

RECON

Building the next generation of statistical tools for outbreak response using R

Thibaut Jombart

28th November 2016

Imperial College London
MRC Centre for Outbreak Analysis and Modelling

Outline

1. Lessons learnt from the Ebola outbreak response
2. The R Epidemics Consortium
3. Up-and-coming RECON packages
4. Methodological dialogue during outbreak response

Ebola response

Lessons learnt from the Ebola response



Lessons learnt from the Ebola response



Lessons learnt from the Ebola response

WHO Ebola response team

Help improving situation awareness

EBOLA OUTBREAK

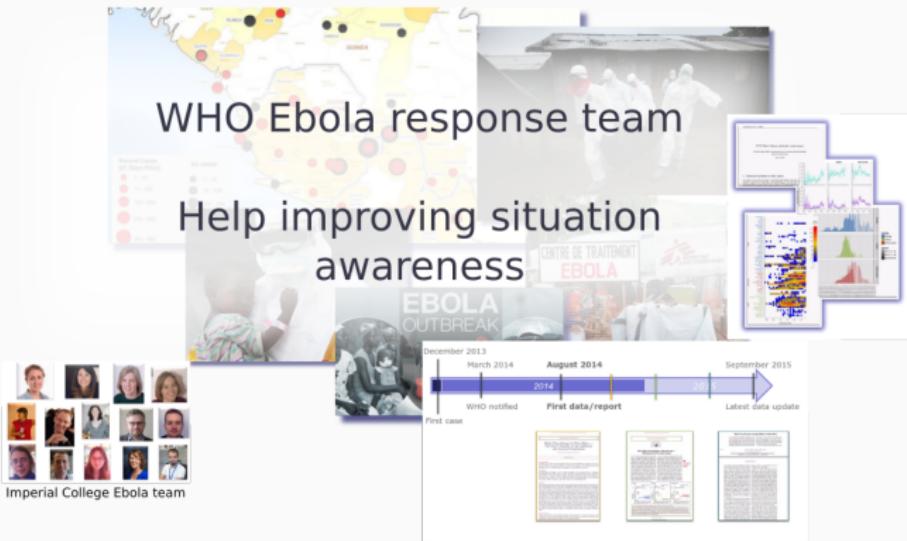
Imperial College Ebola team

December 2013 March 2014 August 2014 September 2015

First case WHO notified First data/report Latest data update

3

Lessons learnt from the Ebola response



Most statistical/modelling tools for situation awareness missing.

What tools do we need?

Some examples:

- **data cleaning:** dictionaries, entry matching
- **graphics:** case incidence in space and time, contact tracing
- **parameter estimation:** key delays, transmissibility
- **estimate / test CFR:** gender, health care workers, treatments effects
- **predictions:** case incidence, mortality, evaluate interventions
- **report:** (semi-)automated situation reports

Who do we need to develop these tools?



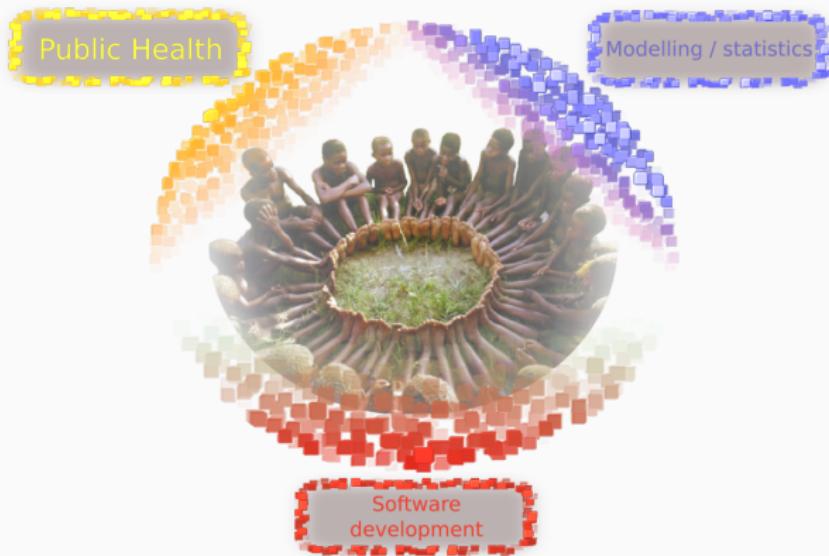
Who do we need to develop these tools?



Who do we need to develop these tools?



Who do we need to develop these tools?



The R Epidemics Consortium

Hackout 3: a hackathon for emergency outbreak response

Last summer at the *rOpenSci* headquarters (Berkeley)



CDC
CENTERS FOR DISEASE
CONTROL AND PREVENTION

World Health Organization

MÉDECINS SANS FRONTIÈRES
DOCTORS WITHOUT BORDERS

Public Health
England

Public Health
Agency of Canada

Centre for
Outbreak Analysis
and Modelling

National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

Imperial College
London

UNIVERSITY OF
CAMBRIDGE

UNIVERSITY OF
OXFORD

LONDON SCHOOL of
HYGIENE & TROPICAL
MEDICINE

BERKELEY
Institute for
Data Science

welcome trust
sanger
institute

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Hackout 3: from ideas to projects to...



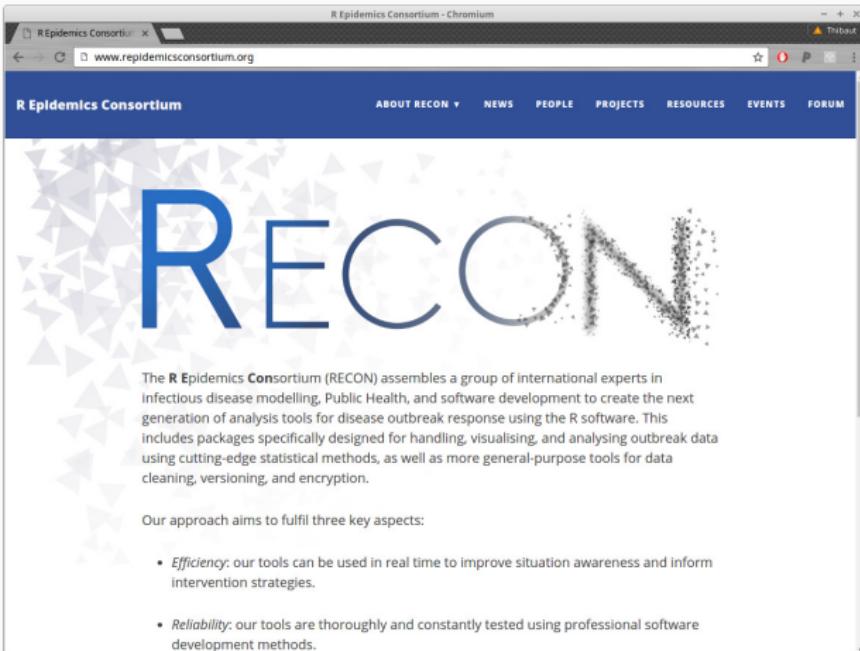
Hackout 3: from ideas to projects to...



How do we keep momentum once the event is over?

RECON: the R Epidemics Consortium

A taskforce to build a new generation of outbreak response tools in  .



The screenshot shows a web browser window displaying the "R Epidemics Consortium" website. The title bar reads "R Epidemics Consortium - Chromium". The address bar shows the URL "www.repidemicsconsortium.org". The page header includes the "R Epidemics Consortium" logo and navigation links for "ABOUT RECON", "NEWS", "PEOPLE", "PROJECTS", "RESOURCES", "EVENTS", and "FORUM". The main visual is a large, stylized word "RECON" where each letter is composed of numerous small, dark grey triangles, set against a background of a similar triangular pattern. Below the graphic, a text block provides a brief description of the consortium's purpose and activities. A section titled "Our approach aims to fulfil three key aspects:" lists three bullet points: "Efficiency: our tools can be used in real time to improve situation awareness and inform intervention strategies.", and "Reliability: our tools are thoroughly and constantly tested using professional software development methods.".

The R Epidemics Consortium (RECON) assembles a group of international experts in infectious disease modelling, Public Health, and software development to create the next generation of analysis tools for disease outbreak response using the R software. This includes packages specifically designed for handling, visualising, and analysing outbreak data using cutting-edge statistical methods, as well as more general-purpose tools for data cleaning, versioning, and encryption.

Our approach aims to fulfil three key aspects:

- *Efficiency*: our tools can be used in real time to improve situation awareness and inform intervention strategies.
- *Reliability*: our tools are thoroughly and constantly tested using professional software development methods.

www.repidemicsconsortium.org

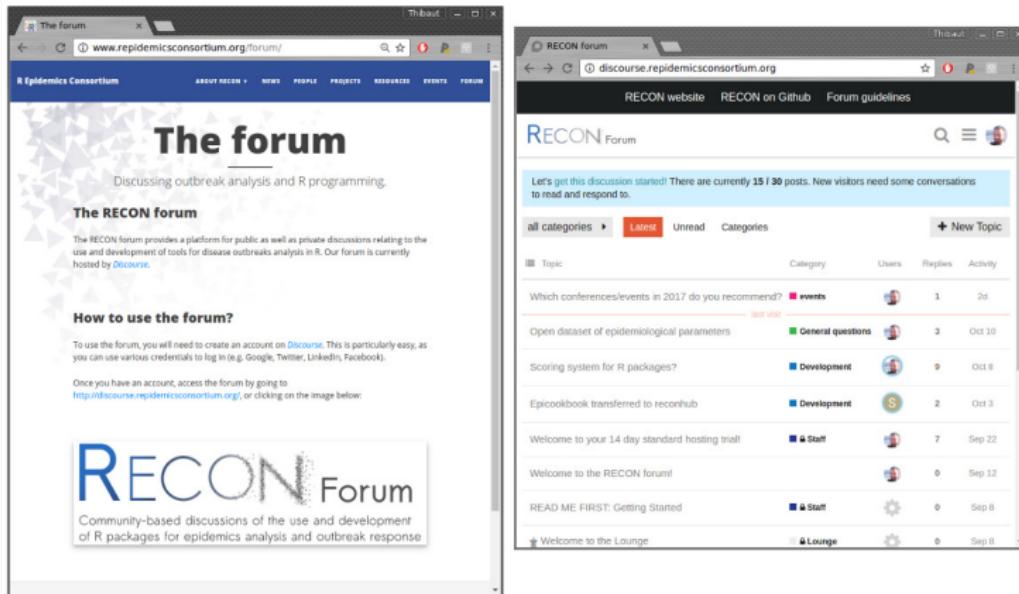
RECON

www.repidemicsconsortium.org

- started 6th September 2016
- 46 people (41 members, 5 board)
- 10 countries, > 20 institutions
- ~ 10 new packages coming
- **public forum**, blog, online resources

The RECON forum

A platform for discussing epidemics analysis in  .



The image shows two side-by-side browser windows. The left window displays the official RECON forum website at www.repidemicsconsortium.org/forum/. It features a dark blue header with navigation links for 'ABOUT RECON', 'NEWS', 'PEOPLE', 'PROJECTS', 'RESOURCES', 'EVENTS', and 'FORUM'. Below the header is a large banner with the text 'The forum' and 'Discussing outbreak analysis and R programming.' A section titled 'The RECON forum' explains the purpose of the forum and mentions it is hosted by Discourse. Another section, 'How to use the forum?', provides instructions for creating an account. At the bottom is a logo for 'RECON Forum' with the subtitle 'Community-based discussions of the use and development of R packages for epidemics analysis and outbreak response'.

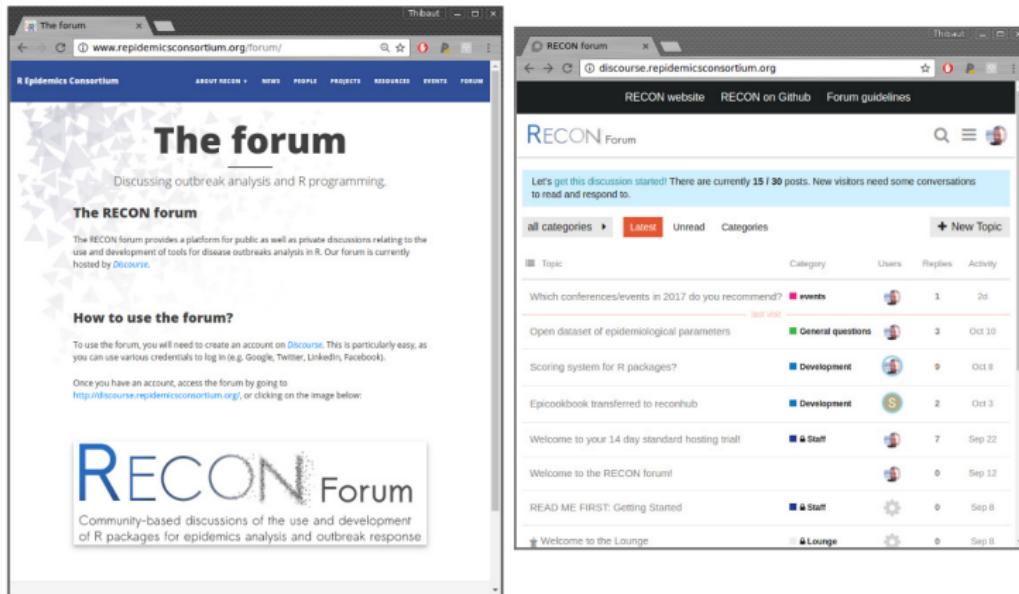
The right window shows the Discourse instance at discourse.repidemicsconsortium.org. It has a similar header with links for 'RECON website', 'RECON on Github', and 'Forum guidelines'. The main content area is titled 'RECON Forum' and includes a search bar and user profile icons. A message encourages users to start conversations. Below is a list of forum topics:

Topic	Category	Users	Replies	Activity
Which conferences/events in 2017 do you recommend?	events	1	2d	Oct 10
Open dataset of epidemiological parameters	General questions	3	Oct 10	
Scoring system for R packages?	Development	9	Oct 8	
Epicookbook transferred to reconhub	Development	2	Oct 3	
Welcome to your 14 day standard hosting trial!	Staff	7	Sep 22	
Welcome to the RECON forum!	Staff	0	Sep 12	
READ ME FIRST: Getting Started	Staff	0	Sep 8	
Welcome to the Lounge	Lounge	0	Sep 8	

www.repidemicsconsortium.org/forum

The RECON forum

A platform for discussing epidemics analysis in  .



The image shows two side-by-side browser windows. The left window displays the official RECON forum website at www.repidemicsconsortium.org/forum/. It features a dark blue header with the RECON logo and navigation links for About RECON, News, People, Projects, Resources, Events, and Forum. The main content area has a light gray background with a geometric pattern and contains sections for 'The forum' (Discussing outbreak analysis and R programming), 'The RECON forum' (describing its purpose and hosting on Discourse), and 'How to use the forum?' (instructions for creating an account). A large 'RECON Forum' logo is at the bottom. The right window shows the Discourse instance at discourse.repidemicsconsortium.org. It has a white header with links to the RECON website, GitHub, and Forum guidelines. The main area is titled 'RECON Forum' and shows a list of topics. Topics include recommendations for conferences/events (1 event), an open dataset of epidemiological parameters, a scoring system for R packages, the transfer of the Epidicookbook to recohub, a welcome message for a 14-day hosting trial, a welcome to the forum, a 'READ ME FIRST: Getting Started' guide, and a 'Welcome to the Lounge'. The interface includes a search bar, user icons, and activity counts.

www.repidemicsconsortium.org/forum/

Join us!

RECON package: what do we aim for?

- efficiency: useful for improving situation awareness in real time; **cutting-edge, computer-efficient statistical methods**

RECON package: what do we aim for?

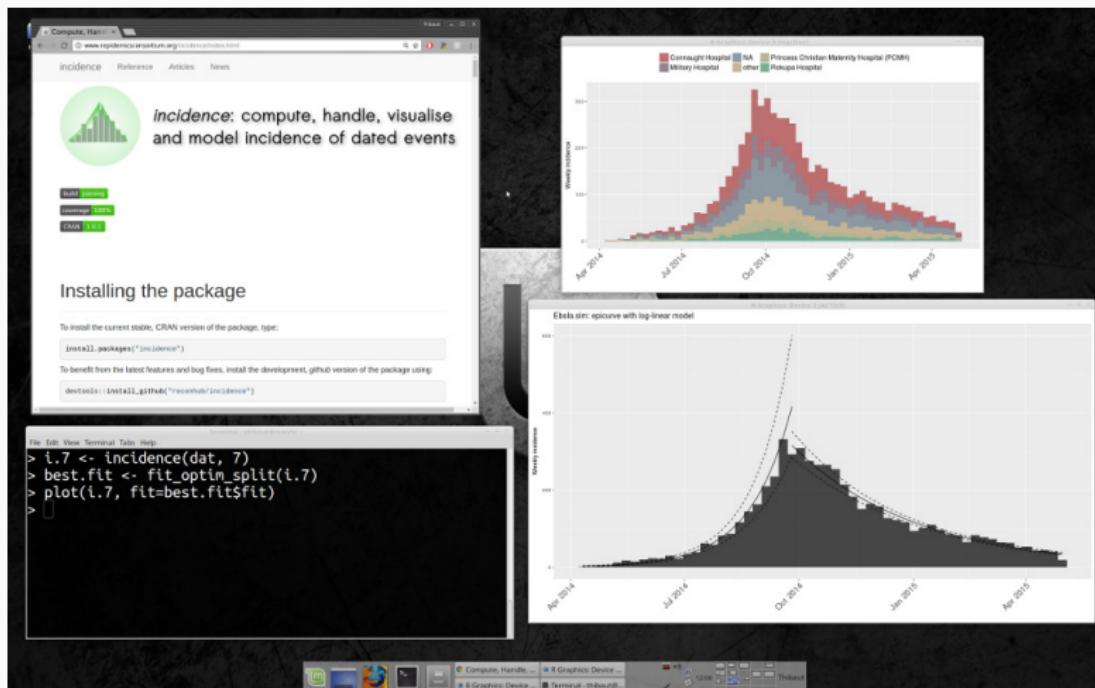
- **efficiency**: useful for improving situation awareness in real time; **cutting-edge, computer-efficient statistical methods**
- **reliability**: outputs can be trusted; **continuous integration, extensive unit testing, code review, good practices**

RECON package: what do we aim for?

- **efficiency**: useful for improving situation awareness in real time; **cutting-edge, computer-efficient statistical methods**
- **reliability**: outputs can be trusted; **continuous integration, extensive unit testing, code review, good practices**
- **accessibility**: widely available, easy learning curve; **extensive documentation, tutorials, websites, forum**

Up-and-coming RECON packages

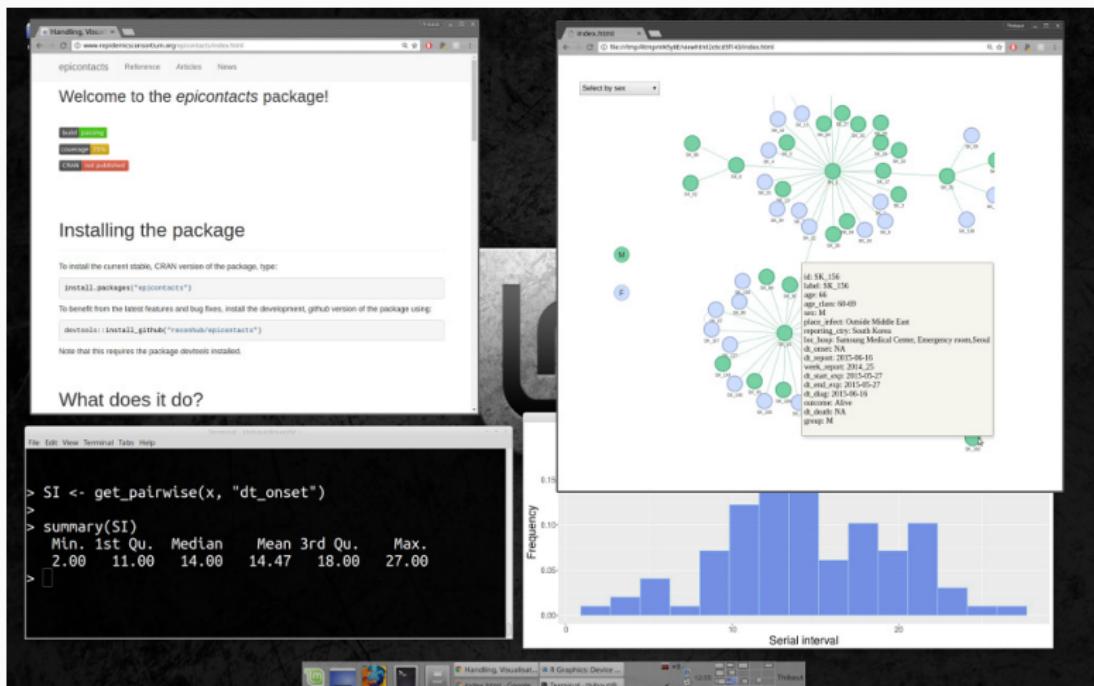
incidence: computation, handling, visualisation and modelling of epicurves



www.repidemicsconsortium.org/incidence

[released]

epicontacts: handling, visualisation and analysis of epidemiological contacts



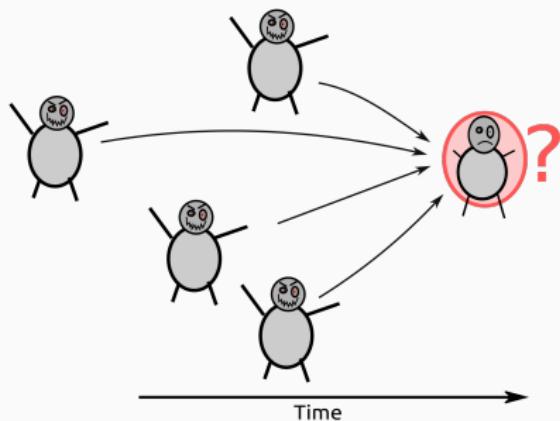
www.repidemicsconsortium.org/epicontacts

[release December 2016]

outbreaker2: inferring who infects whom in an outbreak

Original *outbreaker* model: timing of symptoms and pathogen genomes to infer infectors

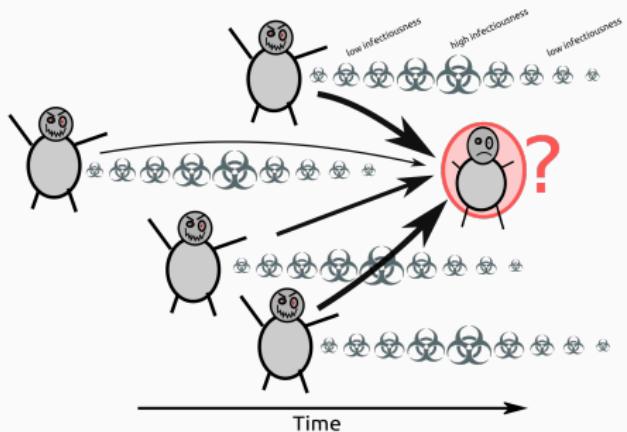
(Jombart *et al*, PLoS Comp Biol, 2014)



outbreaker2: inferring who infects whom in an outbreak

Original *outbreaker* model: timing of symptoms and pathogen genomes to infer infectors

(Jombart et al, PLoS Comp Biol, 2014)

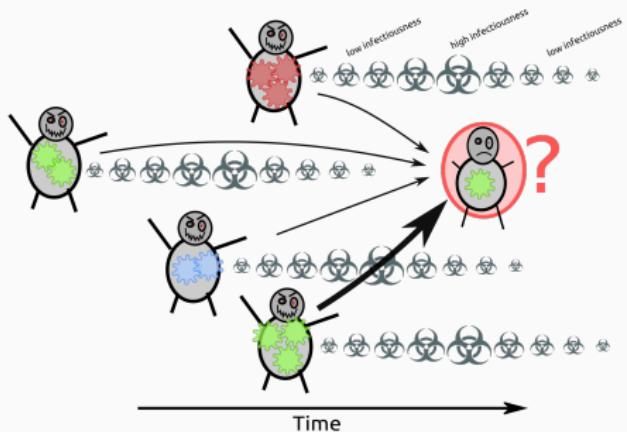


Since *outbreaker*: new models, data, and questions.

outbreaker2: inferring who infects whom in an outbreak

Original *outbreaker* model: timing of symptoms and pathogen genomes to infer infectors

(Jombart et al, PLoS Comp Biol, 2014)



Since *outbreaker*: new models, data, and questions.

But: methodological niche fragmented.

Are different methods really... different?

Are different methods really... different?



Are different methods really... different?



Different models can lead to very similar implementations.
Can we find a **general formulation**?

What do these model look like?

- a, b, c : different types of data
- θ : parameters / augmented data

Data are often assumed to be *conditionally independent*:

$$p(a, b, c|\theta) = p(a|\theta)p(b|\theta)p(c|\theta)$$

What do these model look like?

- a, b, c : different types of data
- θ : parameters / augmented data

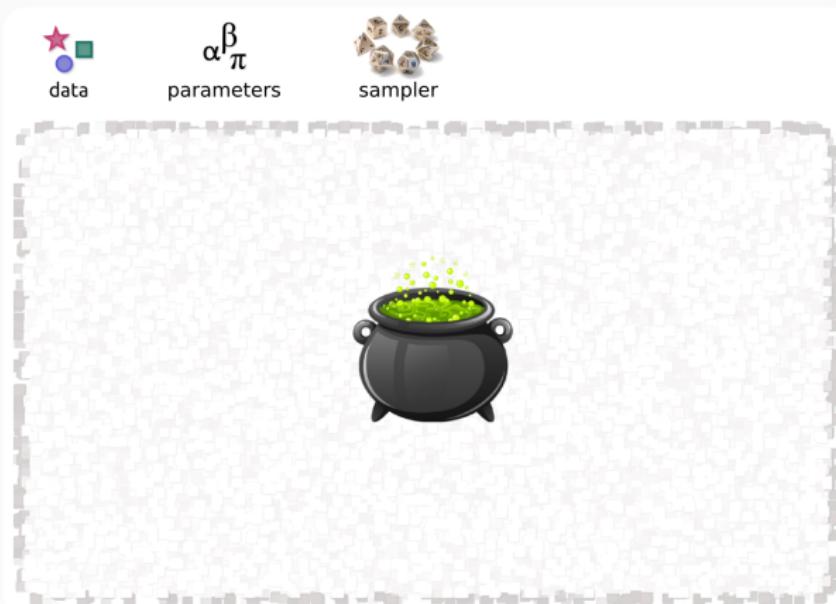
Data are often assumed to be *conditionally independent*:

$$p(a, b, c|\theta) = p(a|\theta)p(b|\theta)p(c|\theta)$$

Components can be treated as **plugins**.

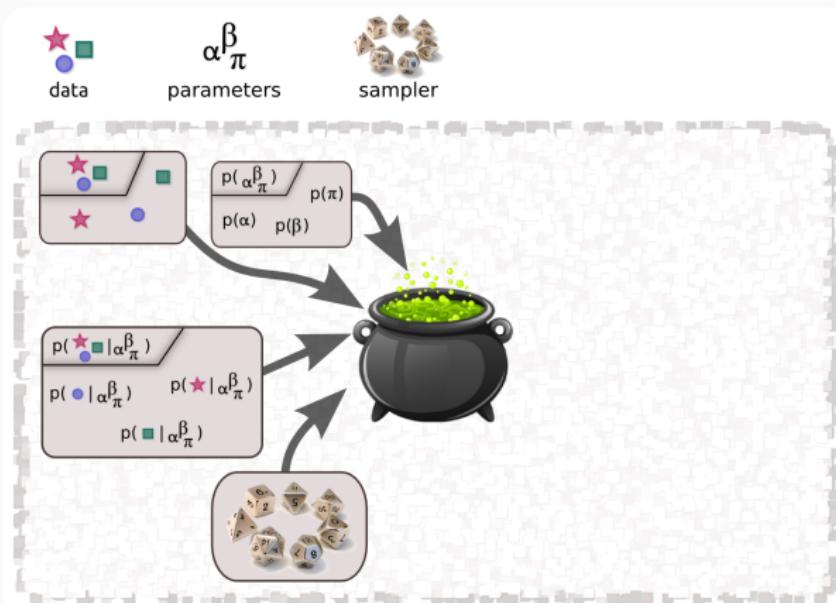
outbreaker2: a general cauldron for cooking methods

Use-your-own: data type, likelihood, prior, MCMC.



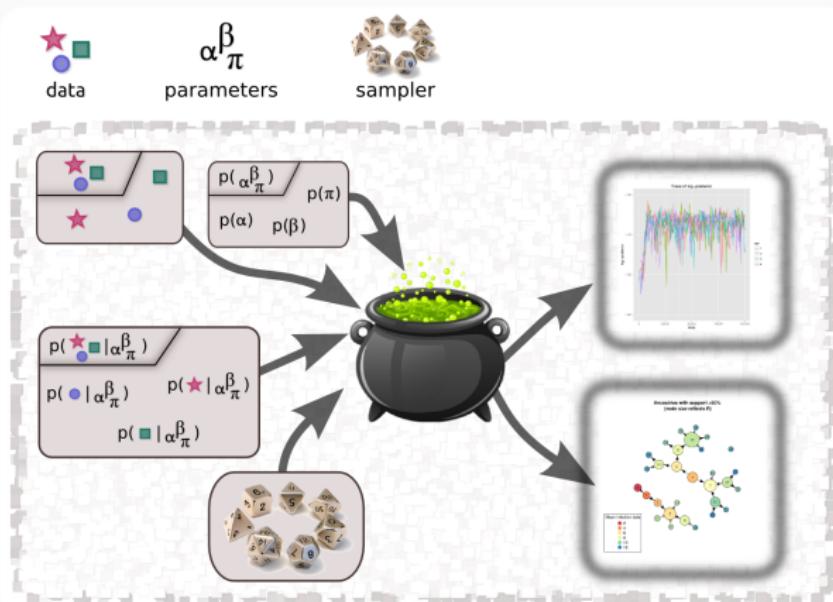
outbreaker2: a general cauldron for cooking methods

Use-your-own: data type, likelihood, prior, MCMC.



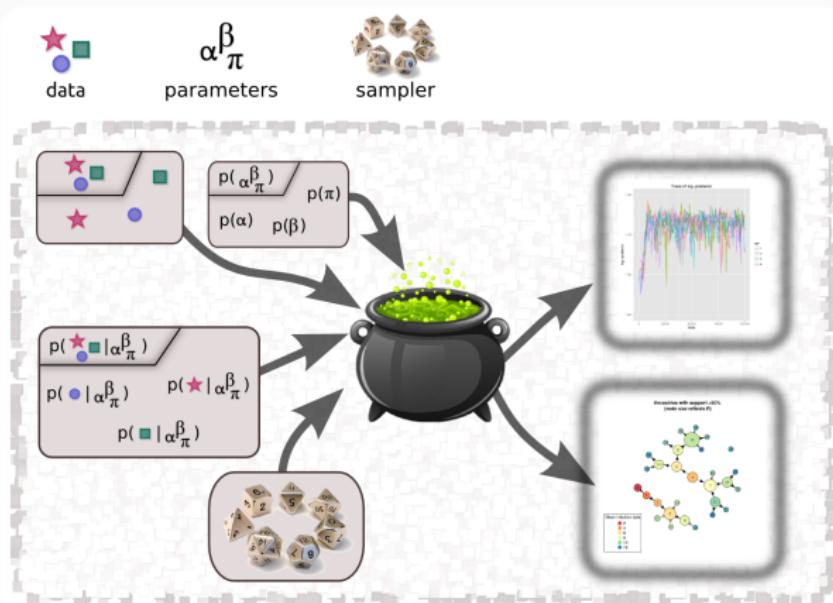
outbreaker2: a general cauldron for cooking methods

Use-your-own: data type, likelihood, prior, MCMC.



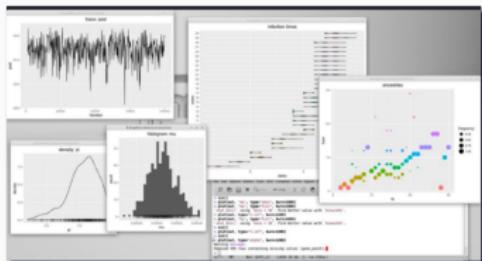
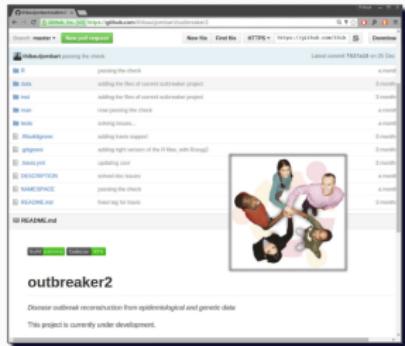
outbreaker2: a general cauldron for cooking methods

Use-your-own: data type, likelihood, prior, MCMC.

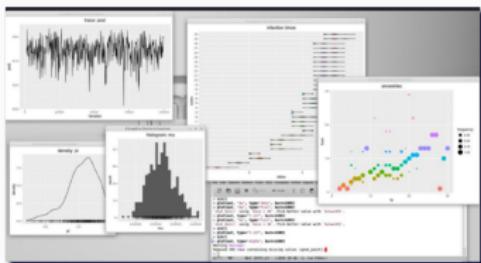
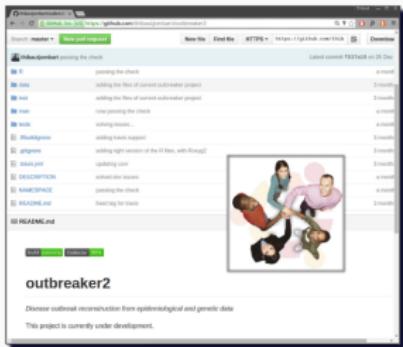


Modularity is key to generalising approaches

outbreaker2: a general tool for outbreak reconstruction



outbreaker2: a general tool for outbreak reconstruction



- **modularity:** customise model and sampler, optional Rcpp
- **reliability:** continuous integration, extensive unit testing (aiming for 100% coverage)
- **prettier:** plot methods using *ggplot2*, interactive networks visualisation
- enable **contributions from the community**
- release planned for early 2017

Methodological dialogue

Methodological development relies on an interdisciplinary dialogue

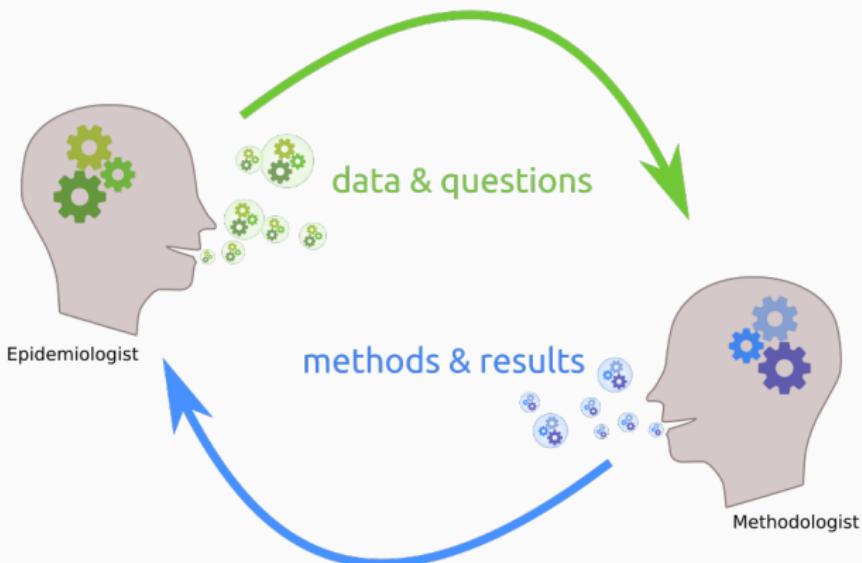


Epidemiologist

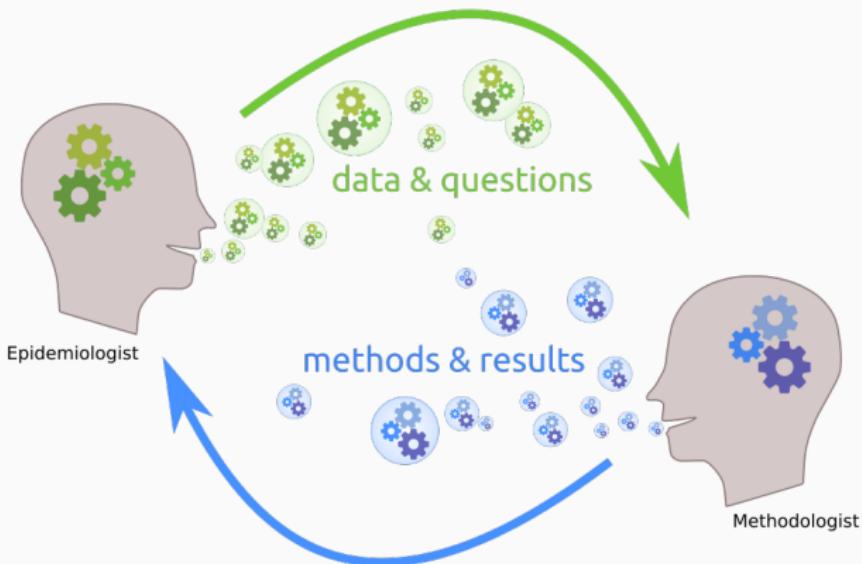


Methodologist

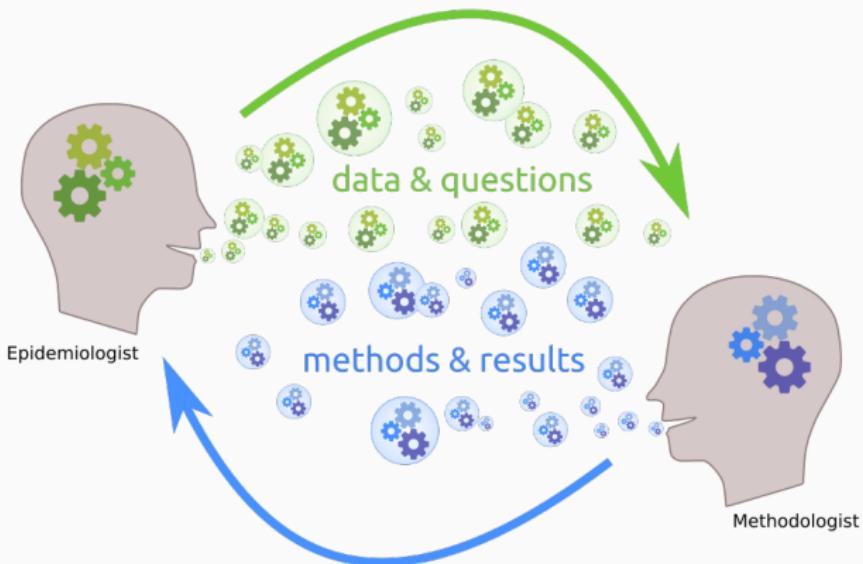
Methodological development relies on an interdisciplinary dialogue



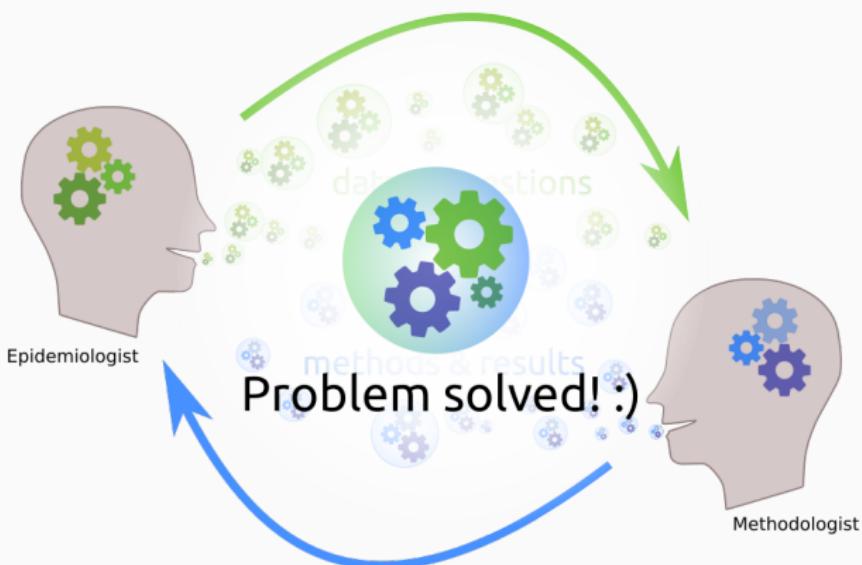
Methodological development relies on an interdisciplinary dialogue



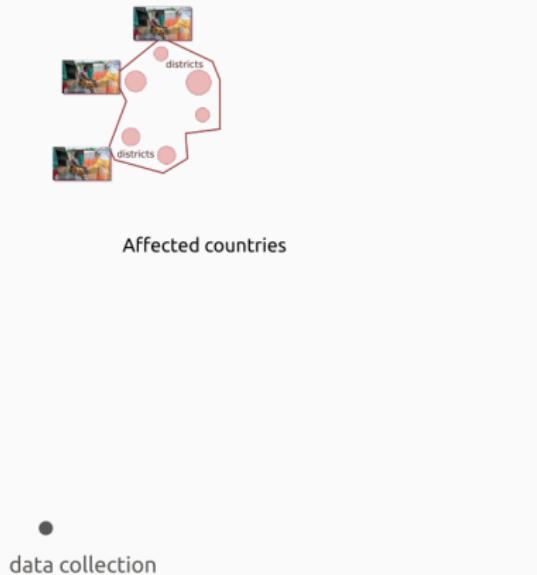
Methodological development relies on an interdisciplinary dialogue



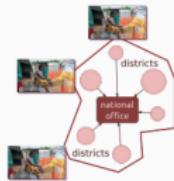
Methodological development relies on an interdisciplinary dialogue



Outbreak response context creates distance and delays



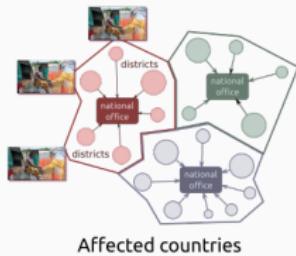
Outbreak response context creates distance and delays



Affected countries

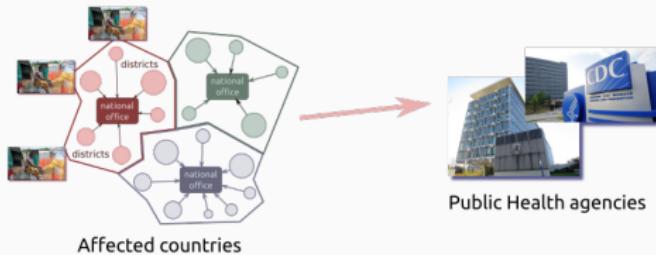


Outbreak response context creates distance and delays

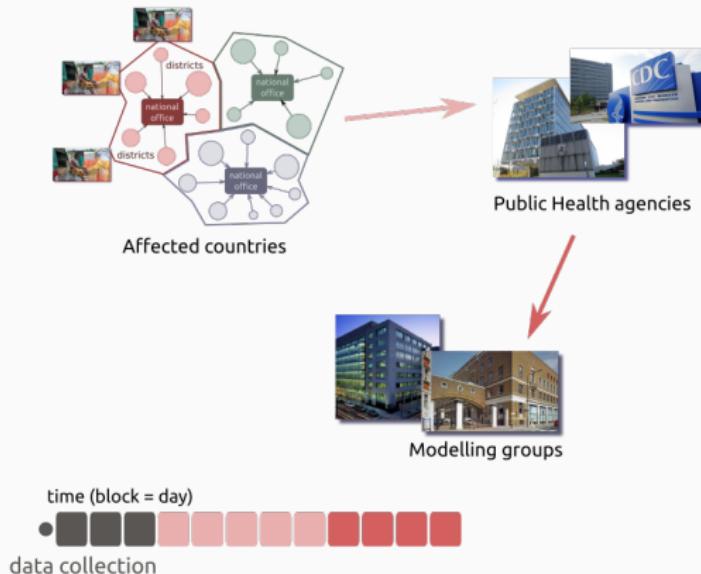


time (block = day)
• data collection

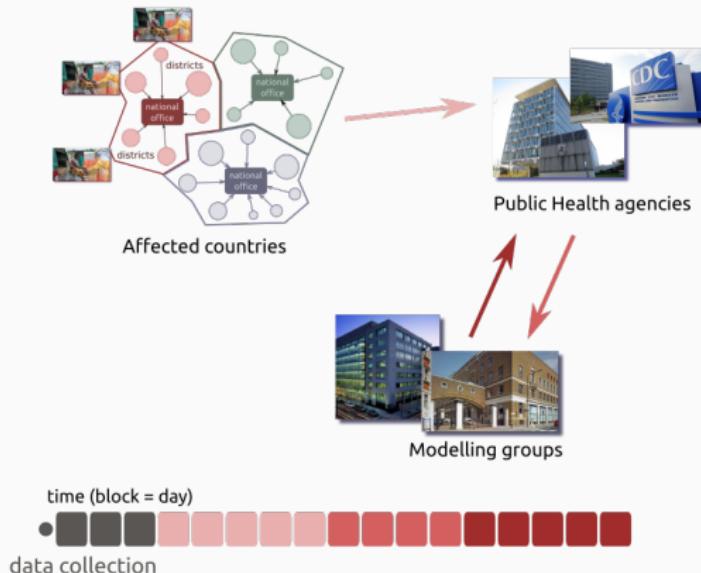
Outbreak response context creates distance and delays



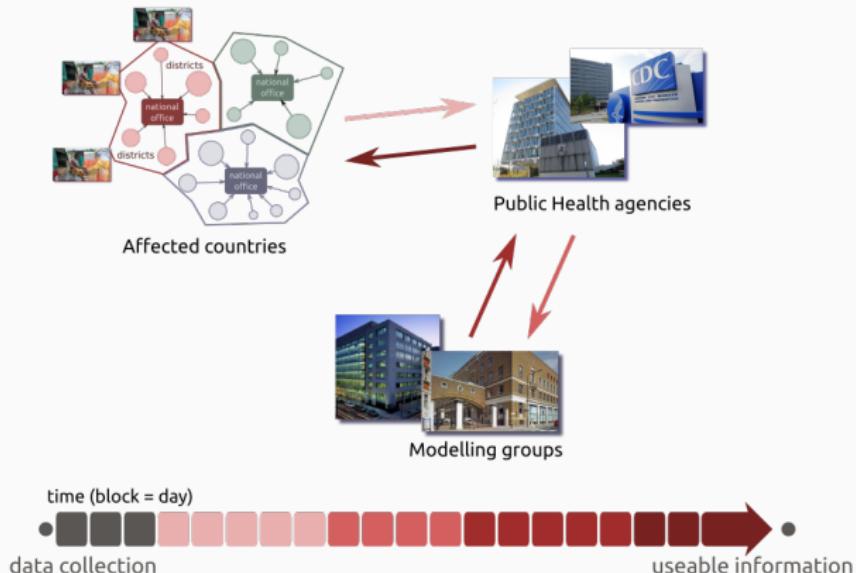
Outbreak response context creates distance and delays



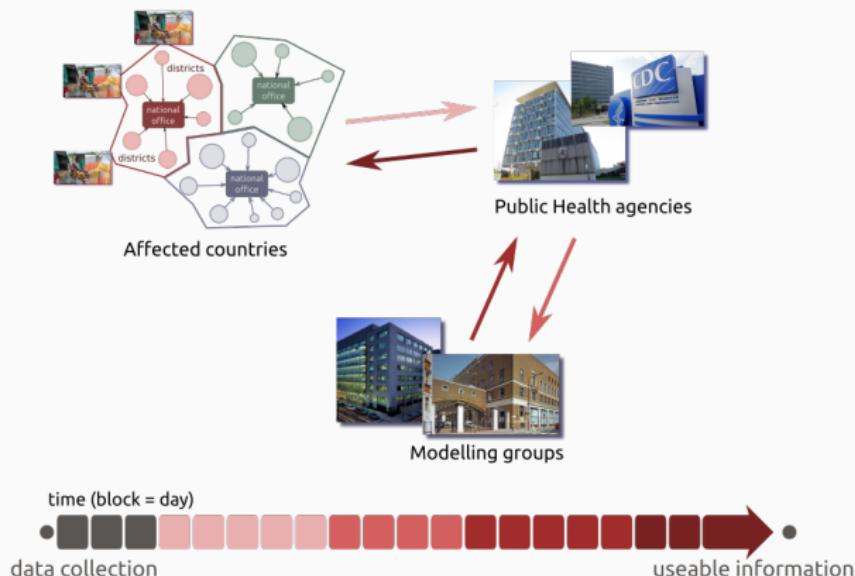
Outbreak response context creates distance and delays



Outbreak response context creates distance and delays

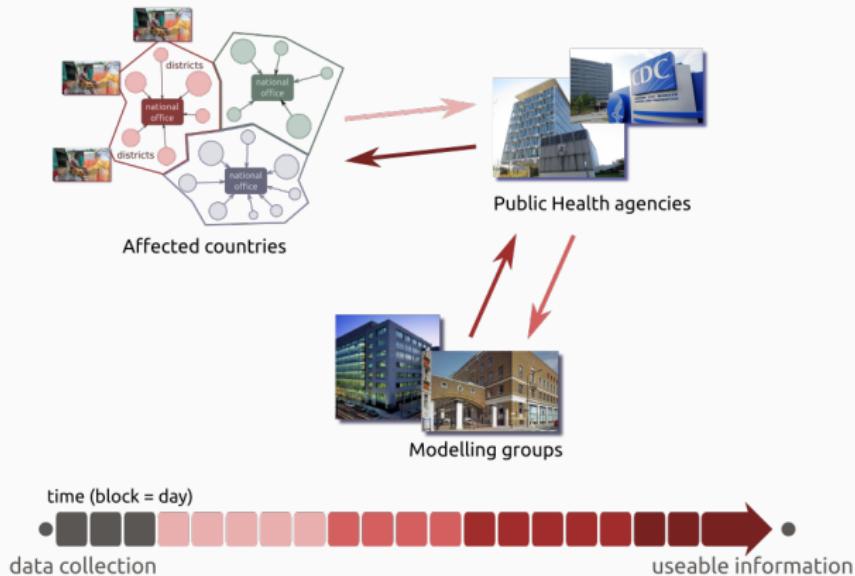


Outbreak response context creates distance and delays



- efficient tools can shorten delays

Outbreak response context creates distance and delays



- efficient tools can shorten delays
- potential of **embedding methodologists in outbreak response teams**

Thanks to...

- **ESCAIDE** organisers
- **Imperial College:** Neil Ferguson, Rich Fitzjohn, Anne Cori, Finlay Campbell, Evgenia Markvardt, James Hayward
- **UC Berkeley:** Karthik Ram
- **Groups:** WHO Ebola Response Team, Hackout 1/2/3, RECON members
- **funding:** HPRU-NIHR, MRC

More on:

www.repidemicsconsortium.org

Questions?