**Emerging Infectious Disease Repository (EIDR): A web app dedicated to emerging infectious disease events**

Zachary S. Gold1, Andrew G. Huff1, Amy M. Slagle1, Jonathan K. Goley1, Toph Allen1, Nathaniel Breit1, Nico Preston, Carla G. Tilchin, & Peter Daszak1

1EcoHealth Alliance, 460 West 34th Street, 17th Floor, New York, New York 10001, USA

**Abstract**

Emerging Infectious Disease Repository (EIDR) combines a curated and transparent database on past emerging infectious disease (EID) events occurring between 1940 and 2013. EIDR is an interactive web application built to disseminate and discuss this information. The EIDR database contains 350 EID events, the majority of which (296) derive from the database created by Jones et al. (2008). The composition of the EIDR database largely validates the findings of Jones et al. (Jones et al. 2008), but also reveals the limitations of compiled EID case studies. The challenge associated with collecting comprehensive information on complex historical events is considerable. Furthermore, precisely defining and discerning key aspects of EID events is a difficult task.

**Introduction**

In 2014, the World Health Organization estimated that in 2012 infectious and parasitic diseases accounted for 15.8% of all disability-adjusted life-years (DALYs) (WHO 2014). Although this is a 4% decrease from an estimate of 19.5% of all DALYs in 2000, the rate of pathogen emergence is increasing, and the opportunity for pandemic emergence remains high (WHO 2014; Woolhouse et al. 2008). Globalization has produced an efficient mechanism to spatially distribute EIDs. A mere 20 years after HIV’s discovery, HIV was the fourth highest cause of death worldwide (Smolinski, Hamburg, and Lederberg 2003). More recently, the vulnerability of global health security was revealed in the ongoing Ebola Virus Disease epidemic in West Africa. The impact of the Ebola Virus Disease epidemic has been catastrophic, causing well over 8,000 deaths in one year, and incurring an estimated total financial impact of between $3.8 and $32.6 billion (USD) internationally (World Bank Group 2014). To combat these EID threats we must understand the driving factors behind their emergence so that informed and effective prevention, preparation, and response strategies can be developed.

Studying the origins of EIDs is a complex task that has been approached from many angles including qualitative and quantitative analyses including segmented regression models (Smolinski, Hamburg, and Lederberg 2003; Taylor, Latham, and Woolhouse 2001; Weiss and McMichael 2004; Jones et al. 2008; Woolhouse et al. 2012; Morens, Folkers, and Fauci 2004; Funk et al. 2013). One approach has focused soley on identifying and studying EID events (Jones et al. 2008; Grace D 2012). EID events represent the earliest known emergence of EIDs (Jones et al. 2008). This approach has been useful, notably yielding a map of ‘hotspots’ of disease emergence developed by Jones et al. Despite the face validity of the method, the method may have limited utility in furthering the study of EIDs. The specific information surrounding historic emergence events is often complex, difficult to find, and even more difficult to validate (Funk et al. 2013). EIDR (the Emerging Infectious Disease Repository) was developed from the database used by Jones et al. (Jones et al. 2008) to explore EID events in greater detail by creating an expanded, highly curated, database of EID events.

**Methods**

***EID Event Collection***

For the purpose of EIDR, an EID event is defined as the original case or cluster of cases representing the emergence of an infectious disease in human populations (Jones et al. 2008). Emergence is defined as the development of any of the following with respect to a given microorganism: (1) earliest instance of natural human infection, (2) reappearance after control or elimination, (3) new or increasing drug resistance, (4) new or expanding geographic region, (5) increasing incidence, or (6) increasing virulence. Events are not counted amongst EID events unless some clinical significance and relevance is attributable to the pathogen in question. Potential EID events were evaluated based on these definitions by emerging infectious disease experts at EcoHealth Alliance and classified as EID events if they met any of the above criteria. See Appendix A for detailed descriptions of these emergence categories.

The events in EIDR date back to 1940, a cut-off chosen by Jones et al. (Jones et al. 2008), and informed by the Institute of Medicine’s resources on EIDs (Smolinski, Hamburg, and Lederberg 2003). Potential EID events were collected from a review of meta-analyses on disease emergence, or through an internal literature review. All events between 1940 and 2004 derive from Jones et al. (Jones et al. 2008), which relied heavily on the review by Taylor and colleagues (Taylor, Latham, and Woolhouse 2001). Events between 2004 and 2013 derive from a recent effort to map emerging zoonoses (Grace et al. 2012), a review of trends in viral discovery (Rosenberg et al. 2013), or were compiled through a review of the literature.

***Data Collection and Review***

For each EID event data were collected on a set of variables identified as important by a team of EcoHealth Alliance experts. These variables are designed to capture critical spatial, temporal, clinical, epidemiologic, economic, pathogen, and host information. Data were also collected on potential drivers associated with each EID event, like war and famine, antimicrobial use, or proximity to wildlife. Driver categories are based on those found in Smolinski et al. (Smolinski, Hamburg, and Lederberg 2003) and Lederberg et al. (Lederberg, Shope, and Oaks 1992), but some categories are removed and others are broken down further. Emergence locations are resolved to the most specific spatial information available, frequently a polygon representing the smallest administrative region associated with an event. Rarely, multiple potential emergence locations are provided for a single event due to insufficient temporal information within the available literature. Events were independently reviewed at least twice by EcoHealth Alliance veterinarians, disease ecologists, epidemiologists, and public health specialists. A list of all variables and their sub-categories can be found in Appendix A.

Short abstracts are included for all events. When possible, direct language from text was captured to justify values for subjective variables. If no information could be found on a particular variable this absence was captured. General contextual information for each event was acquired from various sources, many unrelated to EID events. For example, taxonomic information is from the National Center for Biotechnology Information (NCBI 2015), and economic information is from the World Bank (World Bank Group 2015).

**Results**

***EID Events***

EIDR contains 350 EID events; 296 from Jones et al. (Jones et al. 2008), 38 from Grace et al. (Grace D 2012), 6 from Rosenberg et al. (Rosenberg et al. 2013), and 10 from the internal literature review. For a list of the 39 events deriving from the Jones et al. (Jones et al. 2008) database that were excluded from the EIDR database see Appendix B. EIDR is dominated by EID events caused by bacteria (50.0%) and viruses (31.7%). Vector-borne diseases are associated with 22.0% of EID events in EIDR. EID events occurred primarily in North America (31.7%), Europe (24.3%) and Asia (18.3%), although no adjustment has been made to offset reporting bias.

Events representing the earliest instance of natural human infection by a microorganism are the most numerous (55.7%), followed by events representing new or expanding drug resistance (20.9%), increasing incidence (6.9%), new or expanding region (6.3%), increased virulence (5.7%), and reemergence after control or elimination (5.1%). The most commonly identified cause of EID events is the use of antimicrobial agents (20.3%). Other significant causes include human susceptibility to infection (18.9%), proximity to wildlife (11.1%), human behavior (11.1%), and ecosystem change (10.6%). Notably, no cause could be identified for 28.9% of EID events.

The majority of EID events involve known zoonotic pathogens (63.1%). Just over half of these zoonotic EID events involve a specifically documented instance of transmission of a microorganism from animals to humans during the event (52.0% of zoonotic EID events, 32.9% of all EID events). The type of transmission that occurred during EID events that involve zoonotic pathogens, but lack evidence of a specifically documented instance of transmission of a microorganism from animals to humans, is primarily unknown (63%), although nosocomial transmission occurred in a notable number (13.9%).

***EIDR Web Application***

An interactive web-application displays the information stored in EIDR (http://eidr.ecohealthalliance.org/). Through the website EID events can be explored in a variety of ways. The “Emergence Events” view contains a table of EID events. Users can choose which EIDR variables they would like to view (Fig. 1), and perform specific searches using a filter feature. This feature allows users to search for events with a common variable, like a specific host, or pathogen. A map of all EID events is offered through the “Event Map” view (Fig. 2). Additional methods and variable definitions are available on the “About”, and “Variable Definitions” pages respectively.

Users can explore individual EID events in greater detail through individual event pages. Event pages can be accessed by clicking an event in the “Emergence Events” table. Each EID event page contains a detailed report on the event, including a narrative abstract, a map showing the location of the event, tables of additional data, and a discussion board that allows users to comment on the event (Fig. 3). In some cases data is displayed with supporting textual evidence. References for each event are available in the event pages.

**Conclusion**

EIDR is the combination of an expansive, highly curated and transparent database of EID events with a user-friendly, engaging and interactive web application. The composition of the EIDR database largely validates the findings of Jones et al. (Jones et al. 2008). EIDR and the Jones et al. database contain comparable percentages of EID events associated with zoonotic diseases (63.1%, 60.3%), vector-borne diseases (22.4%, 22.8%), bacteria (50.0%, 54.3%), and viruses (31.7%, 25.4%). This paper and Jones et al. (Jones et al. 2008) identify antimicrobial agent use as the most common cause of EID events. Despite these validations, the EIDR initiative uncovers some of the limitations of using EID events to study trends in disease emergence. Specific limitations encountered during the creation of EIDR include the following. Substantial labor is required to create reliable EID event databases. The creation of the EIDR database was undertaken over the course of several years. Future studies can use the EIDR database to reduce labor, however, maintaining the EIDR database so that it reflects the most current definition of emergence, and contains the most recent EID events is a formidable task. Events are often complex and have been subjected to varying levels of investigation. Verifying potential EID events that were noted in previous studies is sometimes difficult. Extensive effort was made to search for these potential EID events, often with limited success. Identifying the root causes of many events is a challenge, particularly for zoonotic disease transmission events that may require additional studies to determine host involvement. Lastly, the classification of events as EID events is shrouded in ambiguity. It is particularly difficult to critically evaluate events that involve increasing incidence, or geographic expansion. Funk et al. (Funk et al. 2013) correctly assert that a more quantifiable definition is needed for EIDs. The transparency and accessibility of the EIDR database is made possible by the EIDR web-application and should spur constructive conversations about the most effective methods to study and define disease emergence objectively. Future research should be conducted to determine and create more objective and comparable definitions for emerging infectious diseases. A concrete definition of disease emergence is needed so that the scientific community can make effective comparisons between diseases.

**Acknowledgment**

Special thanks to Alexa Frank, Elizabeth Eckel, Amelia Averitt, Sarah Elwood, and Emily Hagan for completing EIDR data collection and review. Thank you to Zack Krejci and Mark Gibson for all the development work you provided. Thanks so much to Melinda Rostal, Allison White, Parviez Hosseini, and Carlos Zambrana-Torrelio for your thoughtful reviews. Thank you to Peter Daszak, William Karesh, Jonathan H. Epstein, and Kevin J. Olival for your advice and guidance. Thank you Kate Jones, Nikkita Patel, Marc Levy, Adam Storeygard, Deborah Balk, John Gittleman, and Andrew Dobson for the inspiration for the project. This study was made possible by the generous support of the American people through the United States Agency for International Development (USAID) Emerging Pandemic Threats PREDICT, and through the support of the Defense Threat Reduction Agency (DTRA) through a contract (Contract No. HDTRA1-13-C-0029) awarded to EcoHealth Alliance. The contents are the responsibility of the authors and do not necessarily reflect the views of USAID, DTRA or the United States Government.

**References**

Funk, S., T. L. Bogich, K. E. Jones, A. M. Kilpatrick, and P. Daszak. 2013. 'Quantifying trends in disease impact to produce a consistent and reproducible definition of an emerging infectious disease', *PLoS One*, 8: e69951.

Grace D, Mutua F, Ochungo P, Kruska R, Jones K, Brierley L, Lapar ML, Said M, Herrero M, Phuc PM, Thao NB, Akuku I, Ogutu F. 2012. "Mapping of poverty and likely zoonoses hotspots." In. Nairobi Kenya: ILRI.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. 2008. 'Global trends in emerging infectious diseases', *Nature*, 451: 990-3.

Lederberg, J., R. E. Shope, and S. C. Oaks. 1992. *Emerging Infections: Microbial Threats to Health in the United States* (Institute of Medicine, National Academy Press: Washington (DC)).

Loh, E., Olival K. J., Zambrana-Torellio C., Bogich T. L., Johnson C. K., Mazet J. A. K., Karesh W. B., Daszak P. . In Press. 'Targeting transmission pathways for emerging zoonotic disease surveillance and control. ', *Vector Borne and Zoonotic Diseases*.

Morens, D. M., G. K. Folkers, and A. S. Fauci. 2004. 'The challenge of emerging and re-emerging infectious diseases', *Nature*, 430: 242-9.

NCBI. 2015. 'Home - Taxonomy - NCBI'. <http://www.ncbi.nlm.nih.gov/pubmed/>.

Rosenberg, Ronald, Michael A. Johansson, Ann M. Powers, and Barry R. Miller. 2013. 'Search strategy has influenced the discovery rate of human viruses', *Proceedings of the National Academy of Sciences*, 110: 13961-64.

Smolinski, M. S., M. A. Hamburg, and J. Lederberg. 2003. (National Academies Press: Washington (DC)).

Taylor, L. H., S. M. Latham, and M. E. Woolhouse. 2001. 'Risk factors for human disease emergence', *Philos Trans R Soc Lond B Biol Sci*, 356: 983-9.

Weiss, R. A., and A. J. McMichael. 2004. 'Social and environmental risk factors in the emergence of infectious diseases', *Nat Med*, 10: S70-6.

WHO. 2014. 'WHO | Global Health Estimates', *WHO*.

Woolhouse, M. E., R. Howey, E. Gaunt, L. Reilly, M. Chase-Topping, and N. Savill. 2008. 'Temporal trends in the discovery of human viruses', *Proc Biol Sci*, 275: 2111-5.

Woolhouse, M., F. Scott, Z. Hudson, R. Howey, and M. Chase-Topping. 2012. 'Human viruses: discovery and emergence', *Philos Trans R Soc Lond B Biol Sci*, 367: 2864-71.

'World Bank Group'. 2015. <http://www.worldbank.org/>.

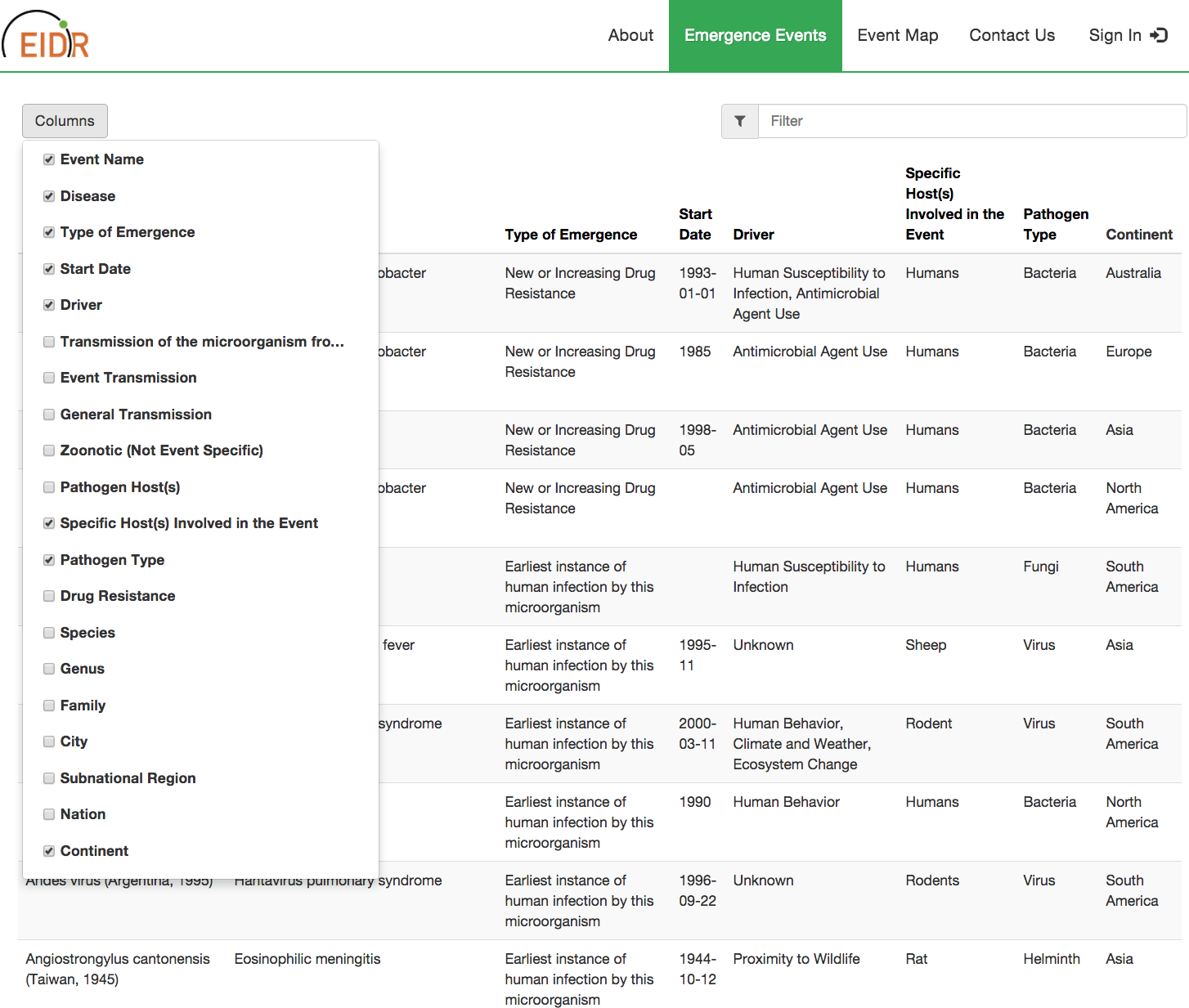


Figure 1. The customizable “Emergence Events” view allows users to sort, explore, and compare EID events by selecting EIDR variables for the table to display. Clicking any of these events brings the user to the “Event page” for that event.

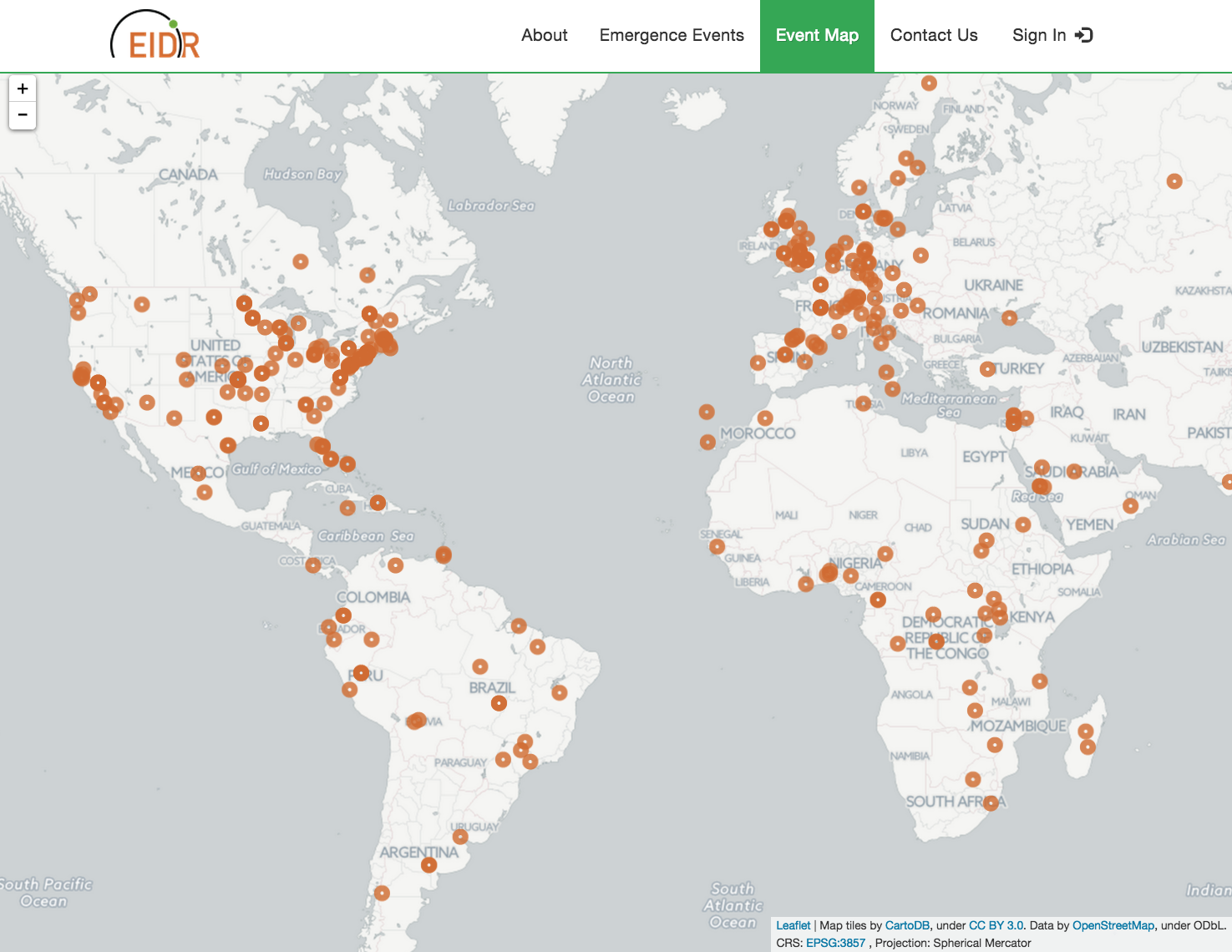


Figure 2. The “Event Map” view depicts the most specific spatial origination of all EID events. Individual events can be accessed by clicking on their corresponding pins.

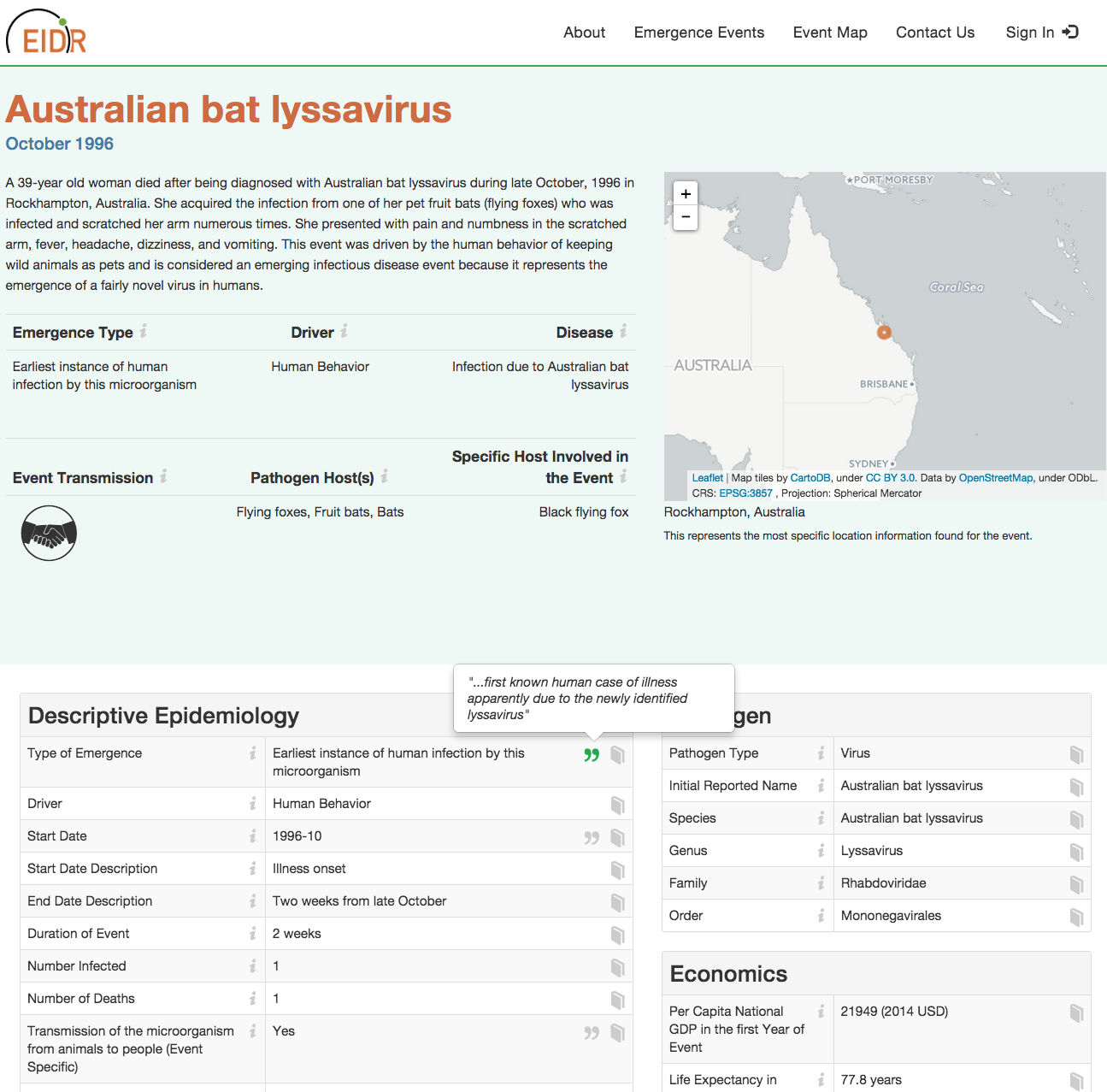


Figure 3. The event page includes an abstract describing the event, a map with the most specific emergence location for the events, and tables displaying different sets of data for a particular event. Users can view short definitions of each variable by clicking on the information icon the to the right of each variable. For certain variables, quotes pertaining to a variable value can be viewed by hovering over the “ icon. The references used to collect data can be found by clicking on the book icons.

**Appendix A. Variable Definitions and Explanations**

**Descriptive Epidemiology Variables**

**Type of Emergence**

*Description:* The explanation as to why this event is considered an emergence event.

*EIDR Values:*

* New or Expanding Region: The pathogen appeared in a region that is significantly distant from any other regions it is found in, or the geographic range of the pathogen showed marked expansion.
* Increasing incidence: There was a marked increase in incidence of the pathogen.
* New or Increasing drug resistance: The pathogen displayed a novel drug resistance, or a rare drug resistance was observed to be increasing in incidence or geographic range. Events that involve the development of drug resistance to a single antimicrobial agent, and events that involve the development of resistance to multiple antimicrobial agents are including.
* Earliest instance of natural human infection by this microorganism: This event pertains to the earliest confirmed case(s) of human infection associated with a microorganism.
* Increasing virulence: The pathogen showed a marked increase in virulence.
* Reappearance after control or elimination: The pathogen reappeared for the first time after a significant period of no cases.

**Driver**

*Description:* Potential factor(s) that may have contributed to the emergence event.

*EIDR Values:*

* Human susceptibility to infection: Any immune deficiency from an underlying disease, condition, age, or scenario.
* Antimicrobial agent use: The overuse, or misuse of antimicrobial agents.
* Agricultural industry changes: Intensification of agriculture practices, use of antibiotics or pesticides, new agricultural practices, new feeds.
* International travel and commerce: Human travel, goods trade, invasive species from trade or travel.
* War and Famine: War, conflict, political unrest, malnutrition, famine, substandard living conditions due to conflict, refugee camp conditions, intent to harm.
* Medical Industry Changes: New vaccination practices, changes in medical protocol, new techniques, new equipment.
* Climate and Weather: Like, climate change, deforestation, drought, rainy season, tsunami, hurricane, typhoon, flood, earthquake, stagnant pools, heat wave.
* Human Behavior: Living conditions, population density, urbanization, migration, food consumption, drug use, human gatherings, daily life routines, recreation.
* Proximity to wildlife: Living near wildlife, suburbanization, being in the wild. Bushmeat: Consuming, hunting, or trading bushmeat.
* Breakdown of public health measures: Like, issues of vaccine production, vaccine use, vaccine enforcement, sanitation, failure to quarantine, toxic chemical exposure, water quality, air quality, hazardous waste management.
* Ecosystem change: Changes to the environment, like, expansion of agricultural lands, or clear-cutting that dramatically alters an area's ecosystem.
* Unspecified: There was insufficient information to determine a driver for the event.

**Disease**

*Description:* The name of the disease or syndrome that was associated with the pathogen during the event. Generalized names, such as “infections due to Enterococcus facecium” are provided when no specific disease was mentioned.

**Start Date**

*Description:* The date of the onset of the event.

**Start Date Description**

*Description:* The context of the start date.

*EIDR Values:* Illness onset, first case presented to official, first report, first date of study, first pathogen isolation, treatment initiation, initial hospitalization, beginning of increasing incidence, first exposure

**End Date**

*Description:* The date of the end of the event.

**End Date Description**

*Description:* The context of the end date.

*EIDR Values:* Death of single case patient, death of last patient, symptom resolution and no further cases, study discontinued or completed, last pathogen isolation, last report, last illness onset, chronic illness, end of increasing incidence, last exposure

**Duration of Event**

*Description:* The length of the event as determined by the start and end dates reported.

**Number Infected**

*Description:* The total number of confirmed human infections during the event. Occasionally, suspected or probable cases were included in this variable.

**Number of Deaths**

*Description:* The total number of deaths due to the pathogen during the event.

**Zoonosis (Not Event Specific)**

EIDR Values: Yes/No

*Description:* Is the microorganism associated with this event the etiologic agent of infection in humans and animals? This variable is not specific to the event and does not entail transmission of the microorganism from animals to humans during the event.

**Event Specific: Transmission of the microorganism from animals to people.**

EIDR Values: Yes/No/Uncertain

*Description:* Is there behavioral or textual evidence of direct or indirect transmission of the microorganism from animals to humans during the event?

**Pathogen Host(s)**

*Description:* The suspected host(s) of the pathogen, if applicable. Hosts may be reservoir, parenteral, dead-end, or accidental.

**Specific Host(s) Involved in the Event**

*Description:* The name of the host(s) involved in the event.

**Event Transmission**

*Description:* The mode of transmission by which the infected persons acquired the pathogen during this event. Many of these definition were obtained from Loh et al. (Loh In Press).

*EIDR Values:*

* Direct transmission: The pathogen was transmitted through skin-to-skin contact; the bite of an infected animal; contact with body fluids, organs, and tissues; direct contact with large droplet >5 µm.
* Sexual transmission: The pathogen was transmitted during sexual contact.
* Vertical transmission: The pathogen was transmitted from the mother to her child in the womb.
* Vector transmission: The pathogen was transmitted by the bite of, or mechanical transfer from, an arthropod.
* Nosocomial transmission: The pathogen was transmitted in a hospital or other health care facility.
* Airborne transmission: The pathogen was transmitted through aerosolized particles  <5 µm.
* Oral transmission: The pathogen was transmitted through the consumption of contaminated food or water.
* Contamination: The pathogen was transmitted by indirect contact with soil or vegetation, contact with water, or indirect transmission from contaminated inanimate objects.
* Transmission from infected animals: The pathogen was acquired from direct or indirect contact with an infected animal.
* Unknown: The mode of transmission is unknown

**General Transmission**

*Description:* The general modes of transmission by which the pathogen can be acquired. Many of these definition were obtained from Loh et al. (Loh In Press).

*EIDR Values:*

* Direct transmission: The pathogen is transmitted through skin-to-skin contact; the bite of an infected animal; contact with body fluids, organs, and tissues; direct contact with large droplet >5 µm.
* Sexual transmission: The pathogen is transmitted during sexual contact.
* Vertical transmission: The pathogen is transmitted from the mother to her child in the womb.
* Vector transmission: The pathogen is transmitted by the bite of, or mechanical transfer from, an arthropod.
* Nosocomial transmission: The pathogen is transmitted in a hospital or other health care facility.
* Airborne transmission: The pathogen is transmitted through aerosolized particles  <5 µm.
* Oral transmission: The pathogen is transmitted through the consumption of contaminated food or water.
* Contamination: The pathogen is transmitted by indirect contact with soil or vegetation, contact with water, or indirect transmission from contaminated inanimate objects.
* Transmission from infected animals: The pathogen is acquired from direct or indirect contact with an infected animal.

**Reported Symptoms**

*Description:* The event specific symptoms reported. Percentages reported reflect the prevalence of each symptom within the infected individuals from the event. Bracketed terms reflect symptoms severity.

**Testing Method**

*Description:* The testing method used to determine the pathogen. Note, variable values and data for this variable are being re-collected.

*EIDR Values:* Serology: The pathogen was identified through serological antibody techniques. Direct blood: The pathogen was directly identified from a blood sample without serologic markers. Direct fecal: The pathogen was directly identified from a fecal sample. Direct other: The pathogen was directly identified from a sample taken from any other source, such as a tissue sample. Genetic sequencing: The pathogen was identified through genetic sequencing.

**Host Use**

*Description:* The description of the interaction or relationship between the infected human and the host involved in the event.

*EIDR Values:* Hunted: The host was hunted. Eaten: The host was eaten. Markets or traded: The host was traded or was encountered in a market. Sex: The host was used for sexual intercourse. Pet: The host was a pet; medical: The host was used in a medical setting, for example an organ transplant or blood transfusion; other: The use of the host in pathogen transmission can not be defined by any of the other categories; none: There was no use of the host involved in pathogen transmission; unknown: The use of the host is unknown in pathogen transmission.

**Host Age**

*Description:* The age of the host involved in the event.

**Drug Resistance**

*Description:* Whether or not drug resistance was reported in the event and related to inclusion as an EID event.

*EIDR Values: Yes/No*

**Occupation**

*Description:* Epidemiologically relevant occupations of infected persons.

**Average Age of Infected**

*Description:* The average age of the infected persons.

**Average Age of Death**

*Description:* The average age of the infected persons who died.

**Pathogen Taxonomy Variables**

**Taxonomy:** All taxonomic information, except Pathogen Type, and Initial Reported Name, was acquired from the NCBI taxonomy browser (NCBI 2015).

**Pathogen Type**

Classification as a bacteria, fungi, helminth, prion, protozoa, rickettsia, or virus.

**Initial Reported Name**

The name of the pathogen exactly as it was reported in the literature at the time of the event.

**Spatial Variables**

The most specific spatial information identified for the origin of the emergence event.

Emergence locations are resolved to the most specific spatial information available, frequently a polygon representing the smallest administrative region associated with an event. Rarely, multiple potential emergence locations are provided for a single event due to insufficient temporal information within the available literature.

**Economics Variables**

**Per Capita National GDP in the first Year of Event**

*Description:* The average per capita GDP of the country during the first year of the event as documented by the World Bank ('World Bank Group' 2015). Information is not available for events prior to 1960.

**Life Expectancy in Country in the first Year of Event**

*Description:* The average life expectancy of people in the country during the first year of the event as documented by the World Bank ('World Bank Group' 2015). Information is not available for events prior to 1960.