (2000). Emerging infectious diseases of wildlife--threats to biodiversity and human health. *science*, 287(5452), 443-449.
41. Karesh, W. B., Cook, R. A., Bennett, E. L., & Newcomb, J. (2005). Wildlife trade and global disease emergence. *Emerg Infect Dis*, 11(7), 1000-1002.

## **CONSORTIUM AND CONTRACTUAL ARRANGEMENTS**

This project is led by EcoHealth Alliance, New York (Dr. Huff, PI), which will subcontract funds to two organizations, The Clango Group (Steven J. Hawkins) and the International Society for Disease Surveillance (ISDS; Dr. Laura Streichert).

The applicant organization (EcoHealth Alliance) is justified in taking the lead on this project because it specializes in technological solutions for disease surveillance and data management for research. The subcontractees will work on specific issues and areas in which they have proven expertise. These areas are: data security and privacy (Clango Group) and collaborative One Health disease surveillance methods (ISDS).

Dr. Huff has experience in managing partnering organizations in collaborative research technology projects including a partnership with DIT (Clango's parent company) on a 2 year Defense Threat Reduction Agency biosurveillance research grant.



MARK A. MUSEN, MD, PhD

*Professor of Medicine (Biomedical Informatics)  
Chief, Biomedical Informatics Research  
(650) 725-3390 Fax (650) 725-7944  
email: musen@stanford.edu*

September 30, 2015

Dr. Andrew Huff  
Associate Vice President  
EcoHealth Alliance  
460 W 34th St. 17th Floor  
New York, NY 10001

Dear Andrew,

All of us at the Center for Expanded Data Annotation and Retrieval (CEDAR) are eager to collaborate with EcoHealth Alliance and its developers in their application for the U01 entitled "Big Data to Knowledge (BD2K) Development of Software Tools and Methods for Biomedical Big Data in Targeted Areas of High Need." We look forward to assisting you in the development of metadata annotation software to be developed by EcoHealth Alliance to further health research.

With years of experience in health informatics, and metadata and ontology management, CEDAR's researchers are uniquely positioned to provide guidance on metadata software development. Our goal to create a unified framework that researchers can use to create consistent, easily searchable metadata aligns well with EcoHealth Alliance's objectives to use connected ontologies and metadata for biosurveillance and health research. We are looking forward to advising EcoHealth to help build a unified framework that researchers in biosurveillance and other biomedical disciplines can use to combine disparate data.

Sincerely,

Mark A. Musen, M.D., Ph.D.  
Professor of Medicine (Biomedical Informatics)  
Principal Investigator,  
Center for Expanded Data Annotation and Retrieval

Medical School Office Building • 1265 Welch Road • Room X271 • Stanford CA 94305-5479



September 30, 2015

Dr. Andrew Huff  
Associate Vice President  
EcoHealth Alliance  
460 W 34th St. 17th Floor  
New York, NY 10001

Dear Dr. Huff,

On behalf of the International Society for Disease Surveillance (ISDS), I am pleased to express our strong interest in collaborating with you on the NIH funded U01 project entitled "Big Data to Knowledge (BD2K) Development of Software Tools and Methods for Biomedical Big Data in Targeted Areas of High Need". The Mantle software platform offers tremendous promise for biosurveillance and disease forecasting abilities. We believe you have built an excellent team with Dr. Peter Daszak and other EcoHealth Alliance staff to further develop the Mantle platform.

Mantle aligns with ISDS' dedication to the improvement of population health by advancing the science and practice of disease surveillance. To assist with this project, we will provide expertise in the field of One Health biosurveillance. ISDS' membership represents professional and academic subject matter experts in the fields of public health surveillance, clinical practice, health informatics, health policy, and other areas related to national and global health surveillance. ISDS works toward a vision of timely, effective, and coordinated disease prevention and response among a skilled public health workforce through programs that position us at the vanguard of the disease surveillance field.

We will contribute valuable insight into the development of Mantle's biosurveillance tools, models, and disease forecasting capabilities. Our involvement in the biosurveillance community will guide the development of Mantle to maximize usability and meet current and future scientific demand. We believe that by engaging end-users in the development process from the outset, there is a much greater chance that the software platform is practical and has real utility to the public health surveillance professional.

ISDS will provide all necessary consultation and support to Dr. Huff and his team to accomplish the development of Mantle to further our common missions.

Sincerely,

A handwritten signature in blue ink, reading "Laura C. Streichert".

Laura C. Streichert, PhD, MPH  
Executive Director, ISDS  
617-779-7227; [lstreichert@syndromic.org](mailto:lstreichert@syndromic.org)

---

**International Society for Disease Surveillance**

26 Lincoln St., Suite # 3, Brighton, MA 02135 (617)  
779-0880 [www.syndromic.org](http://www.syndromic.org)



7701 France Ave S, Suite 400Edina, MN 55435  
1-877-817-4712

Dr. Andrew Huff  
Associate Vice President  
EcoHealth Alliance  
460 W 34th St. 17th Floor  
New York, NY 10001

To Whom It May Concern,

On behalf of Clango, I would like to express my full support for Dr. Andrew Huff and EcoHealth Alliance (EHA) in their application for the U01 entitled " Big Data to Knowledge (BD2K) Development of Software Tools and Methods for Biomedical Big Data in Targeted Areas of High Need". Clango is interested in working with EHA to help provide secure Big Data solutions in order to further human, animal, and environmental health research.

With 15 years experience in identity and access management, anti-fraud solutions, governance, and advisory services, Clango will contribute to the development of Mantle's data security and privacy features. Clango has deployed numerous identity and access management (IAM) capabilities like user registration and lifecycle management; adaptive and federated authentication; privileged administration and access governance, and has supported global organizations in finance, healthcare, higher education, and across federal, state, and local governments. Clango uses the IAM framework to assure that authorized identities have access to the right data at the right time.

We at Clango are looking forward to our collaboration with the EcoHealth Alliance team and working further on this exciting project.

Sincerely

Duane Graham  
CEO

## RESOURCE SHARING PLAN

Mantle will be distributed under the Apache 2.0 open-source license. The software will be open source, and data will be open access to Mantle users, with all code hosted on repositories on GitHub. Specifically, Mantle will be made freely available to biomedical researchers, and educators in the nonprofit sector within 1 year of project completion. Dissemination of enhanced or customized versions of the software will be ongoing, beginning 1 month after project completion. Components of Mantle with discrete functionality will be hosted as separate software packages so that other projects can benefit from their advances. For example, the part of the software that matches datasets with ontologies will be developed as a package with API end-points so that it can be used in other software projects, potentially with customizations. Eventually, organizations will also be able to host their own server running an Mantle instance to manage their datasets. These instances will be able to interface with other instances through APIs. The EcoHealth Alliance-hosted instance of Mantle will be made publicly available. EcoHealth Alliance will maintain the original and any EcoHealth created versions of Mantle, but other users may modify source code and share modifications on GitHub.