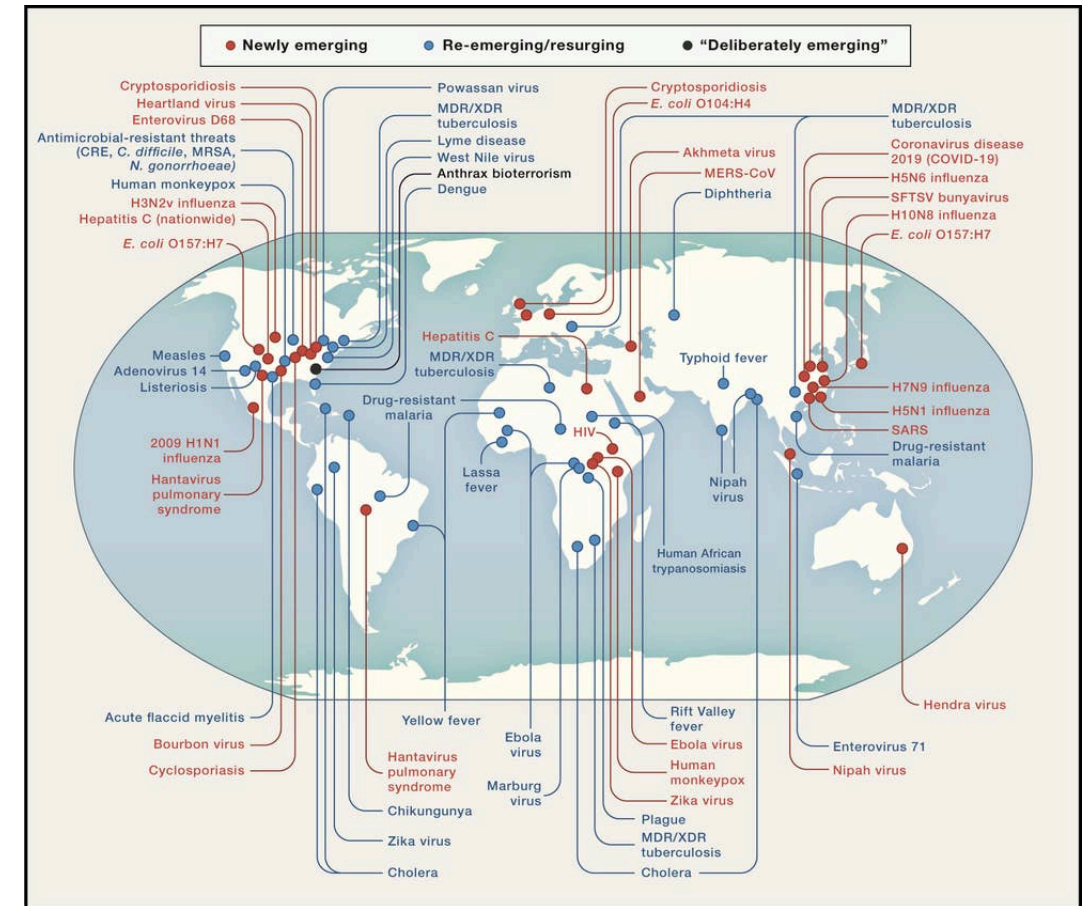# Data Science in the Field

David Kennedy,
Data Scientist, UK-PHRST and Research Fellow, LSHTM

The role of data during an outbreak

We can use data to generate evidence for answering the key epidemiological questions:

- **Who?**
  - Who is the population at risk
- **What?**
  - What is happening to them?
- **Where?**
  - Where are the affected people?
- **When?**
  - When did this begin?

Morens DM, Fauci AS. Emerging Pandemic Diseases: How We Got to COVID-19 [published online ahead of print, 2020 Aug 15]. Cell. 2020;S0092-8674(20)31012-6. doi:10.1016/j.cell.2020.08.021

Lots to consider!

- Is there an existing surveillance system in place? Is it functional?

- What type of event? What do we need to monitor? Is the tool used for the right purpose?

Lots to consider!

- What are the resource implications (costs, human resources) for set up and training?

- How quickly can it be rolled out?

- Is there technical capacity to implement and maintain the system/tool?

- Can it be easily adapted? Does it use open source software?

Lots to consider!

- What are the barriers to getting good data?

- Data security: Where is the data stored? Where is the data sent to?

- Data visualisation: How can the interpretation of the data be communicated to the target audience?

Using data for outbreak research

Our World in Data

Articles by topic

Q Search...

Latest   About   Donate

OXFORD MARTIN SCHOOL   UNIVERSITY OF OXFORD   GCDL

All charts   Sustainable Development Goals Tracker

Coronavirus pandemic: daily updated research and data.   Read more

How experts use data to identify emerging COVID–19 success stories

This is a guest post from epidemiologists David Kennedy, Anna Seale, and Daniel Bausch (UK Public Health Rapid Support Team), alongside Hannah Ritchie and Max Roser (Our World in Data) as part of the Exemplars in Global Health platform.
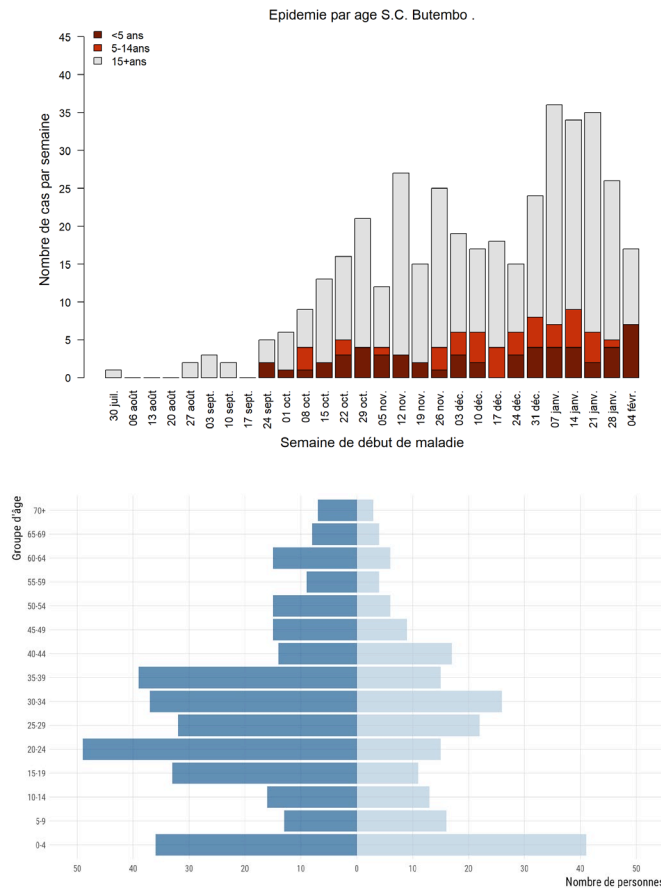
- Work with partners to identify knowledge gaps

- Develop protocols to answer questions

- Analyse a variety of data sources to build up a clear picture

Challenges in data analytics

EVD outbreak, North Kivu, DRC, Aug – Feb 2018
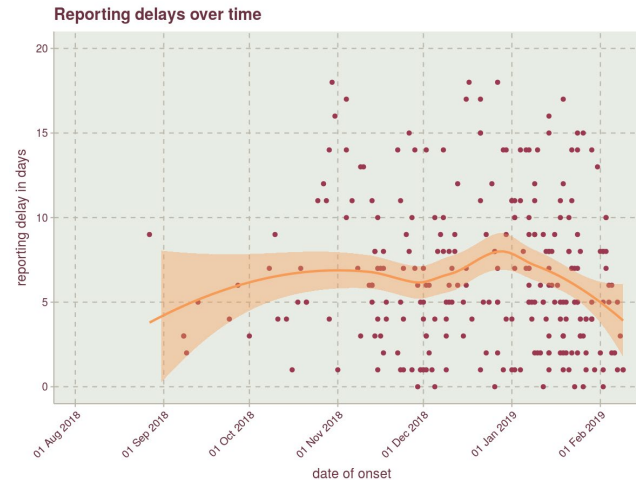
- # Excel, STATA, ArcGIS, R, …

- Often ad-hoc and person-specific

- Sustainability using R remains difficult

- Automation, but do not forget epidemiolgocial interpretation

- Automation: data integration, cleaning
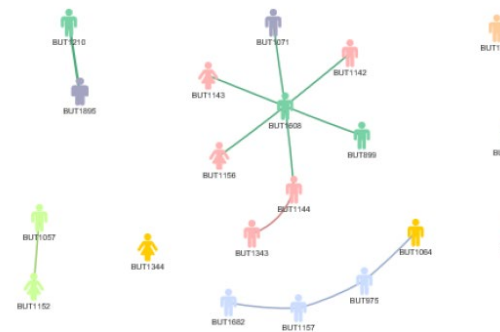
Questions on the roll-out of electronic tools

- Is there an existing surveillance system in place? Is it functional?
- What type of event? What do we need to monitor? Is the tool used for the right purpose?
- What are the resource implications (costs, human resources) for set up and training?
- How quickly can it be rolled out?
- Is there technical capacity to implement and maintain the system/tool?
- Can it be easily adapted? Does it use open source software?
- Can it be easily integrated with other surveillance tools, platforms?
- Data security: Where is the data stored? Where is the data sent to?

Thank you for joining

Data Science in the Field

David Kennedy,
Data Scientist, UK-PHRST and Research Fellow, LSHTM