Background

• Viruses, bacteria, fungi and other agents are constantly changing because of nucleic acid transcription errors and recombinations,
• Most nucleic acid transcription errors and recombinations result in less effective functionality and therefore weaker pathogens.
• Rarely nucleic acid transcription errors and recombinations result in more effective pathogens. Pathogens with the potential to kill thousands if not millions can emerge.
• Information must spread faster. Standard based data collection (SBDC) is one partial solution to the rapid collection and sharing of data.

Two Examples of the Need for Rapid Collection and Sharing of Data

Methods

The author is a board certified residency trained Public Health Physician with 14 years of outbreak investigation experience in two states, informatics training as a DBMI Postdoctoral Scholar and experience as a key partner during the development, alpha and beta testing of the New Jersey Electronic Birth Certificate System.

I benefited from the contributions of a large number of individuals. Rich Tsui PhD suggested REDCap, John Vander suggested using codes for specific questions, Michael Wagner MD, PhD provided his laboratory. Allan John Ashby, Joyce Zeinis and Bill Shirey provided access to the DBMI REDCap.

Hillary Booth and the Oregon Integrated Food Safety Center of Excellence provided an Oregon survey data dictionary with element names used for the examples in this poster. The variables are called KeeneCodes in honor of the late Bill Keene. See http://www.outbreakmuseum.com/dr-bill-keene/

From:
http://wiki.siframework.org/Structured+Data+Capture+Initiative
2. H1N1 Public Health Emergency Response Team http://www.h1n1.org/special/committees/pher/index.cf

For Data Between Local Health Department and State/CDC:
1. The Public Health Informatics Network (PHIN) http://www.cdc.gov/phln/

For Local Case Investigation Data:
Standards based data collection (SBDC) holds the promise of rapid data collection, sharing and analysis of information from persons about exposures relevant to infectious disease investigations.

For instances where new data elements/variables must be generated, jointly used local investigation data elements (JL Codes) would allow investigators to directly add questions to preprogrammed data collection apps and/or web survey forms in a standardized manner.

Potential tools include:
2. REDCap Research Electronic Data Capture - http://project-redcap.org/ *REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies, providing 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources.*
3. Automated algorithms could increase and speed the identification of new outbreaks and their infectious reservoirs.

Discussion

During the Ebola outbreak, some modelers decried the lack of good data. However given the challenges in the field, the data received reflected the difficulties of field data collection under life threatening conditions. Some workers gave their lives.

Standardized data elements are a logical and tested solution to make this process easier. The New Jersey Electronic Birth Certificate system permitted the upload of CSV data from electronic health records and published its first data dictionary 20 years ago – see http://www.nj.gov/health/vital/documents/data_dictionary.pdf. CSV still works today.

JL codes including the Keene Codes already developed by the Oregon Health Department allow the creation of data collection tools rapidly by using pre-existing or newly created questions. When data is collected with uniform fields and variables, they can more easily be shared between local public health partners.

REDCap is one tool that can allow the creation of manual and electronic (web and smart phone) data collection tools quickly. No programming skills are needed. Data collection instruments can be created in less than one hour. Results are in a machine readable format.

Better data helps everyone: Decision makers, Investigators, Modelers and most of all the People.