Misjudgements in the prediction, detection and response to infectious diseases outbreaks mean that epidemics and pandemics remain among humanity’s greatest threats. Increased risk is seen in settings with high population growth, urbanization and connectivity but with compromised health systems and political instability. Beginning in 2014 the Ebola outbreak in West Africa ravaged communities and economies for more than two years, leading to more than 28,616 cases, 11,310 deaths and billions of dollars lost in national gross domestic products (GDPs). Despite the disappearance of Ebola from mainstream media, there have been three additional outbreaks since.

Prevention and control of new outbreaks require working with data from a variety of domains, including population density, travel, trade, environmental factors, land use and human-wildlife interactions. This data is important for the understanding of who is at risk, where at-risk populations are, under which environment they are situated and how they can be reached (2).

Most of this data is available, but remain isolated in silos within the various domain- specific communities, often with their own domain-specific formats, vocabularies and ontologies. Integration of the varied data resources on disease vectors and transmission routes has great potential to identify underlying relationships that determine the trajectory of an outbreak. Each of these are vital to maximise the effectiveness of a response. However, those working to prevent infectious disease outbreaks from reaching epidemic or pandemic levels are unable to access and analyse the critical, disjoint data with the potential to save lives.

In order to optimise outbreak response, there has been a plea for the use of real time modelling, which has the potential to shed light on the factors contributing for the occurrence and sustaining of an outbreak (3). The use of computer automation to speed up the collection and integration of data in such circumstances can be a useful tool to advance the understanding of the spread, help monitoring and suggest interventions in situations where taking the appropriate actions in the shortest period of time is crucial (4). However, the collection and integration of datasets in order to quickly construct models to prevent and/or respond to outbreaks face many challenges. Datasets must not only be accessible, but also documented and described in such ways that allows the integration process to occur efficiently.

Current internet-based platforms for rapid tracking and dissemination of information on infectious diseases outbreaks such as ProMED play an important role in aggregating and sharing data from a variety of sources. The Global Public Health Intelligence Network uses an automated web-based system to scan data from a variety of media reports around the world and alert health authorities on the risk of new outbreaks.

The format and structure mailings themselves require human readers to extract and format the relevant information.

Note: Make the case that the potential of this approach will not only benefit investigations of infectious disease outbreaks, but also different groups who want to answer different questions. Give examples of that: Quality of Medicines, Malaria substandard treatment, etc.