Bree Norlander
6/3/2016
Implementation Plan

Proposal: Data Portal for Emerging Disease Outbreaks

Mission

Provide easily accessible curated collections of high-quality datasets, via an internet portal, that are useful to researchers and other interested parties who are tracking emerging disease outbreaks.

1. Introduction

Zika is an emerging disease of recent concern to researchers globally. The Zika virus is most notably spread to humans by *Aedes aegypti* or *Aedes albopictus* mosquitoes. Zika is also spread through mosquito communities by humans carrying the disease (Centers for Disease Control and Prevention at al., 2016). If researchers have data on local populations of *Aedes aegypti* or *Aedes albopictus* mosquitoes, they can begin to predict whether the area is at risk of the disease spreading. Additional data on local climate, water quality, population density, and occurrences of the disease or related mosquito-borne diseases, can greatly enhance risk assessments.

2. Value

In an effort to engage researchers and data scientists in the study of emerging disease outbreaks, I propose to create a web portal of curated collections of publicly available datasets that have the potential to advance the current research in these areas. There is a dire need for

a one-stop-shop for datasets related to emerging disease outbreaks such as Zika, Chikungunya, and Ebola. This need is so great that Buzzfeed journalist, Jeremy Singer-Vine, took the matter into his own hands by curating datasets related to Zika on a GitHub[1] repository (Kissane and Singer-Vine, 2016). Another group of researchers created a platform for "collaborative science in outbreak situations" specifically for Zika called "ZikaNet" (Borderia et al., 2016). The purpose of the platform is to allow researchers complete control over their own data while and providing an easy way to share the data with other researchers. "This system is designed to reduce barriers to collaborative research typically encountered in highly competitive fields through the facilitation of data sharing, provision of transparency and attribution in research findings and the advancement of science" (Borderia et al., 2016).

I propose a web portal that encompasses all emerging disease outbreaks, that is discoverable by researchers and data scientists, and most importantly, finds and curates collections of datasets that are already publicly available online. I have begun the project by creating a collection of datasets about mosquito populations in the Western United States (see Appendix). Creating this collection has led me to believe that there are vast resources of untapped, long-tail data (see Plale, 2012), that could be used to advance important research on mosquito-borne diseases. A significant contribution of my proposed web portal is to find this untapped data and make it discoverable by packaging it in appropriate collections, adding the metadata (using Dublin Core[2] vocabulary) outlined in the example data collection (see Appendix), and

[1] https://github.com/BuzzFeedNews/zika-data

[2] Dublin Core was chosen because it is widely used and "domain-agnostic" (Digital Curation Centre, n.d.)

linking to the data. I propose working with data producers whose data is at high risk of loss to deposit the data in a repository for preservation. The goal of the web portal is not to create a repository, but to be a discovery tool.

Borderia et al. (2016) state that "reducing the time to action in outbreak situations would be best facilitated through a globally coordinated effort for joint analysis on shared data." My proposed resource would facilitate the analysis and sharing of data and add considerable value to the datasets and the research area. I propose to begin the project with a focus on the United States for two reasons: 1) a smaller scope will allow the team to focus on quality and comprehensiveness and 2) fewer employees (and thus fewer financial resources) will be required to begin the program. However, as the resource becomes a relied upon source for emergent data, I would like to see it expand globally.

3. Work Plan

The first step in building the portal is to create a team and a timeline. The team will consist of one full-time project manager (myself), three half-time student employees, and the equivalent of one full-time intern (3 to 6-month employment stints).

3.1 Project Manager

As project manager, I will oversee the creation of the website, contract with an internet service provider, and purchase a domain, as well as the necessary hardware and software. The design of the website will be contracted out to a design firm. The website must be professional, easy to navigate, and allow for easy entry of metadata and links.

3.2 Student Employees

The second step is to hire the student employees who will begin to create dataset collections such as the collection outlined in the Appendix. The student employees will be responsible for finding datasets and curating meaningful collections that meet the mission of the project. Each employee will contact data producers to request permission to be included in the collection, and to open up a discussion about data deposit if the data is not currently available in a repository. Complex cases may be routed to the project manager. Collections will be reviewed by the project manager prior to publication on the website.

3.3 Interns

The website should be active and populated with several collections within one year. Interns will be hired to assist with public relations and grant writing. It will be necessary to announce the availability of the resource to the research and data science communities.

4. Implementation Challenges

This project faces several challenges. The first challenge is to determine the appropriate level of permission required for a dataset to be included in the portal. It is imperative that data producers receive credit for their work and no resource will be included without the permission of and attribution of the data producer.

The second challenge is to encourage data producers of high loss-risk data to deposit data into a repository and to determine the appropriate repository. Our team will need to be aware of

the costs associated with deposit and the available repositories. We will also need to distinguish trustworthy repositories from those that may themselves be at risk of loss.

The third challenge is announcing the project to the research community and building trust in our product. The project will need to be announced in established publications, conferences, and through appropriate social media channels.

5. Limitations

The main limitation to this project is that it is not a data repository. As such, we will be working to ensure the preservation of the data we include in our collections, but will not be able to guarantee long-term access.

6. Conclusion

In summary, the Data Portal for Emerging Disease Outbreaks (DPEDO), meets a demand that has yet to be filled. Researchers and data scientists are eager to address disease outbreaks to slow their spread and find solutions. Sourcing and providing quality datasets in clear and concise collections on a single data portal will allow these scientists easy access to information with which they may advance current research and risk assessments.

References

Borderia, A., Fontes, M., MacPherson, C., Thomas, S., Mangravite, L., & Omberg, L. (2016, May 24). Collaborative Data Sharing in Outbreak Situations, a possible solution. Retrieved June 02, 2016, from http://blog.f1000research.com/2016/05/24/collaborative-data-sharing-in-outbreak-situations-a-possible-solution/

Centers for Disease Control and Prevention, National Center for Emerging and Zoonotic Infectious Diseases, & Division of Vector-Borne Diseases. (2016, April 26). Zika Virus. Retrieved June 03, 2016, from http://www.cdc.gov/zika/vector/range.html

Digital curation Centre. (n.d.). Dublin Core. Retrieved June 03, 2016, from http://www.dcc.ac.uk/resources/metadata-standards/dublin-core

Kissane, E., & Singer-Vine, J. (2016, February 4). An Open Guide to Zika Data. Retrieved June 02, 2016, from https://source.opennews.org/en-US/articles/open-guide-zika-data/

Plale, B. (2012). Managing the long tail of science: Data and communities. *ACM International Conference Proceeding Series,* ACM International Conference Proceeding Series, 2012.

Appendix – Example Data Collection

Introduction

Mosquitoes are known vectors for pathogenic infections. In the study of epidemiology, "a vector is any agent (person, animal, or microorganism) that carries and transmits an infectious pathogen into another living organism" (Wikipedia, 2016). For example, two species of mosquitoes, *Aedes aegypti* and *Aedes albopictus,* "are the main vectors transmitting dengue and chikungunya viruses" (Kraemer et al., 2015). *Aedes aegypti* mosquitoes are also a vector for the zika virus (Centers for Disease Control, 2016).

According to UCAR Center for Science Education:

> ...climate plays an important role in the seasonal pattern or temporal distribution of diseases that are carried and transmitted through vectors because the vector animals often thrive in particular climate conditions. For example, warm and wet environments are excellent places for mosquitoes to breed. If those breeding mosquitoes happen to be a species that can transmit disease and if there is an infected population in the region, then the disease is more likely to spread in that area. Because they are sensitive to climate, the distribution and number of vectors is also affected by climate change. According to the IPCC Fourth Assessment Report, climate change has already altered the distribution of some disease vectors. There is evidence that the geographic range of ticks and mosquitoes that carry disease has changed in response to climate change. Ticks have extended their range north in Sweden and Canada and into higher altitudes in the Czech Republic. While future climate change is expected to continue to alter the distribution of disease vectors, it is important to recognize that there are several other factors (such as changes in land use, population density, and human behavior) that can also change the distribution of disease vectors as well as the extent of infection. (The University Corporation for Atmospheric Research, n.d.)

In an effort to aid researchers in the study of mosquito-borne diseases, I have compiled a collection of datasets that contain information about current and past populations of mosquitoes in certain areas of the Western United States. These datasets can be used alone or in conjunction with data related to climate conditions, land use, and population density to assess the current conditions of mosquito populations and to make predictions about the future of these populations.

Collection Mission

Provide high-quality datasets that are useful to researchers and other interested parties who are tracking mosquito prevalence and trends in the Western United States. These datasets can be used in conjunction with, for example, climate data, water-quality data, or population-density data, to make predictions and inferences about mosquito populations and mosquito-borne disease prevalence in certain areas.

Selection Criteria

In order to craft my selection criteria, I drew upon the Data Preservation Alliance for the Social Sciences (DataPASS) appraisal guidelines (2005) and Whyte's and Wilson's "A Digital Curation Centre and Australian National Data Service 'working level' guide" (2010).

1. Relevance to Collection Mission – Does the dataset meet the mission of the collection?
2. Scientific/Analytic Value and Potential – Does the dataset offer value to scientist and data analysts now and in the future?
3. Uniqueness – Are the data available elsewhere?
4. Usability – Are the data in a format that is readily accessible to scientists and data analysts?
5. Metadata Availability – Is there sufficient metadata (see metadata criteria)?

The following is not a criterion for inclusion, but a rating that can be used to determine preservation priority. The ratings are subjective, but determined based on the following

guidelines:

1. Preservation State:
    a. High – deposited in a data repository
    b. Medium – not deposited in a repository, but the stability of the server (based on main URL) and the state of the data suggest that the data owner intends to preserve/maintain the data
    c. Low – not deposited in a repository and no suggestion that data owner intends to preserve or maintain the data

Metadata Criteria

In order to craft my metadata criteria, I drew upon Rumsey and Jefferies's "Challenges in Building an Institutional Research Data Catalogue" (2013) and "DataFinder: A Research Data Catalogue for Oxford" (2013). I also drew upon ICPSR's "Guide to Social Science Data Preparation and Archiving Phase 3: Data Collection and File Creation" (n.d.).

The following metadata are **required** for any dataset included in my collection:

1. Data Collector/Producer – this may include individuals or organizations
2. Description – this information may be harvested from data sources or provided by data collector
3. Variable Information Including: Dates (temporal coverage), Units of Measurement, Geographic Coverage
4. File Format/s

The following metadata are **strongly desired** for inclusion in my collection:

1. Funding Sources – name of funding agency and any related metadata such as grant information
2. Data Owner/Publisher – name of individual or organization that holds the copyright to the data

3. Data Collector/Producer Affiliation
4. Data Identifier
5. Publication Year
6. Data Collection Instruments
7. Description of Non-Standard Abbreviations
8. Terms and Conditions
9. Provenance

The Data

The datasets that I describe in this assignment fall into two main categories: 1. Static Data and 2. Dynamic Database. Within each category, data can be further categorized based on preservation rating (high, medium, low – see Selection Criteria).

Category: Static Data

1. http://hdl.handle.net/1773/33850

This is the accompanying data to a dissertation by Julia Weicheld: Weicheld, J., & Treser, Charles D. (2015). *Impact of environmental factors on mosquito population abundance and distribution in King County, Washington. Seattle*: University of Washington.

Selection Criteria	
Relevance to Collection Mission	This data fit well within the mission of this collection.
Scientific/Analytic Value and Potential	This dataset can be used alone to reproduce original data analysis or to use new analytical tools, or can be paired with additional data about the area learn more about the Seattle-area mosquito population and its potential for change.
Uniqueness	I have found no other open-access data resource with this much detail about mosquito populations in the Seattle area.
Usability	The data are available in XLSX format which is easily accessible with various data analysis software. Descriptions of the data are provided in DOCX, PDF, and XLSX formats. The accompanying dissertation describes the data gathering methods and data analysis in great detail.
Metadata Availability	This dataset is metadata-rich. All required metadata elements are provided, and many of the desired elements are provided.
Preservation State	High
Metadata Criteria	
Data Collector/Producer	Weicheld, Julia
Description	Background: Climate, land cover, and other environmental factors have been shown to have a direct impact on the epidemiology of vector-borne diseases. Warming temperatures combined with other effects of climate change

	and changes in land use have the potential to amplify vector mosquito populations and transmission of arboviruses in King County, Washington. This research aims to provide insight into vector populations that may govern vector-borne disease transmission in King County. Methods: Mosquitoes were trapped at selected areas in King County in summer 2014. Additional mosquito data for King County were gathered and assessed for quality and completeness. Identical sites sampled in 2003 and 2014 were directly compared to determine any changes in mosquito abundance and diversity over an 11-year period. Temperature, precipitation, and land cover data were obtained and investigated for their influence on mosquito abundance using correlative and regression analyses. Results: The correlative analysis found mosquito abundance was significantly positively associated with percent med-high developed land cover, maximum temperature, and minimum temperature variables. Mosquito abundance metrics were found to be negatively correlated with percent forested land cover and average weekly precipitation. Mosquito abundance was significantly higher in 2003 than in 2014, but was unexplained by changes in land cover or climate. Conclusions: Mosquito populations appear to be impacted by the climate and land cover variables studied, but other factors not examined in this study may have greater impacts.
Variables:	
Temporal Coverage	2003 and 2014
Units of Measurement	Mosquito Count
Geographic Coverage	Specified Locations in King County, WA
Other	
File Format/s	PDF, DOCX, XLSX
The following metadata elements are highly desired, but not required for inclusion in the collection:	
Funding Source/s	N/A
Data Owner/Publisher	University of Washington
Data Collector/Producer Affiliation	University of Washington
Data Identifier	http://hdl.handle.net/1773/33850
Publication Year	2015
Data Collection Instruments	Encephalitis Virus Surveillance CO_2 light traps
Description of Non-Standard Abbreviations	N/A
Terms and Conditions for Use	Copyright is held by the individual authors.

Provenance	Weicheld, Julia

Category: Static Data

2. http://datadryad.org/resource/doi:10.5061/dryad.47v3c

Data Citation: Kraemer MUG, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, Messina JP, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Wint GRW, Elyazar IRF, Teng H, Hay SI (2015) Data from: The global compendium of Aedes aegypti and Ae. albopictus occurrence. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.47v3c

Original Publication Citation: Kraemer MUG, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, Messina JP, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Wint GRW, Elyazar IRF, Teng H, Hay SI (2015) The global compendium of Aedes aegypti and Ae. albopictus occurrence. Scientific Data 2(7): 150035. http://dx.doi.org/10.1038/sdata.2015.35

Selection Criteria	
Relevance to Collection Mission	This dataset contains information on occurrences of two species of mosquito worldwide. Some of the occurrences are within the Western United States. Geographic information is provided with each occurrence and thus this data fit within the mission of the collection.
Scientific/Analytic Value and Potential	This dataset can be used alone for inference or prediction (especially regarding temporal trends), or can be paired with additional data from the specific geographical areas for inference and or prediction.
Uniqueness	This comprehensive dataset of known occurrences of Ae. aegypti and Ae. albopictus between 1960 and 2014 appears to be one-of-a-kind, though it was compiled from available data sources. It is the compilation of the data that makes it a unique resource.
Usability	The data are available in CSV format which is easily accessible with various data analysis software. Descriptions of the variables are provided in PDF format. The accompanying publication describes the data gathering methods and data analysis in great detail.
Metadata Availability	This dataset is metadata-rich. All required metadata elements are provided, and many of

	the desired elements are provided (though provenance information is mysteriously absent).
Preservation State	High
Metadata Criteria	
Data Collector/Producer	Kraemer, Moritz U. G. Sinka, Marianne E. Duda, Kirsten A. Mylne, Adrian Shearer, Freya M. Brady, Oliver J. Messina, Janey P. Barker, Christopher M. Moore, Chester G. Carvalho, Roberta G. Coelho, Giovanini E. Van Bortel, Wim Hendrickx, Guy Schaffner, Francis Wint, G. R. William Elyazar, Iqbal R. F. Teng, Hwa-Jen Hay, Simon I.
Description	PLEASE NOTE, THESE DATA ARE ALSO REFERRED TO IN ANOTHER PUBLICATION. PLEASE SEE http://dx.doi.org/10.7554/eLife.08347. Aedes aegypti and Ae. albopictus are the main vectors transmitting dengue and chikungunya viruses. Despite being pathogens of global public health importance, knowledge of their vectors' global distribution remains patchy and sparse. A global geographic database of known occurrences of Ae. aegypti and Ae. albopictus between 1960 and 2014 was compiled. Herein we present the database, which comprises occurrence data linked to point or polygon locations, derived from peer-reviewed literature and unpublished studies including national entomological surveys and expert networks. We describe all data collection processes, as well as geo-positioning methods, database management and quality-control procedures. This is the first comprehensive global database of Ae. aegypti and Ae. albopictus occurrence, consisting of 19,930 and 22,137 geo-

	positioned occurrence records respectively. Both datasets can be used for a variety of mapping and spatial analyses of the vectors and, by inference, the diseases they transmit.
Variables:	
Temporal Coverage	1960 to 2014
Units of Measurement	Each occurrence constitutes a row in the CSV
Geographic Coverage	Global
Other	
File Format/s	PDF, CSV
The following metadata elements are highly desired, but not required for inclusion in the collection:	
Funding Source/s	M.U.G.K. is funded by the German Academic Exchange Service (DAAD) through a graduate scholarship. M.E.S. is funded by a project grant from the Bill and Melinda Gates Foundation via the VecNet consortium (http://vecnet.org). O.J.B. is funded by a BBSRC studentship. J.P.M. and G.R.W.W. are funded by the International Research Consortium on Dengue Risk Assessment Management and Surveillance (IDAMS, European Commission 7th Framework Programme (21803), :http://www.idams.eu, Publication #28). I.R.F.E. is funded by the Wellcome Trust (#B9RZGS0). VBORNET is an ECDC funded project (contract number ECDC/09/018). S.I.H. is funded by a Senior Research Fellowship from the Wellcome Trust (#095066) which also supports A.M. and K.A.D. S.I.H. and C.M.D. also acknowledge funding support from the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. FMS is funded by the Rhodes Trust.
Data Owner/Publisher	University of Washington
Data Collector/Producer Affiliation	1. **Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK** ○ Moritz U. G. Kraemer ○ , Kirsten A. Duda ○ , Jane P. Messina ○ , G. R. William Wint ○ & Simon I. Hay 2. **Wellcome Trust Centre for Human Genetics,University of Oxford, Oxford, UK**

	○ Marianne E. Sinka ○ , Adrian Mylne ○ , Freya M. Shearer ○ & Oliver J. Brady **3. Institute for Health Metrics and Evaluation, University of Washington, Seattle, USA** ○ Marianne E. Sinka ○ , Adrian Mylne ○ , Freya M. Shearer ○ & Oliver J. Brady **4. Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California, Davis, CA, USA** ○ Christopher M. Barker **5. Center for Vectorborne Diseases, University of California, Davis, CA, USA** ○ Christopher M. Barker **6. Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA** ○ Christopher M. Barker ○ & Simon I. Hay **7. Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO, USA** ○ Chester G. Moore **8. National Dengue Control Program, Ministry of Health, Brasilia, DF, Brazil** ○ Roberta G. Carvalho ○ & Giovanini E. Coelho **9. European Centre for Disease Prevention and Control, Stockholm, Sweden** ○ Wim Van Bortel **10. Avia-GIS, Zoersel, Belgium** ○ Guy Hendrickx ○ & Francis Schaffner **11. Environmental Research Group Oxford Ltd, Department of Zoology, South Parks Road, Oxford OX1 3PS, UK** ○ G. R. William Wint **12. Eijkman-Oxford Clinical Research Unit, Jakarta, Indonesia** ○ Iqbal R. F. Elyazar **13. Center for Research, Diagnostics and Vaccine Development, Centers for Disease Control, Taipei, Taiwan (ROC)**

	○ Hwa-Jen Teng
Data Identifier	http://dx.doi.org/10.5061/dryad.47v3c
Publication Year	2015
Data Collection Instruments	
Description of Non-Standard Abbreviations	Available in accompanying PDF.
Terms and Conditions for Use	This work is licensed under a Creative Commons Attribution 4.0 International License.
Provenance	Provenance was supposedly available in the dryad repository: "Published literature or unpublished sources with reference ID that corresponds to the full list of references in Data Citation 1: Dryad Digital Repository http://dx.doi.org/10.5061/dryad.47v3c" but I could not locate this information.

Category: Static Data

3. http://parasitesandvectors.biomedcentral.com/articles/10.1186/1756-3305-7-276#MOESM3

Citation: Guerra, C., Reiner, R., Perkins, T., Lindsay, S., Midega, J., Brady, O., Barker, C., Reisen, W., Harrington, L., Takken, W., Kitron, U., Lloyd, A., Hay, S., Scott, T. and Smith, D. (2014). A global assembly of adult female mosquito mark-release-recapture data to inform the control of mosquito-borne pathogens. *Parasites & Vectors*, 7, 276.

Selection Criteria	
Relevance to Collection Mission	This dataset contains information on mosquito bionomic parameters. It differs from previous datasets in that the mosquitoes were released by researchers and then recaptured. The data are not indicators of what species reside in certain locations, but provide information on behavior and survival patterns of the released species. The data are global, but some of the data come from within the Western United States. Geographic information is provided with each occurrence and thus this data fit within the mission of the collection.
Scientific/Analytic Value and Potential	This dataset can be used alone for inference or prediction (especially regarding temporal trends), or can be paired with additional data from the specific geographical areas for inference and or prediction.
Uniqueness	This is a compilation of research data from 1913 to 2010. This is a unique comprehensive resource of mark-release-recapture data.
Usability	The data are available in CSV format which is easily accessible with various data analysis software. Descriptions of the variables are provided in PDF format. The accompanying publication describes the data gathering methods and data analysis in great detail.
Metadata Availability	This dataset is metadata-rich. All required metadata elements are provided, and many of the desired elements are provided.
Preservation State	High
Metadata Criteria	
Data Collector/Producer	• Carlos A Guerra • Robert C Reiner Jr, • T Alex Perkins, • Steve W Lindsay,

	• Janet T Midega, • Oliver J Brady, • Christopher M Barker, • William K Reisen, • Laura C Harrington, • Willem Takken, • Uriel Kitron, • Alun L Lloyd, • Simon I Hay, • Thomas W Scott • David L Smith
Description	**Background** Pathogen transmission by mosquitos is known to be highly sensitive to mosquito bionomic parameters. Mosquito mark-release-recapture (MMRR) experiments are a standard method for estimating such parameters including dispersal, population size and density, survival, blood feeding frequency and blood meal host preferences. **Methods** We assembled a comprehensive database describing adult female MMRR experiments. Bibliographic searches were used to build a digital library of MMRR studies and selected data describing the reported outcomes were extracted. **Results** The resulting database contained 774 unique adult female MMRR experiments involving 58 vector mosquito species from the three main genera of importance to human health: *Aedes*, *Anopheles* and *Culex*. Crude examination of these data revealed patterns associated with geography as well as mosquito genus, consistent with bionomics varying by species-specific life history and ecological context. Recapture success varied considerably and was significantly different amongst genera, with 8, 4 and 1% of adult females recaptured for *Aedes*, *Anopheles* and *Culex* species, respectively. A large proportion of experiments (59%) investigated dispersal and survival and many allowed disaggregation of the release and recapture data. Geographic coverage was limited to just 143 localities around the world.
Variables:	
Temporal Coverage	1913 to 2010

Units of Measurement	Individual mosquitoes
Geographic Coverage	Global
Other	
File Format/s	TIF, PDF, CSV
The following metadata elements are highly desired, but not required for inclusion in the collection:	
Funding Source/s	This work was supported by the Research and Policy for Infectious Disease Dynamics (RAPIDD) program of the Science and Technology Directory, Department of Homeland Security and Fogarty International Center, National Institutes of Health (NIH), including funding for data assembly and workshops that led to the writing of this manuscript. ALL acknowledges support from the following grants: NIH R01AI091980 and NSF DMS 1246991. SIH is funded by a Senior Research Fellowship from the Wellcome Trust (#095066).
Data Owner/Publisher	
Data Collector/Producer Affiliation	Carlos A Guerra **Affiliated with** • Fogarty International Center, National Institutes of Health • Center for Disease Dynamics, Economics & Policy Robert C Reiner Jr **Affiliated with** • Fogarty International Center, National Institutes of Health • Department of Entomology, University of California T Alex Perkins **Affiliated with** • Fogarty International Center, National Institutes of Health • Department of Entomology, University of California Steve W Lindsay **Affiliated with** • Fogarty International Center, National Institutes of Health • Department of Disease Control, London School of Hygiene and Tropical Medicine • School of Biological and Biomedical Sciences, Durham University Janet T Midega **Affiliated with** • Pathogen, Vector and Human Biology Unit, KEMRI–University of Oxford–Wellcome Trust Research Programme • Department of Life Sciences, Imperial College London

	Oliver J Brady **Affiliated with** • Spatial Ecology and Epidemiology Group, Department of Zoology, Tinbergen Building, University of Oxford Christopher M Barker **Affiliated with** • Fogarty International Center, National Institutes of Health • Center for Vectorborne Diseases, Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California William K Reisen **Affiliated with** • Fogarty International Center, National Institutes of Health • Center for Vectorborne Diseases, Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California Laura C Harrington **Affiliated with** • Department of Entomology, Cornell University Willem Takken **Affiliated with** • Laboratory of Entomology, Wageningen University Uriel Kitron **Affiliated with** • Fogarty International Center, National Institutes of Health • Department of Environmental Sciences, Emory University Alun L Lloyd **Affiliated with** • Fogarty International Center, National Institutes of Health • Department of Mathematics and Biomathematics Graduate Program, North Carolina State University Simon I Hay **Affiliated with** • Fogarty International Center, National Institutes of Health • Spatial Ecology and Epidemiology Group, Department of Zoology, Tinbergen Building, University of Oxford Thomas W Scott **Affiliated with** • Fogarty International Center, National Institutes of Health • Department of Entomology, University of California David L Smith **Affiliated with**

	• Fogarty International Center, National Institutes of Health • Center for Disease Dynamics, Economics & Policy • Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University
Data Identifier	**DOI:** 10.1186/1756-3305-7-276
Publication Year	2014
Data Collection Instruments	
Description of Non-Standard Abbreviations	Additional file 1: **List and description of all relevant data fields in tables within Additional file** 3. (PDF 63 KB)
Terms and Conditions for Use	This article is published under license to BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Provenance	This is a compilation of data from several publications. Original sources of data are contained within "Additional file 2: List of references from which the MMRR data were extracted.(PDF 181 KB)"

Category: Static Data

4.

http://www.co.coos.or.us/Departments/CoosHealthWellness/PublicHealth/EnvironmentalHealth/Mosquito-Vector.aspx

Coos County, Oregon Department of Environmental Health

Elements of this dataset:

2015 Operations Report from Vector Disease Control International:

http://www.co.coos.or.us/Portals/0/Public%20Health/Environmental%20Health%20Services/Bandon%202015%20Report.pdf?ver=2016-01-07-132348-813

Mosquito Status Report Bandon Marsh Vicinity – Year End 2015:

http://www.co.coos.or.us/Portals/0/Public%20Health/Environmental%20Health%20Services/Memo%20Mosquito%20report%202015.pdf?ver=2016-01-07-132348-937

Oregon State University Veterinary Diagnostic Laboratory Report:

http://www.co.coos.or.us/Portals/0/Public%20Health/Environmental%20Health%20Services/Vector%20Control/OSUVDL%20results%207-6-15.pdf

2015 Mosquito Trapping Survey:

http://www.co.coos.or.us/Portals/0/Public%20Health/Environmental%20Health%20Services/Vector%20Control/2015%20Mosquito%20Adult%20Trap%20Data.xlsx

2014 Mosquito Trapping Survey:

http://www.co.coos.or.us/Portals/0/Public%20Health/Environmental%20Health%20Services/2014%20Mosquito%20Survey%20Results.xlsx

Adult Mosquito Graphs:

http://www.co.coos.or.us/Portals/0/Public%20Health/Environmental%20Health%20Services/Adult%20Mosquito%20Data.pdf

Selection Criteria	
Relevance to Collection Mission	This data fit well within the mission of this collection.
Scientific/Analytic Value and Potential	This dataset can be used alone to reproduce original data analysis or to use new analytical tools, or can be paired with additional data about

	the area learn more about the Coos County mosquito population and its potential for change.
Uniqueness	This dataset is not archived in a repository. It is at high risk for loss. It is unique not only because of its particular location, but also because it is provided by a County rather than a research institution or data repository.
Usability	The data are available in XLSX format which is easily accessible with various data analysis software.
Metadata Availability	All required metadata elements are provided and many of the desired metadata elements are available by extracting information from the materials.
Preservation State	Low
Metadata Criteria	
Data Collector/Producer	Coos County, OR Department of Environmental Health In partnership with United States Fish and Wildlife Service (USFWS) "Adult Mosquito Data Graph" provided by: Bill Bridgeland, USFWS Wildlife
Description	Adult mosquito traps are periodically set in the vicinity of the Bandon Marsh to measure the presence of the salt water mosquito Aedes dorsalis.
Variables:	
Temporal Coverage	2014 and 2015
Units of Measurement	Mosquito Count by Species
Geographic Coverage	Specified Locations in Coos County, OR
Other	
File Format/s	PDF, XLSX
The following metadata elements are highly desired, but not required for inclusion in the collection:	
Funding Source/s	
Data Owner/Publisher	Coos County, OR
Data Collector/Producer Affiliation	Coos County, OR Department of Environmental Health In partnership with United States Fish and Wildlife Service (USFWS)
Data Identifier	
Publication Year	2014 and 2015
Data Collection Instruments	Larval mosquitoes are collected with standard "dipping"' procedures. The dipper consists of a white plastic cup, 400ml in volume, with a wooden dowel handle. Dipping for mosquito larvae is used as a survey tool simply to

	determine the presence or absence of larvae and involves taking several water samples from designated areas and then counting the larvae captured in each dip.
Description of Non-Standard Abbreviations	
Terms and Conditions for Use	Copyright Coos County 2015

Category: Static Data

5. Oregon Department of Fish and Wildlife

Elements of this dataset:

Oregon Department of Fish and Wildlife 2014 Mosquito Sampling Memo:

http://www.co.coos.or.us/Portals/0/Public%20Health/Environmental%20Health%20Services/Vector%20Control/ODFW%20Mosquito%20Sampling%202014.pdf

ODFW 2014 Mosquito Sampling Data:

http://www.co.coos.or.us/Portals/0/Public%20Health/Environmental%20Health%20Services/Vector%20Control/ODFW%202014%20Mosquito%20Data.xls

Selection Criteria	
Relevance to Collection Mission	This data fit well within the mission of this collection.
Scientific/Analytic Value and Potential	This dataset can be used alone to reproduce original data analysis or to use new analytical tools, or can be paired with additional data about the area learn more about Coquille-area mosquito population and its potential for change.
Uniqueness	This dataset is not archived in a repository. It is at high risk for loss. It is unique not only because of its particular location, but also because it is provided by a local governmental agency rather than a research institution or data repository.
Usability	The data are available in XLS format which is easily accessible with various data analysis software.
Metadata Availability	All required metadata elements are provided and some of the desired metadata elements are available by extracting information from the materials.
Preservation State	Low
Metadata Criteria	
Data Collector/Producer	Oregon Department of Fish and Wildlife
Description	The Oregon Department of Fish and Wildlife initiated sampling in 2014 to develop general mosquito abundance baseline data and species composition information for the Department owned properties near Coquille, OR
Variables:	
Temporal Coverage	2014

Units of Measurement	Mosquito Count by Species
Geographic Coverage	OR Department of Fish and Wildlife Winter Lake Tract and Beaver Slough Tract, and two other sites: Hinch Bridge on South Slough, and near Catching Slough
Other	
File Format/s	PDF, XLS
The following metadata elements are highly desired, but not required for inclusion in the collection:	
Funding Source/s	
Data Owner/Publisher	Oregon Department of Fish and Wildlife
Data Collector/Producer Affiliation	Oregon Department of Fish and Wildlife
Data Identifier	
Publication Year	2015
Data Collection Instruments	Bioquip Heavy Duty EVS CO_2 Mosquito Traps
Description of Non-Standard Abbreviations	
Terms and Conditions for Use	

Category: Dynamic Database

6. http://django.msu.montana.edu/MTmosquito/

Montana Mosquito Surveillance Database

Selection Criteria	
Relevance to Collection Mission	This data fit well within the mission of this collection.
Scientific/Analytic Value and Potential	The data available in this database can be used alone or can be paired with additional data about the area learn more about Montana-area mosquito populations and their potential for change.
Uniqueness	While this is a database rather than a static dataset, it is a comprehensive resource of mosquito trapping and identification throughout Montana state and thus a valuable asset to this collection.
Usability	Raw data are available within each county's data link as a CSV. The CSV format is ideal for many types of data analysis software.
Metadata Availability	Required metadata is provided and some of the desired metadata can be extracted from the website and the CSV files.
Preservation State	Medium
Metadata Criteria	
Data Collector/Producer	Montana State University
Description	The purpose of this website is to document the species and distribution of mosquitoes collected in Montana. Following the arrival of West Nile virus in Montana in August 2002, mosquito surveillance began through a cooperative project involving Montana State University's Veterinary Entomology Lab under the guidance of Dr. Greg Johnson and Marni Rolston, Dr. Grant Hokit and Dr. Sam Alvey of Carroll College, and the Montana Department of Public Health and Human Services. The information presented through this website is a result of these monitoring efforts and includes mosquito species collected, yearly and seasonal population dynamics, in-state species distribution, notes on biology and West Nile virus presence.
Variables:	
Temporal Coverage	2004 to 2015
Units of Measurement	Count of Mosquitoes by Species

Geographic Coverage	Montana
Other	
File Format/s	CSV
The following metadata elements are highly desired, but not required for inclusion in the collection:	
Funding Source/s	Funding for construction of this website came in part from grant #52007534 to Carroll College from the Howard Hughes Medical Institute through the Precollege and Undergraduate Science Education Program and by Grant Number P20 RR16455-09 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH).
Data Owner/Publisher	Montana State University
Data Collector/Producer Affiliation	Montana State University
Data Identifier	
Publication Year	
Data Collection Instruments	Mosquitoes were collected using battery-operated CDC light traps with CO_2 as an attractant. Light traps were run either one night per week or once every two weeks. Samples were shipped to the Veterinary Entomology Lab where the specimens were identified, counted and potential vectors tested for West Nile virus.
Description of Non-Standard Abbreviations	
Terms and Conditions for Use	The information contained herein is provided as a public service, with the understanding that the Montana State University makes no warranties, either expressed or implied, concerning the accuracy, completeness, reliability, or suitability of the information. Montana State University makes no warranties that the information is free of any copyright infringement.

Appendix References

Centers for Disease Control. (2016, May 05). Vector Surveillance and Control. Retrieved May 12, 2016, from http://www.cdc.gov/zika/vector/

Coos County Oregon. (2015). Bandon Marsh Mosquito Control. Retrieved May 11, 2016, from http://www.co.coos.or.us/Departments/CoosHealthWellness/PublicHealth/EnvironmentalHealth/Mosquito-Vector.aspx

DataPASS. (2005). DataPASS Appraisal Guidelines. Retrieved May 9, 2016, from http://www.data-pass.org/sites/default/files/appraisal.pdf

Guerra, C., Reiner, R., Perkins, T., Lindsay, S., Midega, J., Brady, O., Barker, C., Reisen, W., Harrington, L., Takken, W., Kitron, U., Lloyd, A., Hay, S., Scott, T. and Smith, D. (2014). A global assembly of adult female mosquito mark-release-recapture data to inform the control of mosquito-borne pathogens. *Parasites & Vectors*, 7, 276.

ICPSR. Guide to Social Science Data Preparation and Archiving Phase 3: Data Collection and File Creation. (n.d.). Retrieved May 10, 2016, from https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter3docs.html

Kraemer MUG, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, Messina JP, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Wint GRW, Elyazar IRF, Teng H, Hay SI (2015) Data from: The global compendium of Aedes aegypti and Ae. albopictus occurrence. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.47v3c

Kraemer MUG, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, Messina JP, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Wint GRW,

Elyazar IRF, Teng H, Hay SI (2015) The global compendium of Aedes aegypti and Ae. albopictus occurrence. Scientific Data 2(7): 150035. http://dx.doi.org/10.1038/sdata.2015.35

Montana State University. (n.d.). Montana Mosquito Surveillance. Retrieved May 11, 2016, from http://django.msu.montana.edu/MTmosquito/

Rumsey, S. & Jefferies, N. (2013). Challenges in Building an Institutional Research Data Catalogue. *International Journal of Digital Curation, 8*(2), 205-214.

Rumsey, S., & Jefferies, N. (2013, July 9). DataFinder: A Research Data Catalogue for Oxford. *Ariadne*. Retrieved May 10, 2016, from http://www.ariadne.ac.uk/issue71/rumsey-jefferies

The University Corporation for Atmospheric Research. (n.d.). Climate Change and Vector-Borne Disease. Retrieved May 12, 2016, from http://scied.ucar.edu/longcontent/climate-change-and-vector-borne-disease

Weicheld, J., & Treser, Charles D. (2015). Data from: *Impact of environmental factors on mosquito population abundance and distribution in King County, Washington. Seattle*: University of Washington.

Weicheld, J., & Treser, Charles D. (2015). *Impact of environmental factors on mosquito population abundance and distribution in King County, Washington. Seattle*: University of Washington.

Wikipedia. (2016, April 8). Vector (epidemiology). In *Wikipedia, The Free Encyclopedia*. Retrieved 16:19, May 12, 2016, from https://en.wikipedia.org/w/index.php?title=Vector_(epidemiology)&oldid=714158700

Whyte, A. & Wilson, A. (2010). "How to Appraise and Select Research Data for Curation". DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides