

RECON

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

Thibaut Jombart

28th November 2016

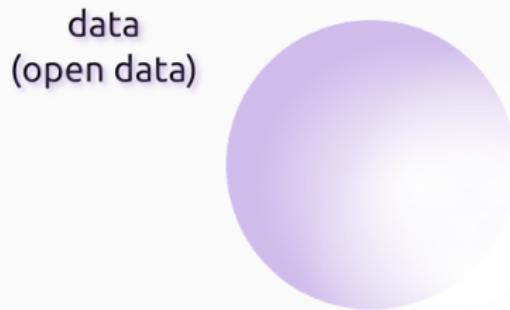
Imperial College London
MRC Centre for Outbreak Analysis and Modelling

Reproducible / open science...

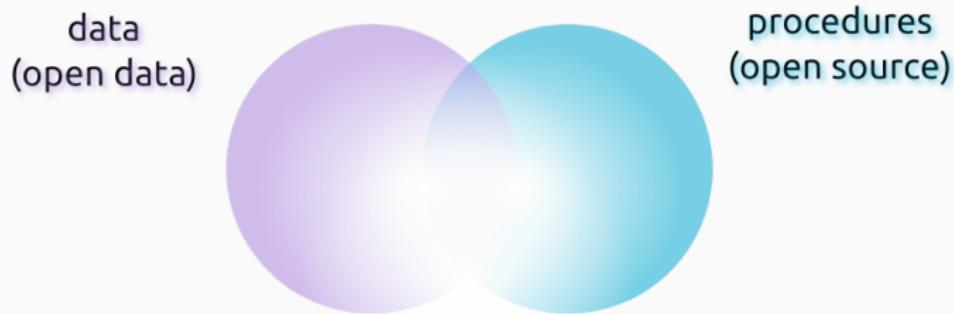
Reproducible / open science...

It's just science!

What do we need for 'reproducible' science?



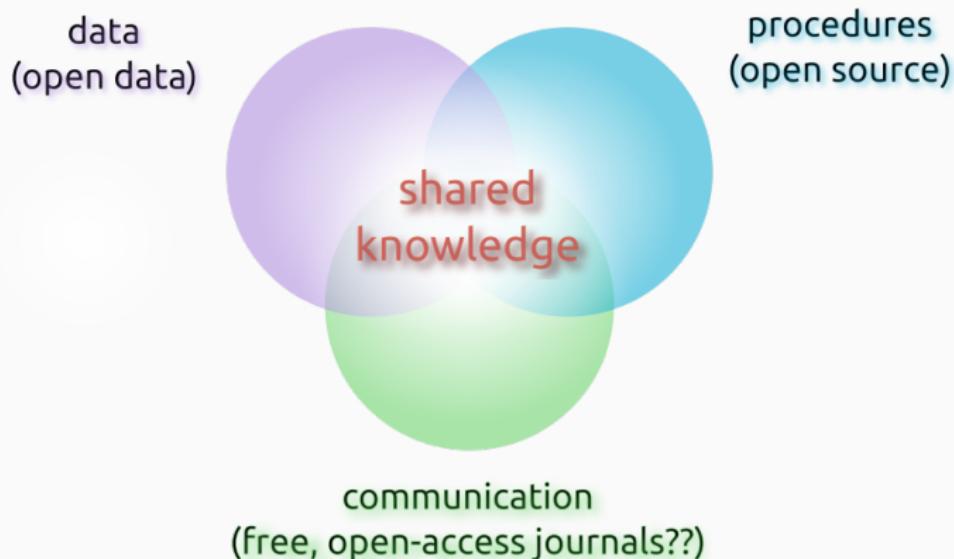
What do we need for 'reproducible' science?



What do we need for 'reproducible' science?



What do we need for 'reproducible' science?

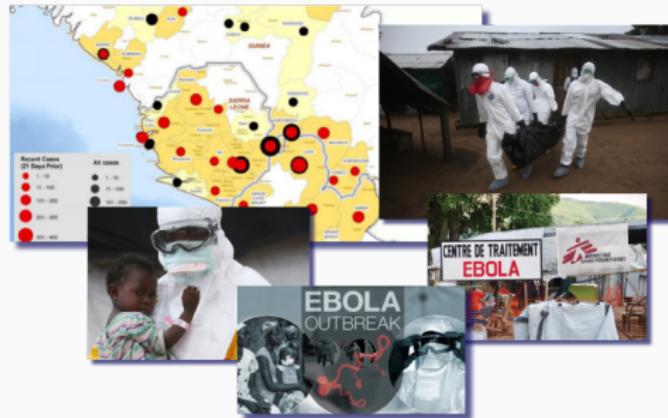


Outline

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



Imperial College Ebola team

Lessons learnt from the Ebola response

WHO Ebola response team

Help improving situation awareness

EBOLA OUTBREAK

December 2013 March 2014 August 2014 September 2015

First case WHO notified First data/report Latest data update

Imperial College Ebola team

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?



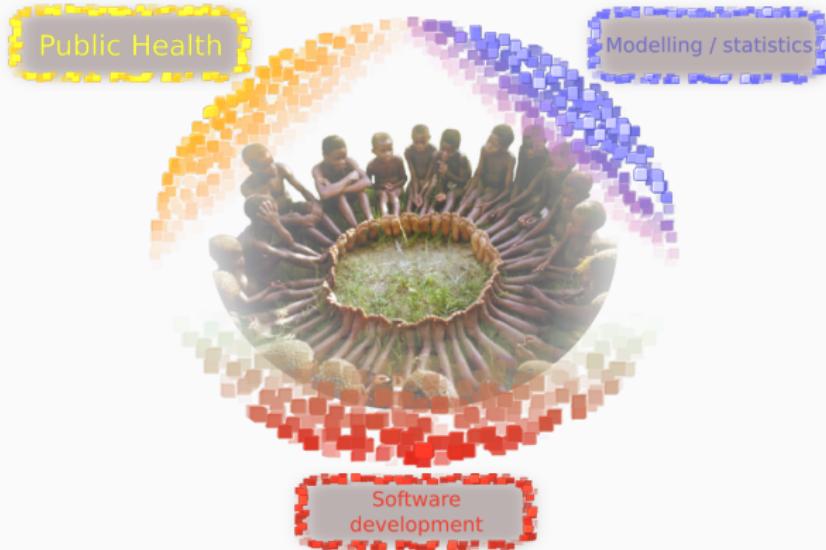
Who do we need to develop these tools?



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)



The image is a collage of logos and photographs related to the hackathon. At the top left is the CDC logo. Next to it is the World Health Organization logo. To the right of that is the Médecins Sans Frontières (Doctors Without Borders) logo. Further right is the Public Health Agency of Canada logo. Below these are logos for the MRC (Centre for Outbreak Analysis and Modelling), the National Institute for Public Health and the Environment (RIVM), Imperial College London, the University of Cambridge, the University of Oxford, and the London School of Hygiene & Tropical Medicine. On the right side are logos for the ECDC (European Centre for Disease Prevention and Control), the Berkeley Institute for Data Science, and the Wellcome Trust Sanger Institute. In the center, there is a large group photograph of approximately 30 people standing in front of a white building. Below this group photo are three smaller photographs showing people working at desks in an office environment.

Hackout 3: from ideas to projects to...



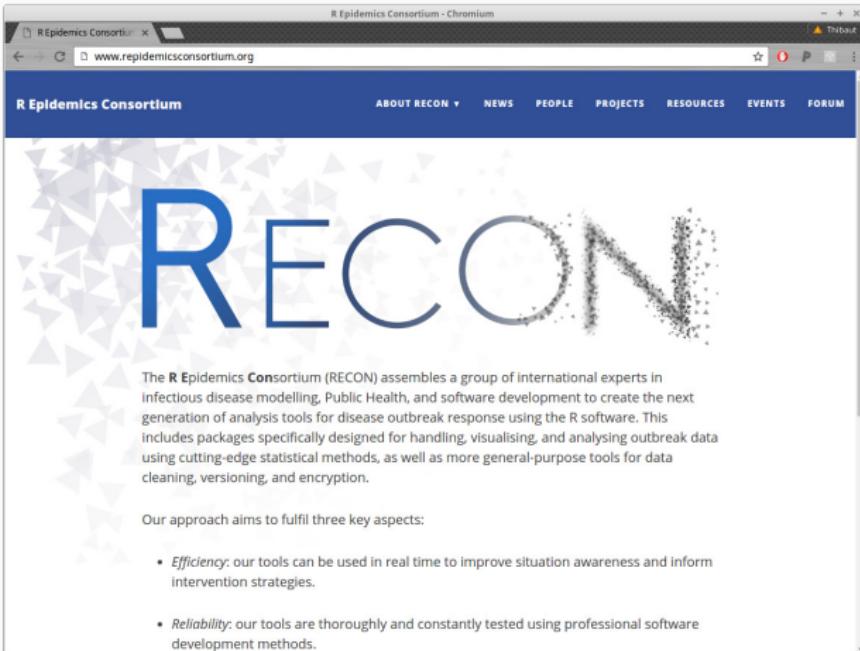
Hackout 3: from ideas to projects to...



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 the homepage of the RECON website. The header features the title "R Epidemics Consortium" and a navigation menu with links for "ABOUT RECON", "NEWS", "PEOPLE", "PROJECTS", "RESOURCES", "EVENTS", and "FORUM". The main visual is a large, stylized word "RECON" composed of numerous small, dark grey triangles, set against a background of a similar geometric pattern. Below the title, a paragraph describes the consortium's mission: "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." 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."

www.repidemicsconsortium.org

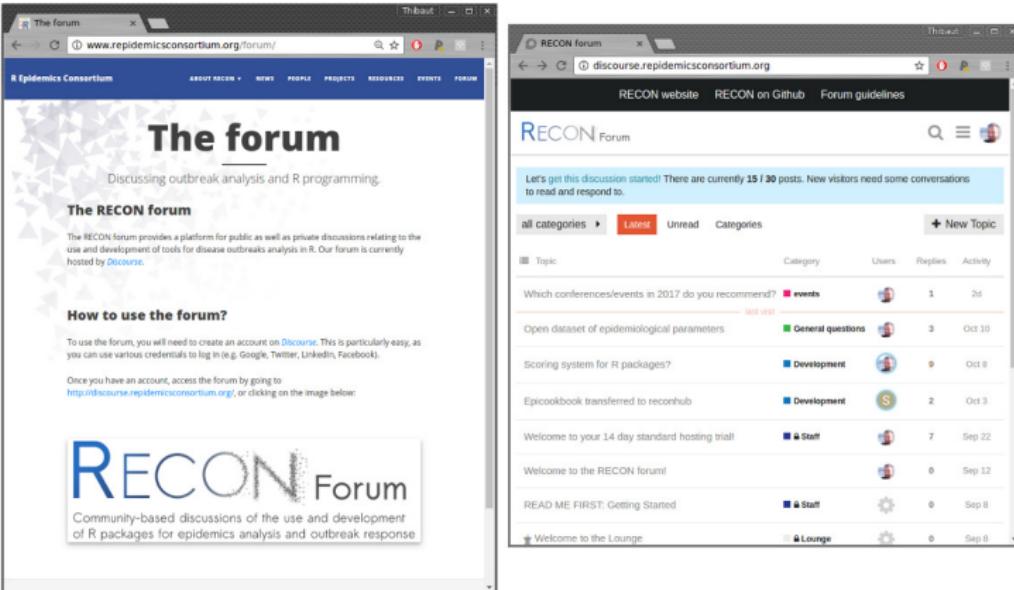
RECON

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 .



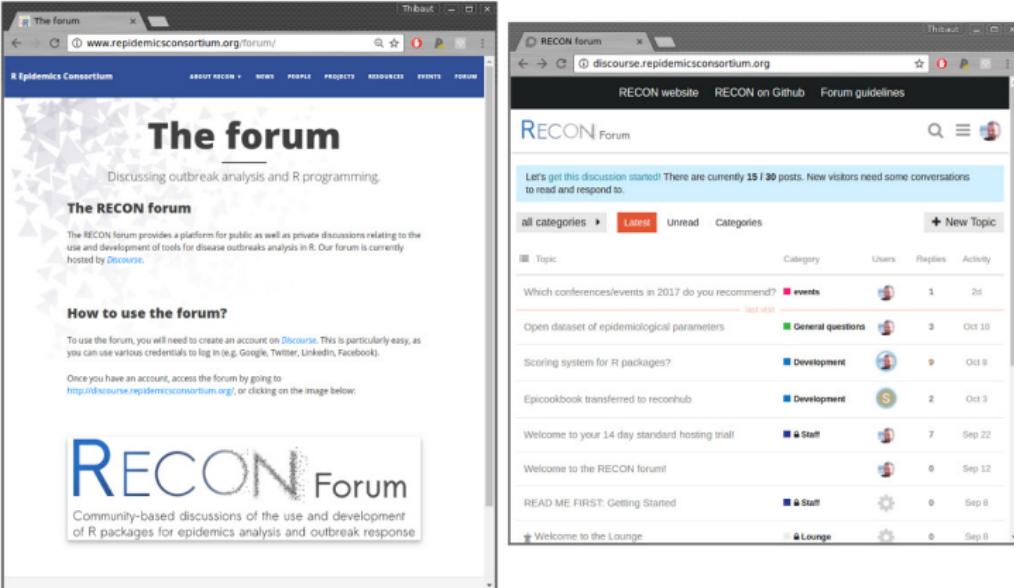
The image shows two side-by-side browser windows. The left window displays the official RECON forum at www.repidemicsconsortium.org/forum/. It features a dark blue header with the 'Epidemics Consortium' logo and navigation links for About RECON, News, People, Projects, Resources, Events, and Forum. The main content area has a light gray background with a geometric pattern and contains the heading 'The forum' and a sub-section titled 'The RECON forum'. The right window shows a discourse instance at discourse.repidemicsconsortium.org. The header includes links to the RECON website, GitHub, and Forum guidelines. The main content area is titled 'RECON Forum' and shows a list of forum topics. The topics listed are:

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 9	
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 RECON forum provides a platform for public as well as private discussions relating to the use and development of tools for disease outbreaks analysis in R. Our forum is currently hosted by [Discourse](#).

To use the forum, you will need to create an account on [Discourse](#). This is particularly easy, as you can use various credentials to log in (e.g. Google, Twitter, LinkedIn, Facebook).

Once you have an account, access the forum by going to <http://discourse.repidemicsconsortium.org/>, or clicking on the image below:


Community-based discussions of the use and development of R packages for epidemics analysis and outbreak response

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?

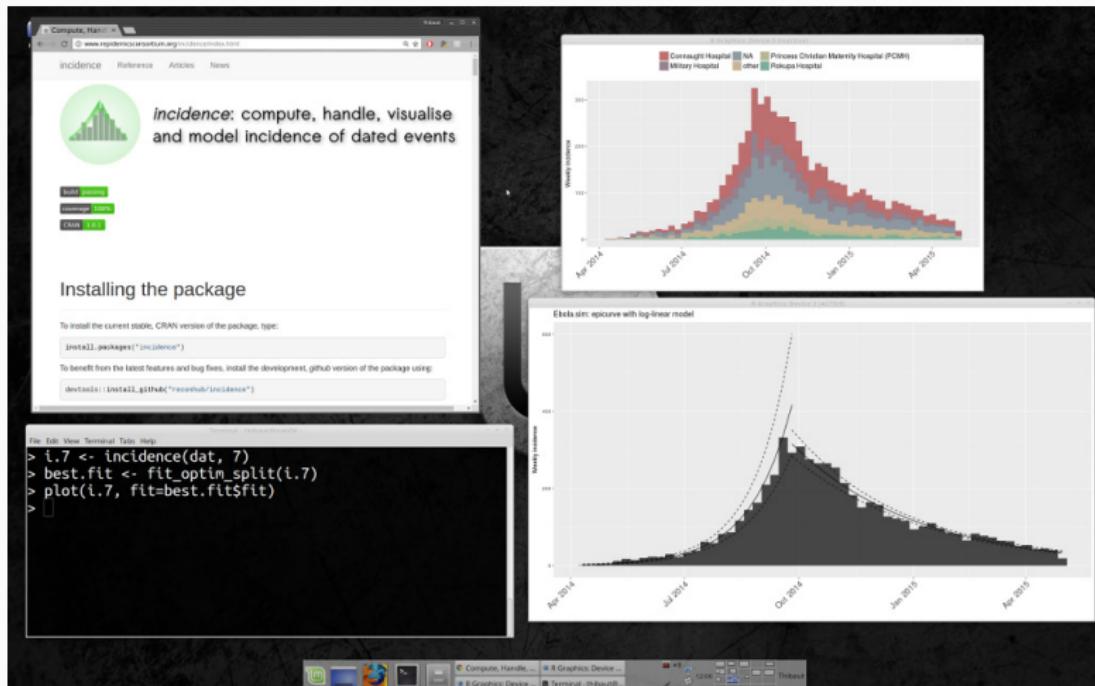
- **efficiency**: useful for improving situation awareness in real time; **cutting-edge, computer-efficient statistical methods**
- **reliability**: outputs can be trusted; **continuous integration, extensive unit testing, code review, good practices**

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

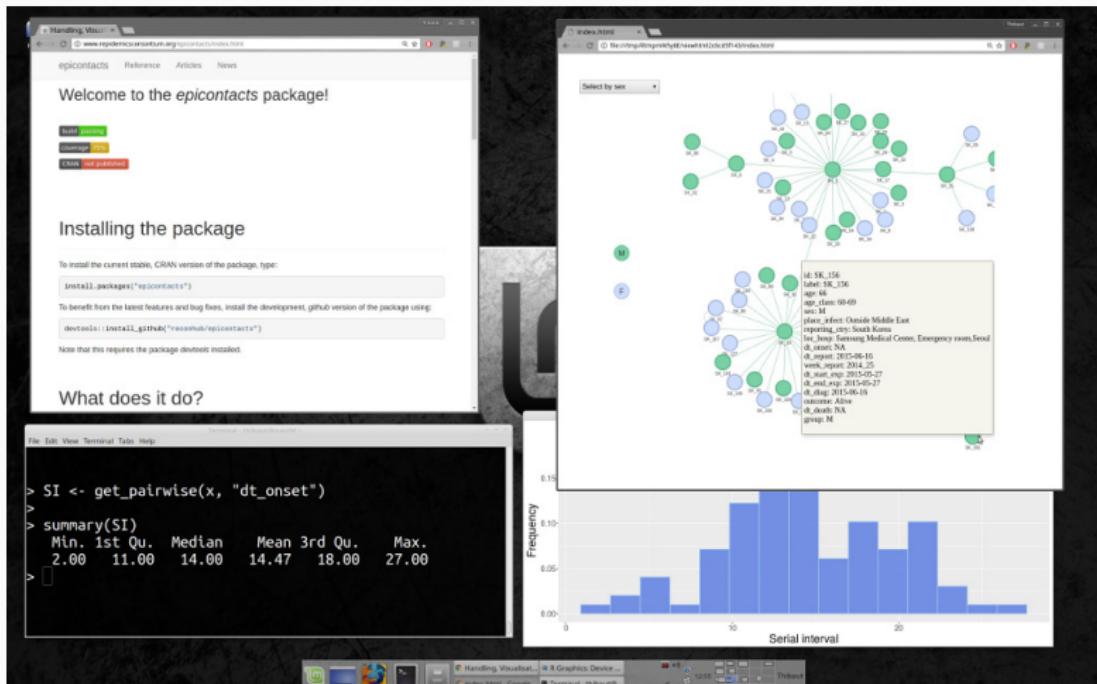
incidence: computation, handling, visualisation and modelling of epicurves



www.repidemicsconsortium.org/incidence

[released]

epicontacts: handling, visualisation and analysis of epidemiological contacts



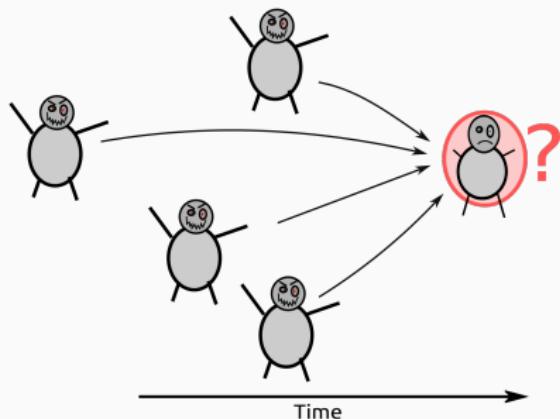
www.repidemicsconsortium.org/epicontacts

[release December 2016]

outbreaker2: inferring who infects whom in an outbreak

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

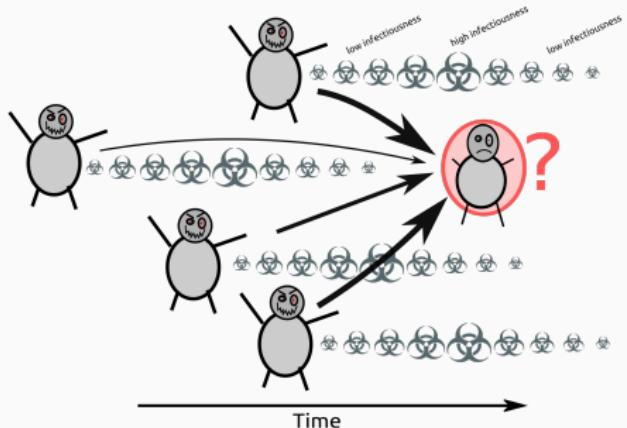
(Jombart et al, PLoS Comp Biol, 2014)



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)

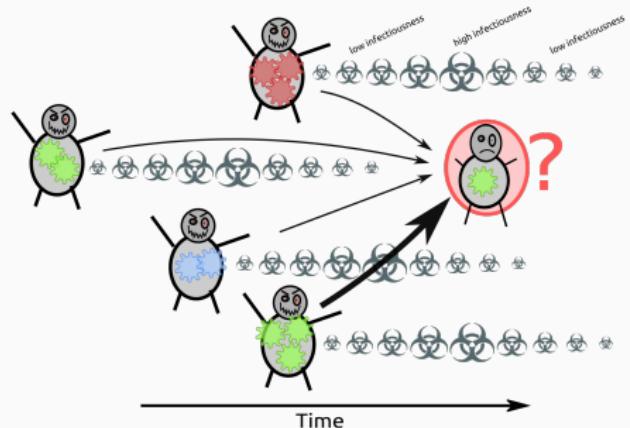


Since *outbreaker*: new models, data, and questions.

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**.

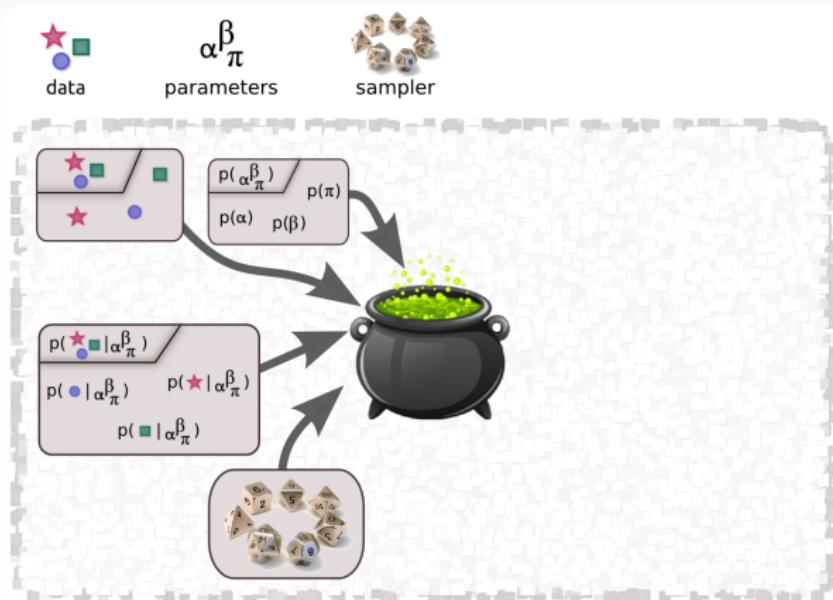
outbreaker2: a general cauldron for cooking methods

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



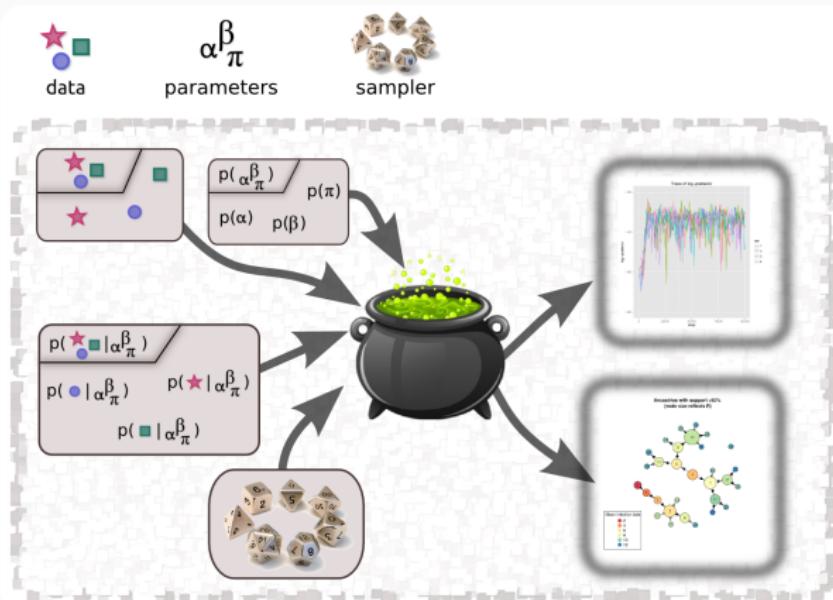
outbreaker2: a general cauldron for cooking methods

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



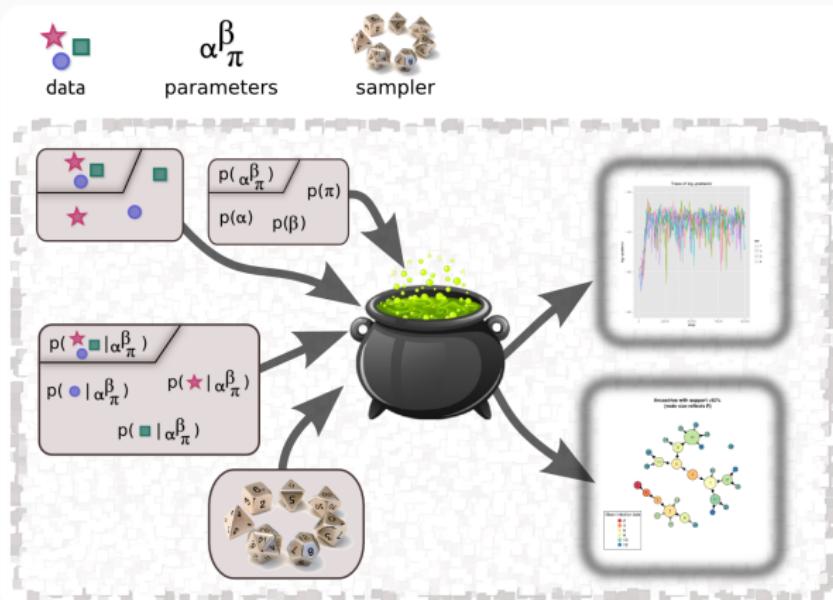
outbreaker2: a general cauldron for cooking methods

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



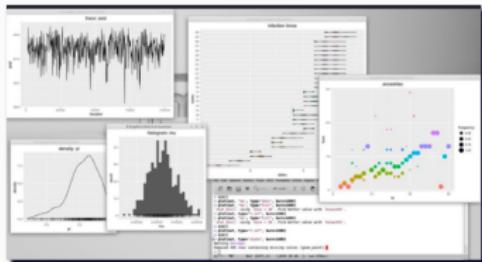
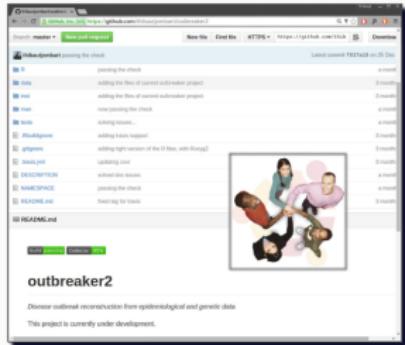
outbreaker2: a general cauldron for cooking methods

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



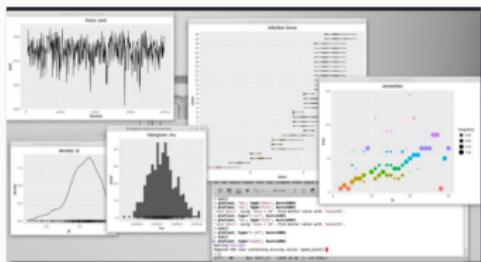
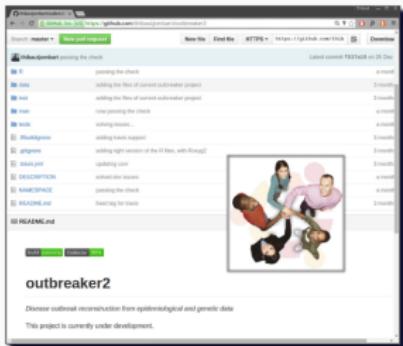
Modularity is key to generalising approaches

outbreaker2: a general tool for outbreak reconstruction

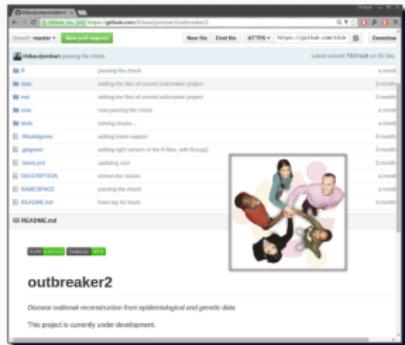


outbreaker2: a general tool for outbreak reconstruction

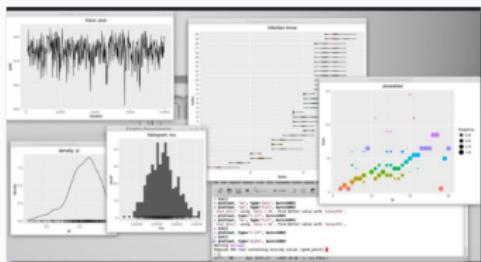
- **modularity:** likelihood, priors, samplers are all modules



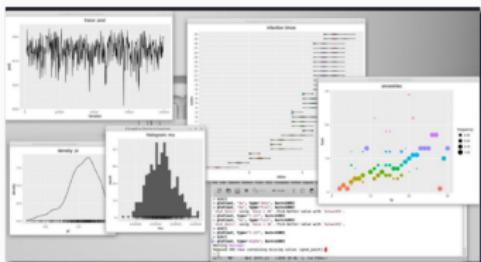
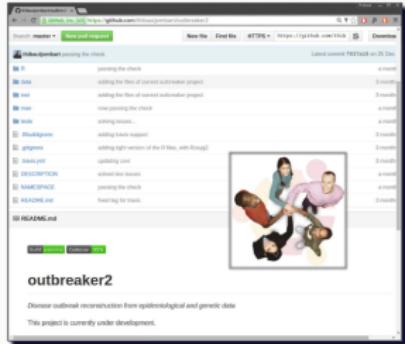
outbreaker2: a general tool for outbreak reconstruction



- **modularity:** likelihood, priors, samplers are all modules
- **new 'extensions':** contact tracing, spatial structure, new MCMC

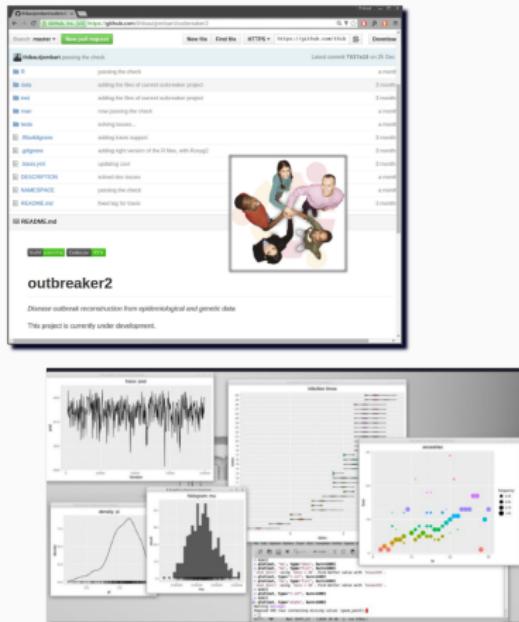


outbreaker2: a general tool for outbreak reconstruction



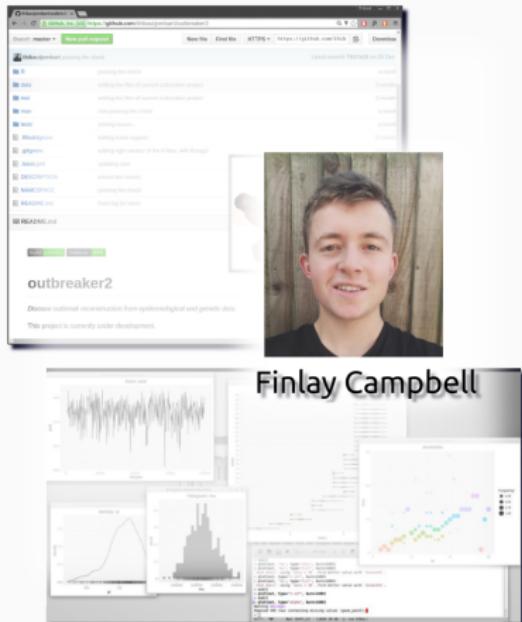
- **modularity**: likelihood, priors, samplers are all modules
 - **new 'extensions'**: contact tracing, spatial structure, new MCMC
 - **reliability**: continuous integration, extensive unit testing (aiming for 100% coverage)

outbreaker2: a general tool for outbreak reconstruction



- **modularity:** likelihood, priors, samplers are all modules
- **new 'extensions':** contact tracing, spatial structure, new MCMC
- **reliability:** continuous integration, extensive unit testing (aiming for 100% coverage)
- **prettier:** plot methods using *ggplot2*, interactive networks visualisation

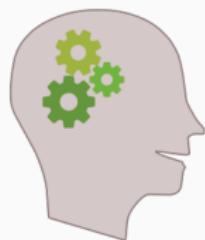
outbreaker2: a general tool for outbreak reconstruction



- **modularity**: likelihood, priors, samplers are all modules
- **new 'extensions'**: contact tracing, spatial structure, new MCMC
- **reliability**: continuous integration, extensive unit testing (aiming for 100% coverage)
- **prettier**: plot methods using *ggplot2*, interactive networks visualisation
- should **facilitate new contributions**

Methodological dialogue

Methodological development relies on an interdisciplinary dialogue

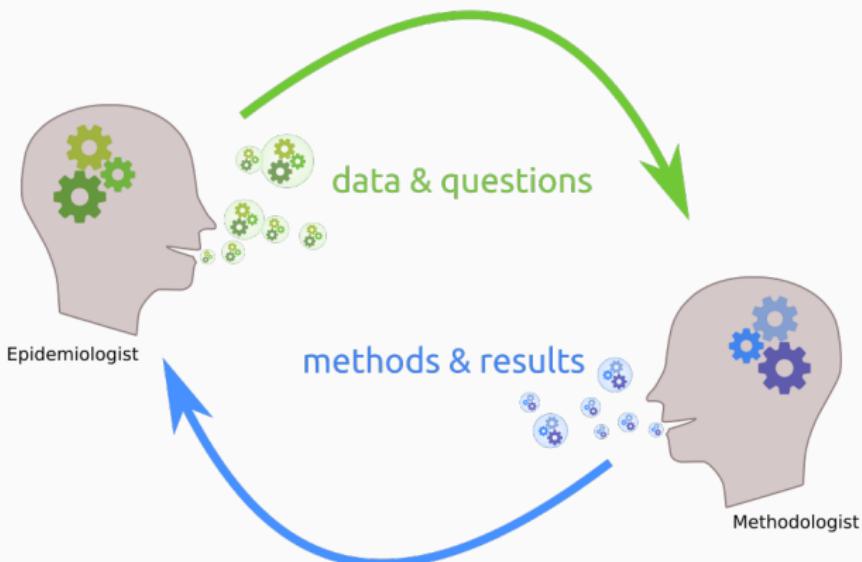


Epidemiologist

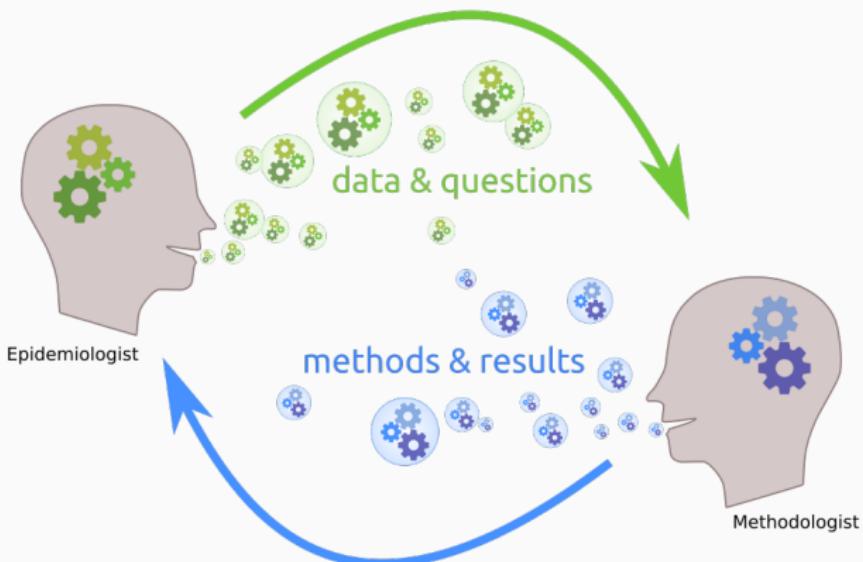


Methodologist

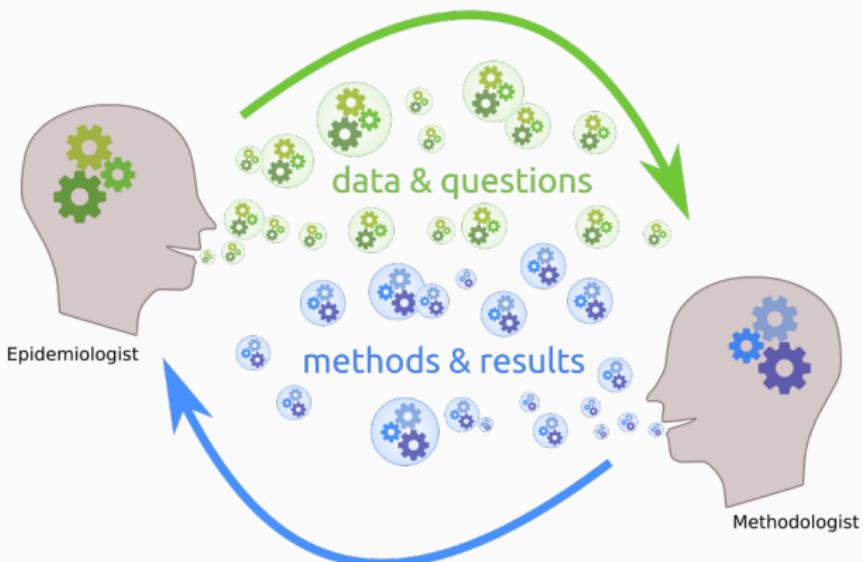
Methodological development relies on an interdisciplinary dialogue



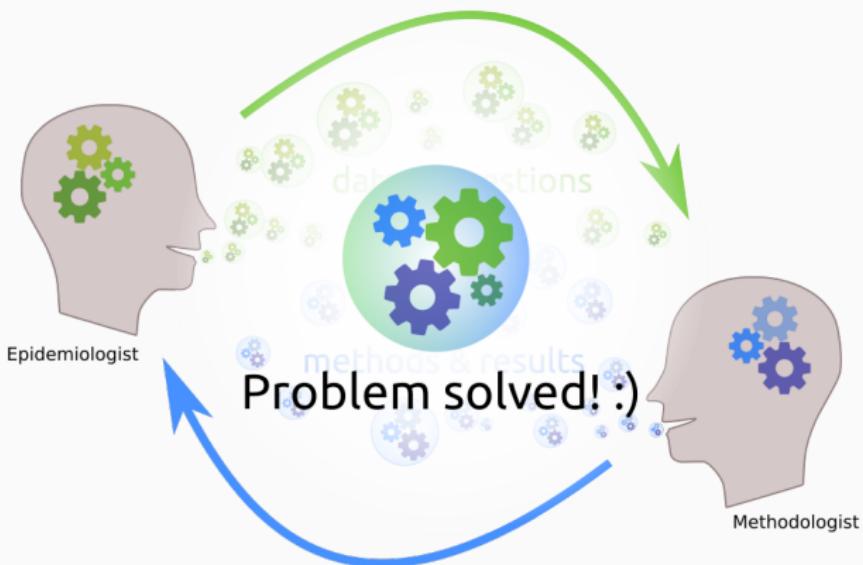
Methodological development relies on an interdisciplinary dialogue



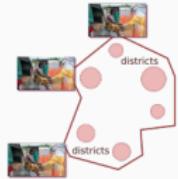
Methodological development relies on an interdisciplinary dialogue



Methodological development relies on an interdisciplinary dialogue



Outbreak response context creates distance and delays

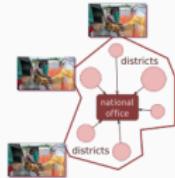


Affected countries



data collection

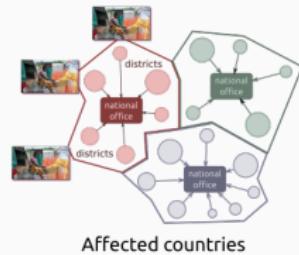
Outbreak response context creates distance and delays



Affected countries

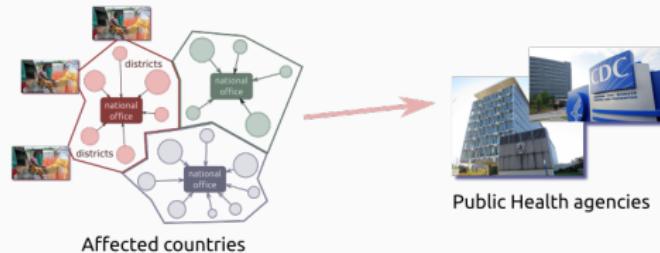
time (block = day)
● 
data collection

Outbreak response context creates distance and delays

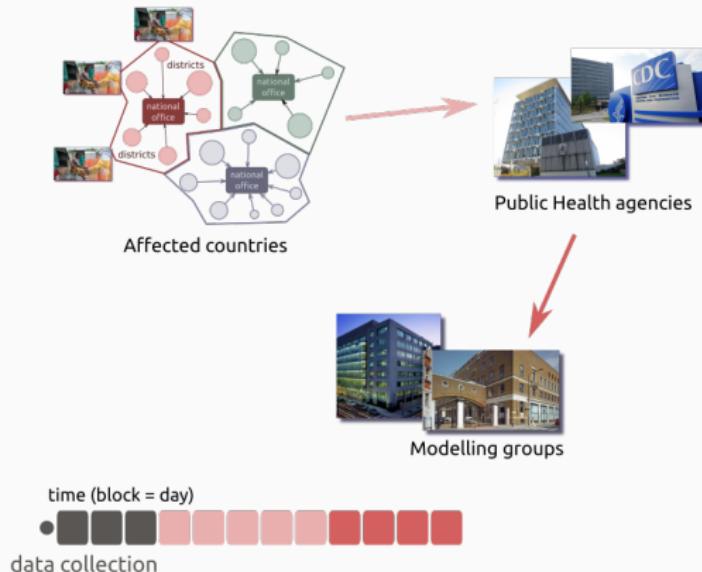


time (block = day)
● data collection

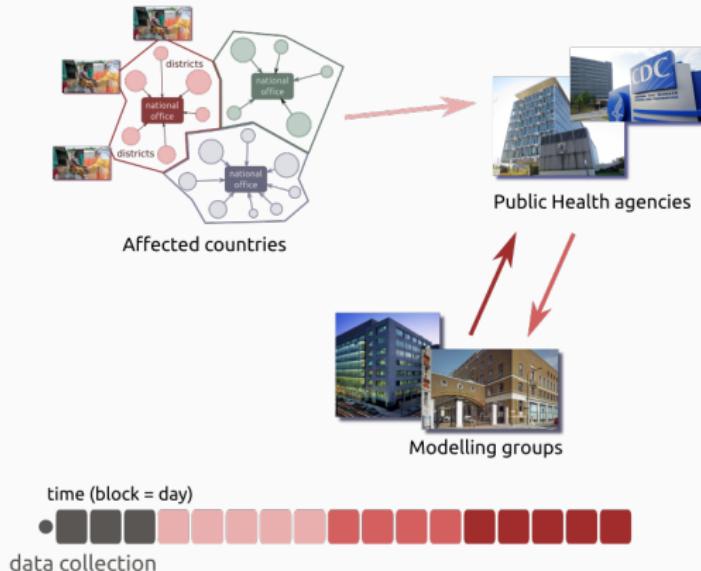
Outbreak response context creates distance and delays



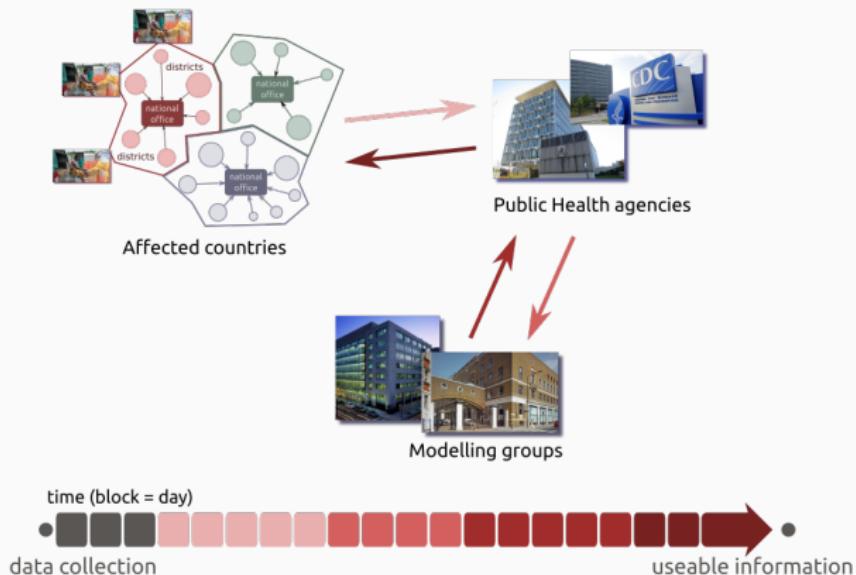
Outbreak response context creates distance and delays



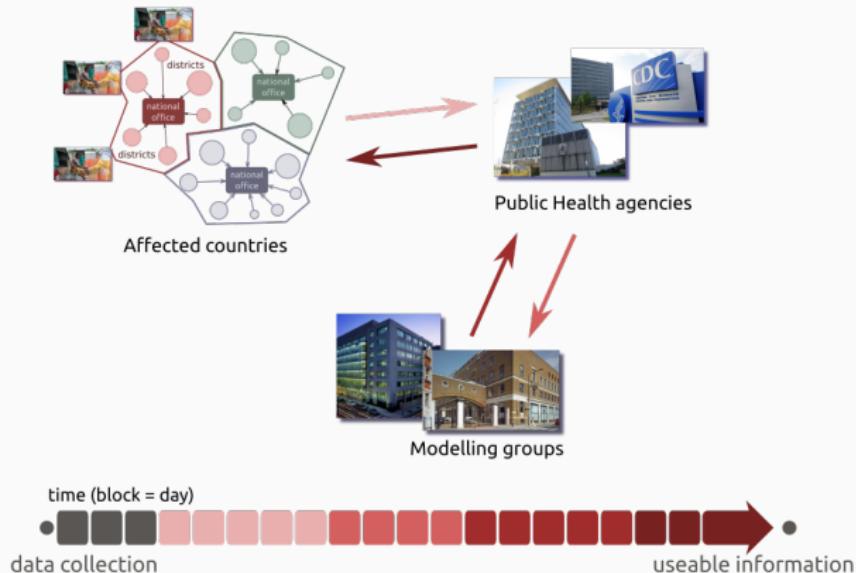
Outbreak response context creates distance and delays



Outbreak response context creates distance and delays

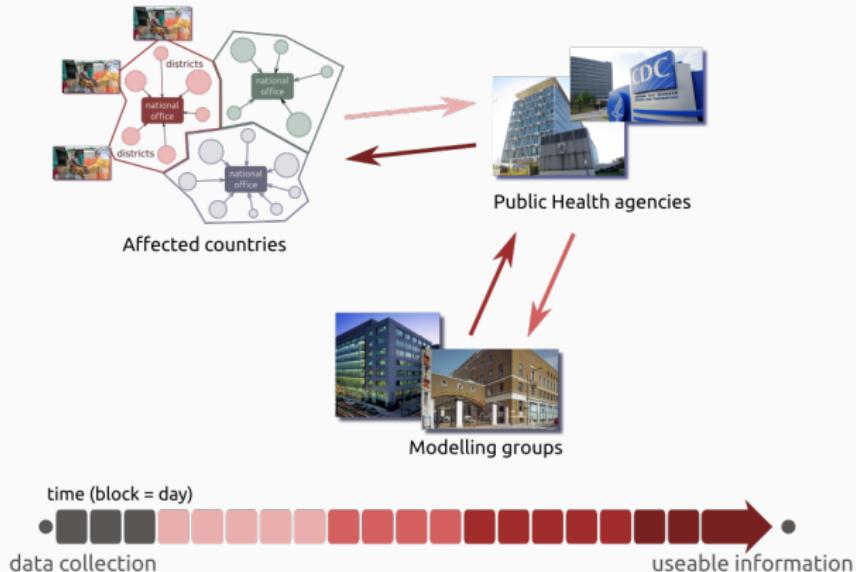


Outbreak response context creates distance and delays



- 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, GOARN
- **funding**: HPRU-NIHR, MRC

More on:

www.repidemicsconsortium.org

Questions?