EIDR Connect

EIDR Connect (EIDR-C) is a BSVE and web app for structuring and visualizing textual data on infectious disease bioevents.

EIDR-C allows analysts to curate important data from reports, browse both curated and automatically extracted events underlying various reports, and view and export visualizations and summaries of the epidemiologic characteristics of a developing event.

BACKGROUND

Listservs (e.g. ProMED-mail), biosurveillance feeds (e.g. The CDC Weekly Report), and other news sources (e.g. Google News) provide up-to-date and readily available information on where and when infectious diseases bioevents are occurring. However, this information is provided across a series of unstructured text documents. This limits the value they can provide to analysts in detecting and assessing emerging disease threats; the data they contain cannot be integrated with other data sources, systematically visualized and summarized, or used for modeling purposes to compute downstream metrics. With the right tools, the underlying information contained in this rich data can be extracted, structured, and harnessed to better mitigate biochemical threats.

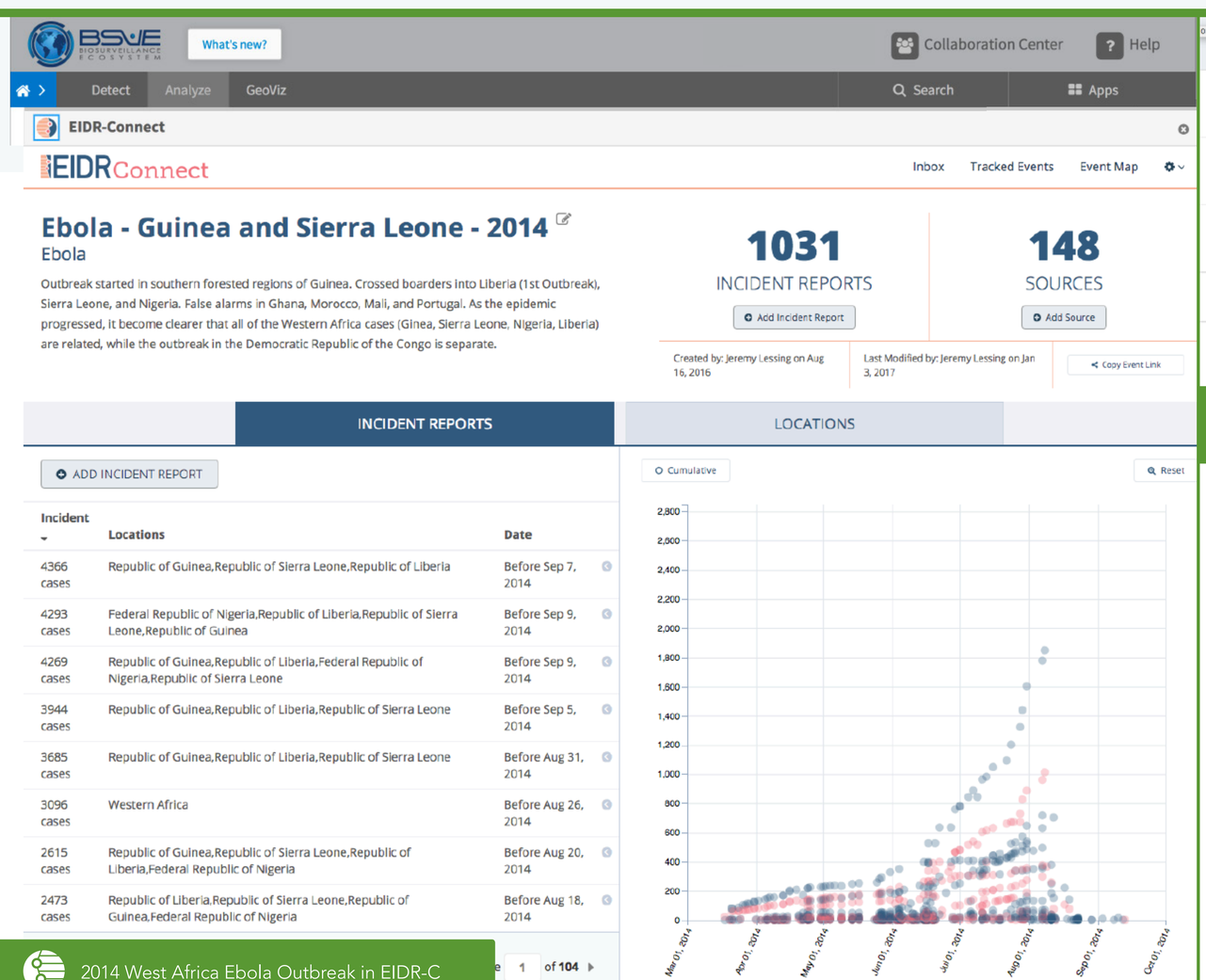
TECHNOLOGY AND INTERFACE

EIDR Connect:

- provides a set of tools for analysts to extract, review, and visualize data on the underlying events described across multiple textual biosurveillance reports
- uses natural language processing (NLP) and machine learning to facilitate analyst workflows
- automatically synthesizes bioevents from textual data (in development)

EIDR-C's data model expresses the underlying structure of bioevent data in text:

- Source – A text document containing information on a bioevent of interest. EIDR-C currently ingests the ProMED-mail feed. It can also receive pasted text and retrieve the contents of submitted URLs.
- Incident Report – A structured unit of information extracted from a Source text. Incident Reports (IRs) predominantly refer to case or death counts but can contain other data, and have an associated location and time/duration. An intelligent algorithm extracts them from text, and they can be confirmed, rejected, or edited by an analyst.
- Event – Events are groups of Incident Reports about a single bioevent. The current collection of events is expert-curated based on epidemiologic criteria (e.g. single chain of transmission).



2014 West Africa Ebola Outbreak in EIDR-C

FEATURES

INBOX:

- Browse a list of Source documents and associated information.
- View extracted Incident Reports and suggested Event associations, with the ability to confirm or reject suggestions.
- Mark Sources as "Reviewed" to hide them from view within the inbox.

EVENTS:

- Browse a table or map of tracked events.
- View, edit, and export detailed information for a specific event, including:
  - Lists of Incident Reports and their Sources
  - Scatterplots of Incident Report data, including case and death counts

RELATED WORK

- GRITS – Houses some of the NLP technology used in EIDR Connect.
- Incident Extractor – Run EIDR Connect's NLP on a single document and export the results as a table.

FUTURE WORK

- A "Send to EIDR Connect" button in BSVE search results, plus a shared set of analyst-created events.
- Additional source data feeds.
- Machine learning algorithm will improve as more data is curated.
- Clustering algorithms to automatically generate events in real-time as text feeds are updated.
- New interactions to speed up the workflow of dynamically creating Events.
- New and improved visualizations of extracted data.

eidr-connect.eha.io

For more information, contact:

Toph Allen
allen@ecohealthalliance.org

Brock Arnold
arnold@ecohealthalliance.org

EcoHealth Alliance