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

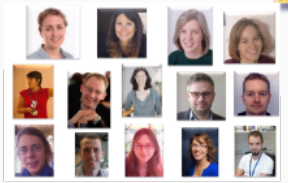
Outline

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



Imperial College Ebola team

Lessons learnt from the Ebola response

Lessons learnt from the Ebola response

Most statistical/modelling tools for situation awareness missing.

What tools do we need?

Some examples:

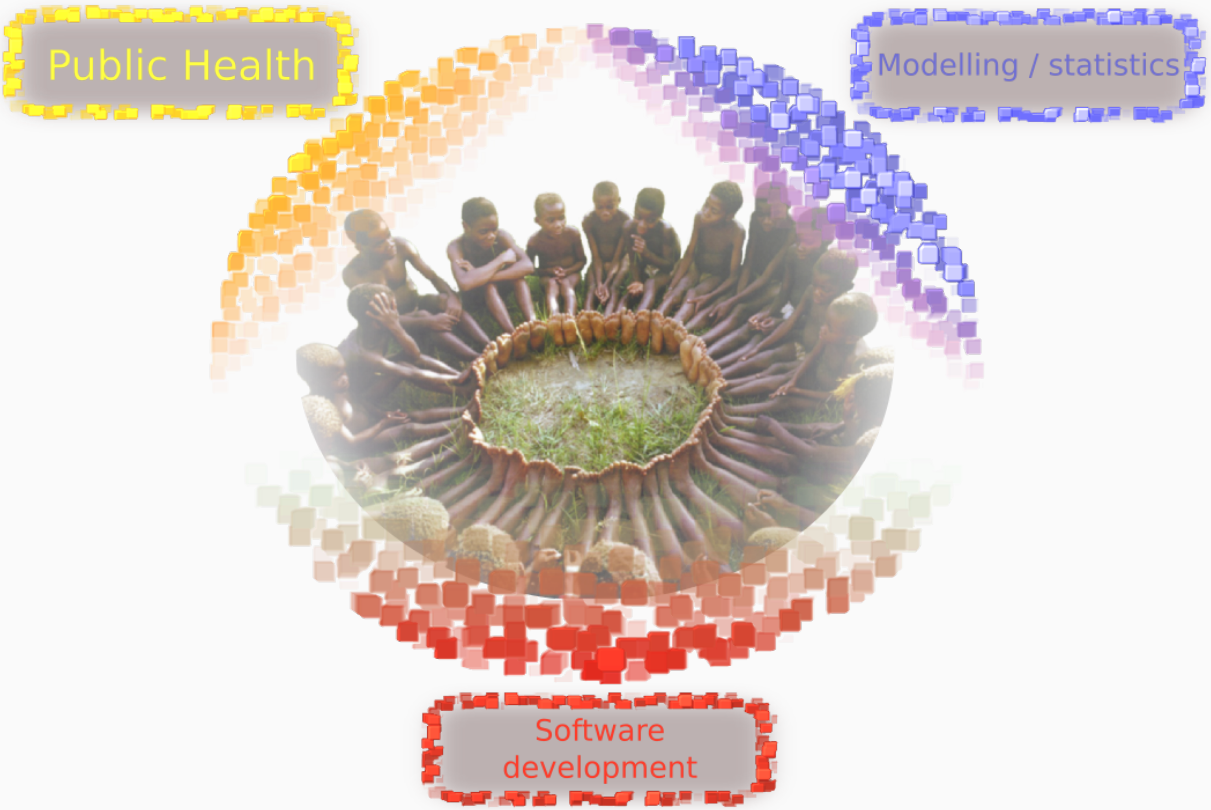
- **data cleaning:** dictionaries, entry matching
- **graphics:** case incidence in space and time, contact tracing
- **parameter estimation:** key delays, transmissibility
- **estimate / test CFR:** gender, health care workers, treaments effects
- **predictions:** case incidence, mortality, evaluate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

Who do we need to develop these tool?

Who do we need to develop these tool?

Who do we need to develop these tool?

Who do we need to develop these tool?



The R Epidemics Consortium

Hackout 3: a R hackathon for emergency outbreak response

Last summer at the *rOpenSci* headquarters (Berkeley)

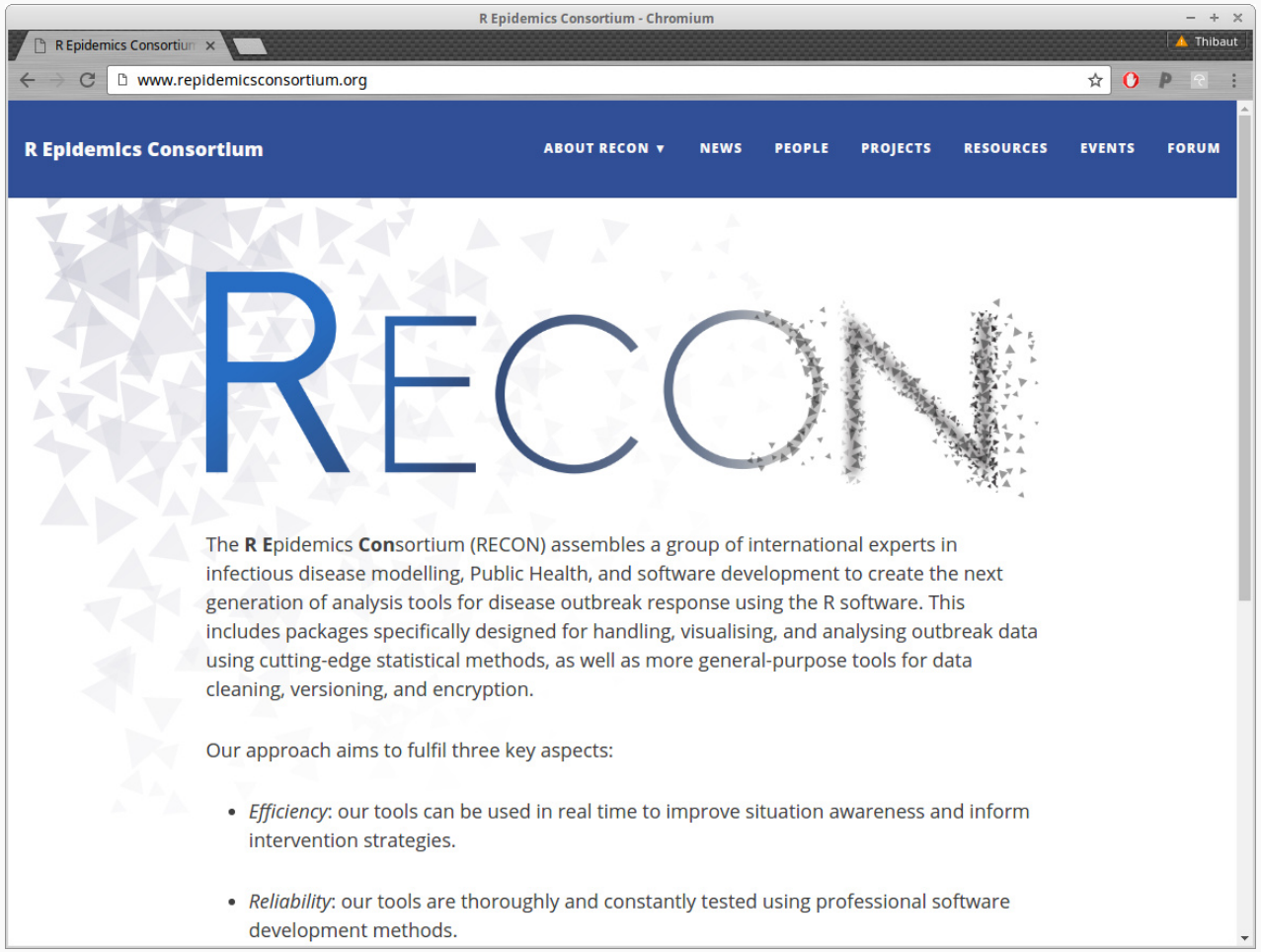
Hackout 3: from ideas to projects to...

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 .



www.repidemicsconsortium.org

In a nutshell

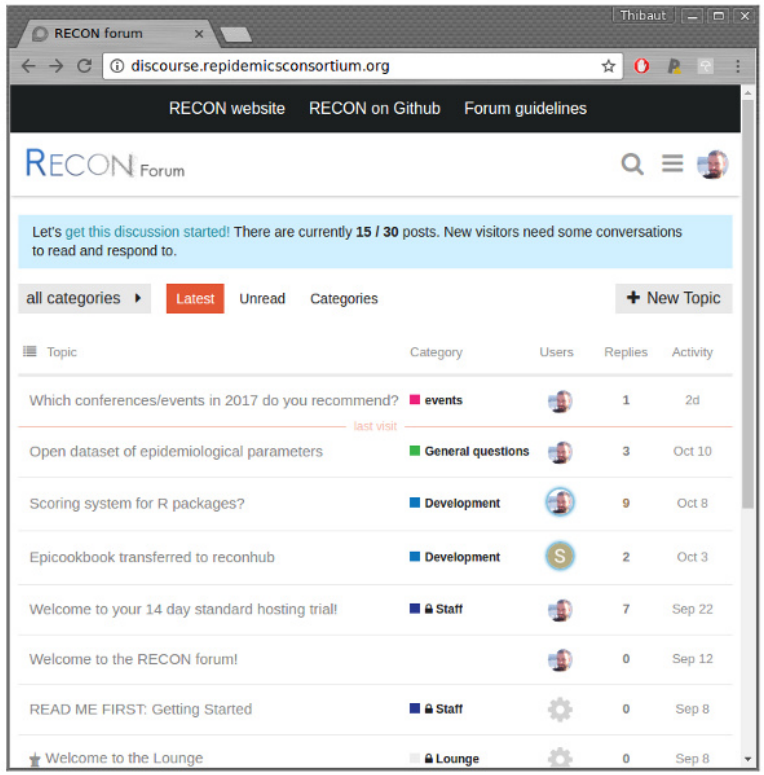
RECON

www.repidemicsconsortium.org

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

The RECON forum

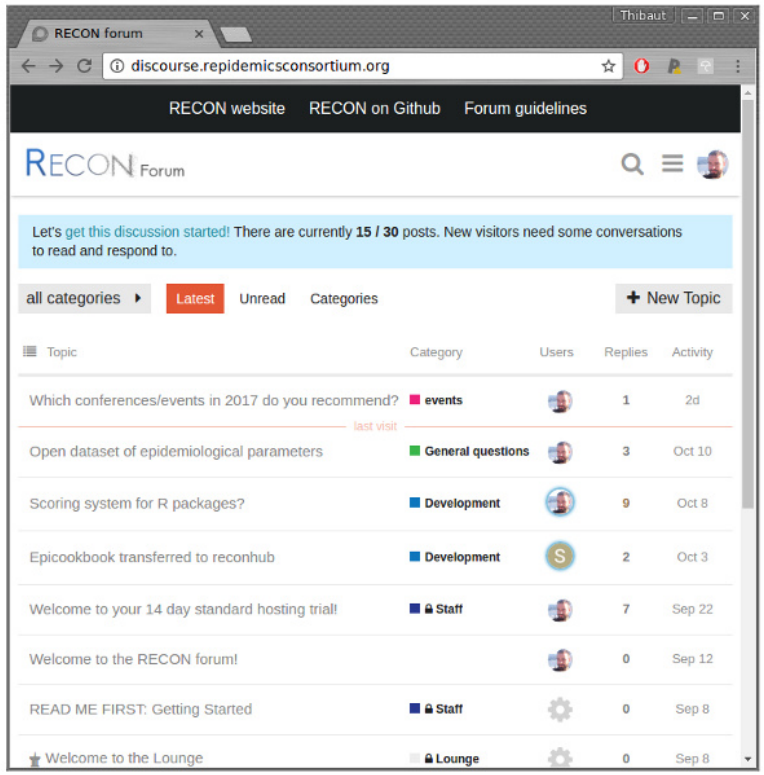
A platform for discussing epidemics analysis in R.



www.repidemicsconsortium.org/forum

The RECON forum

A platform for discussing epidemics analysis in R.



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?

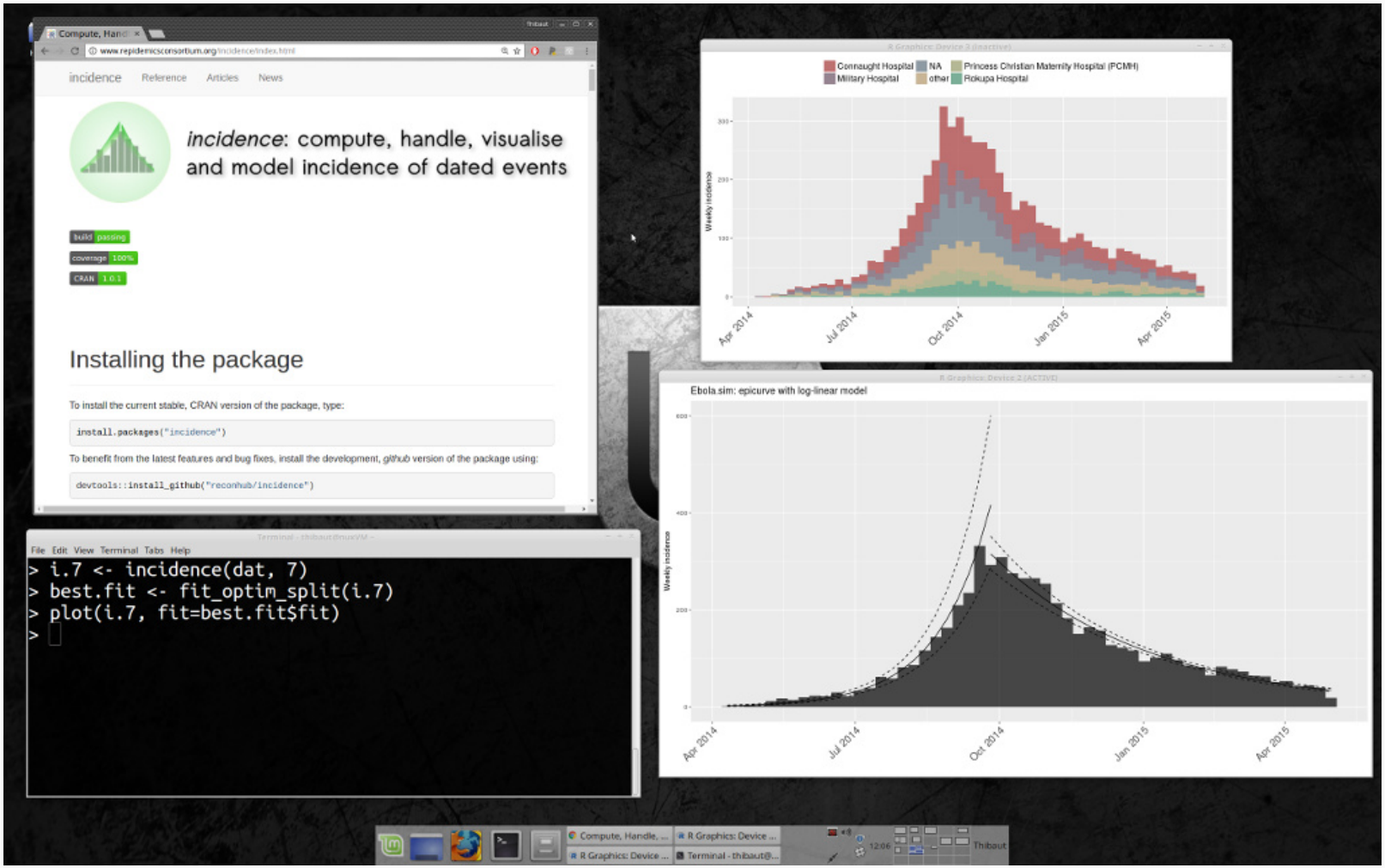
- **efficiency**: useful for improving situation awareness in real time; cutting-edge, computer-efficient statistical methods
- **reliability**: outputs can be trusted; continuous integration, extensive unit testing, code review, good practices

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

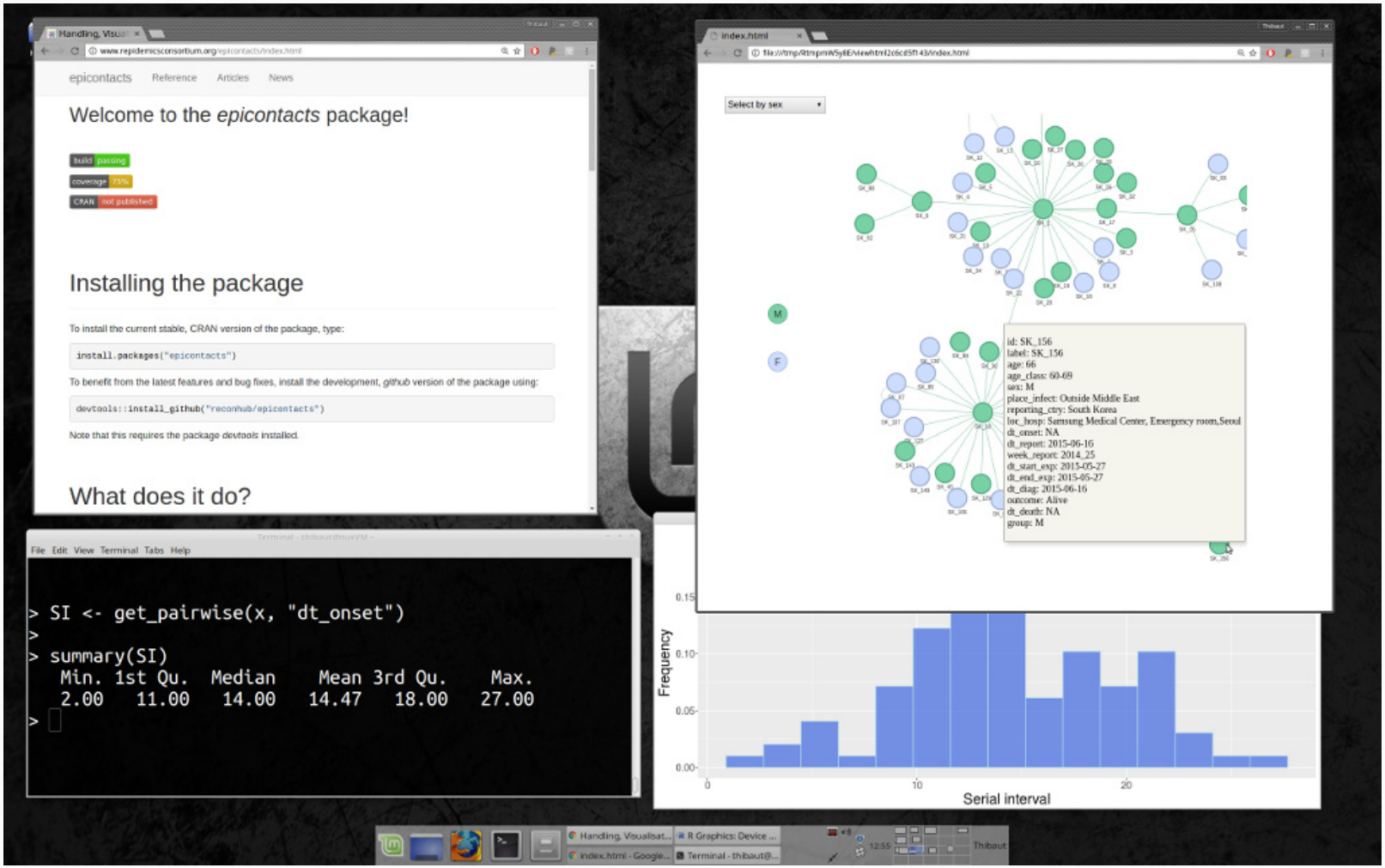
incidence: computation, handling, visualisation and modelling of epicurves



www.repidemicsconsortium.org/incidence

[released]

epicontacts: handling, visualisation and analysis of epidemiological contacts



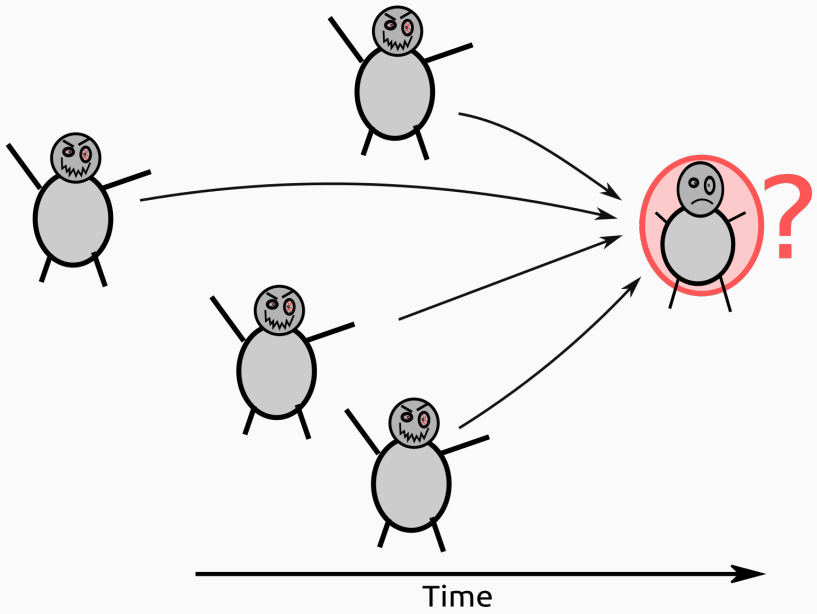
www.repidemicsconsortium.org/epicontacts

[release December 2016]

outbreaker2: inferring who infects whom in an outbreak

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

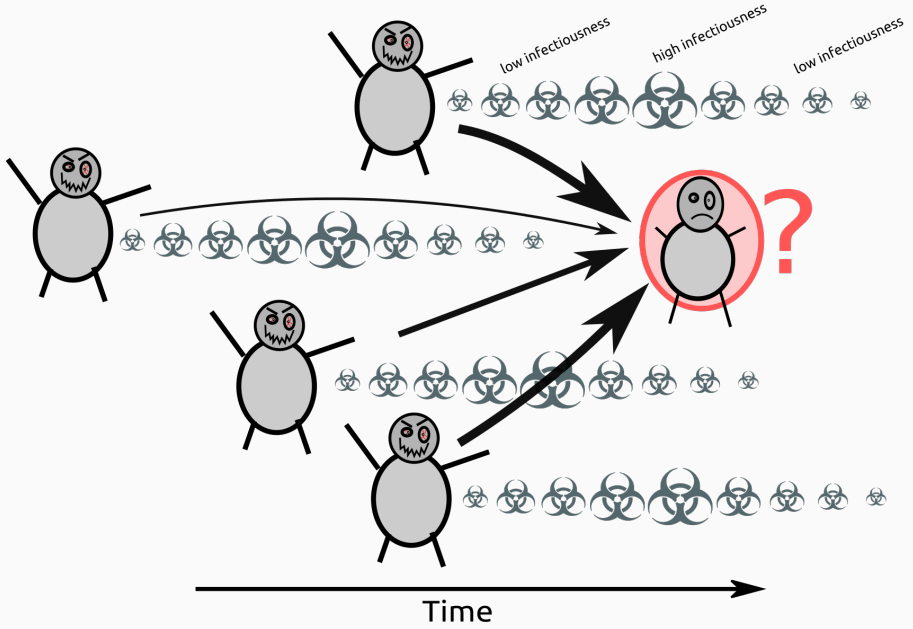
(Jombart *et al*, PLoS Comp Biol, 2014)



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)

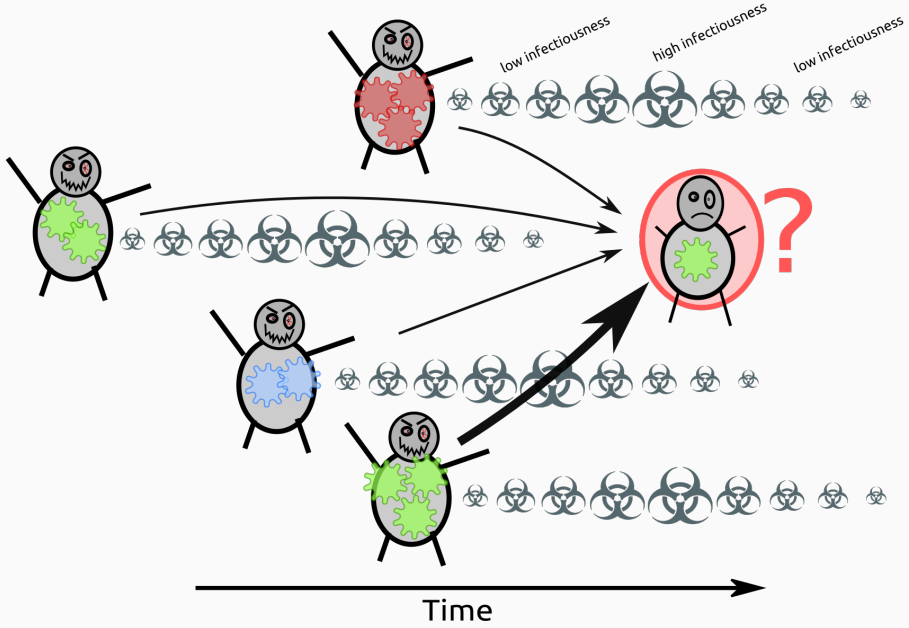


Since *outbreaker*: new models, data, and questions.

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*:

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

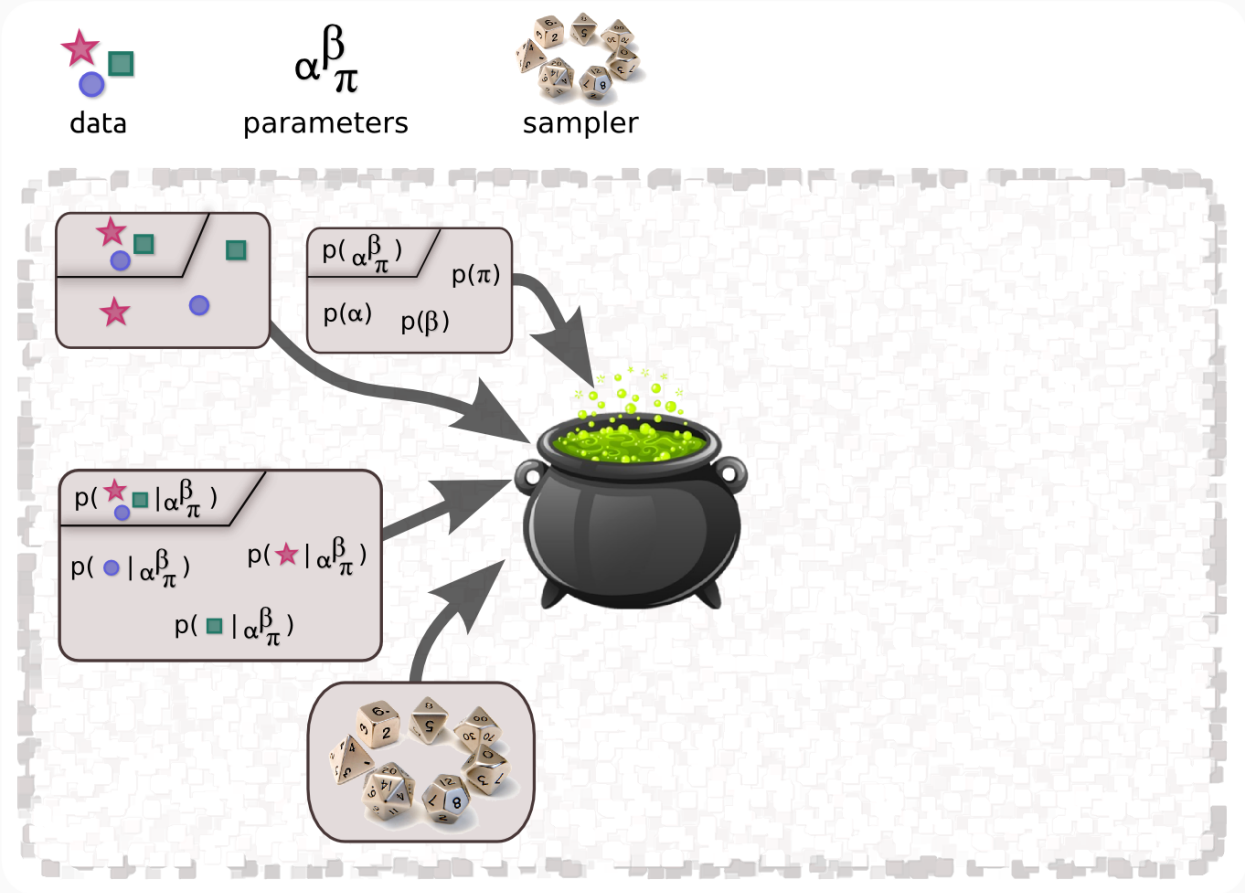
Components can be treated as plugins.

outbreaker2: a general cauldron for cooking methods

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

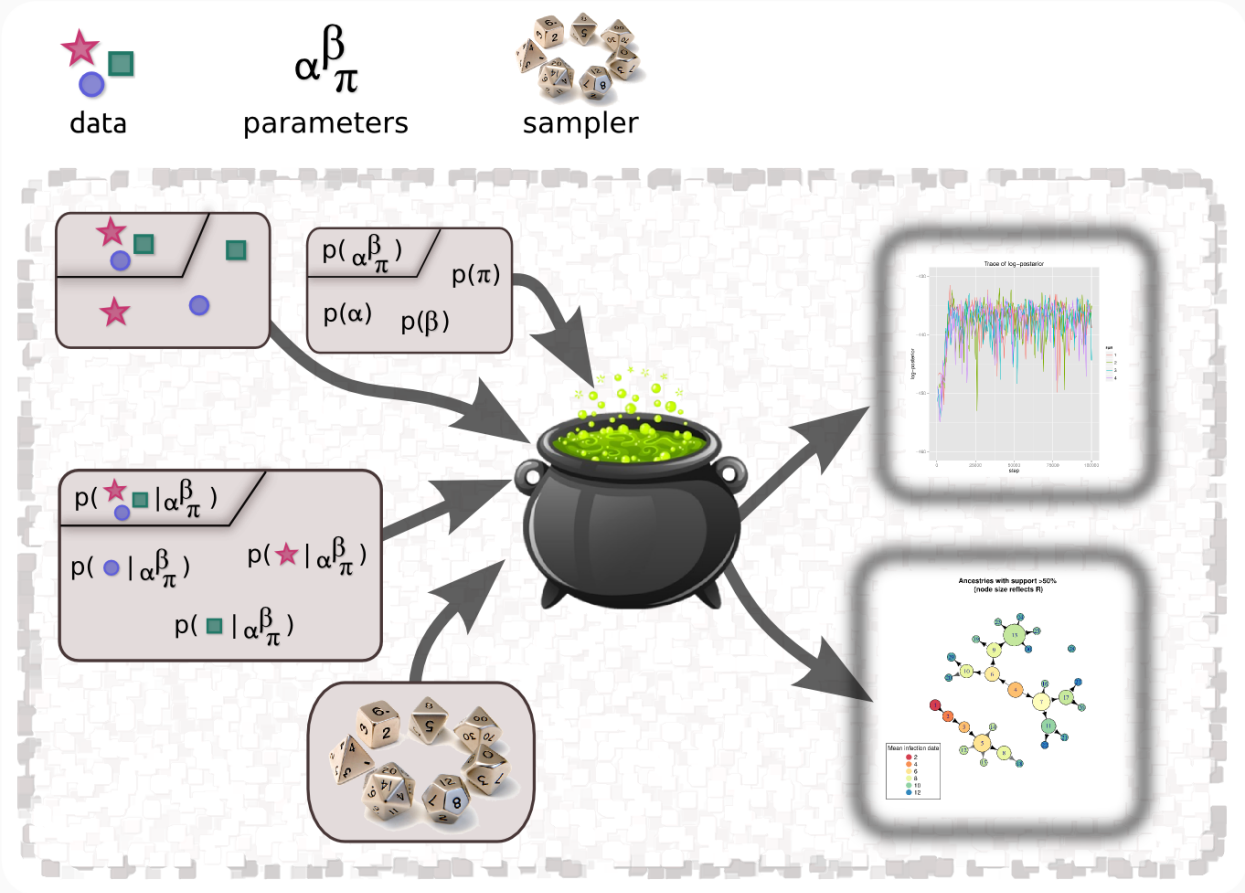
outbreaker2: a general cauldron for cooking methods

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



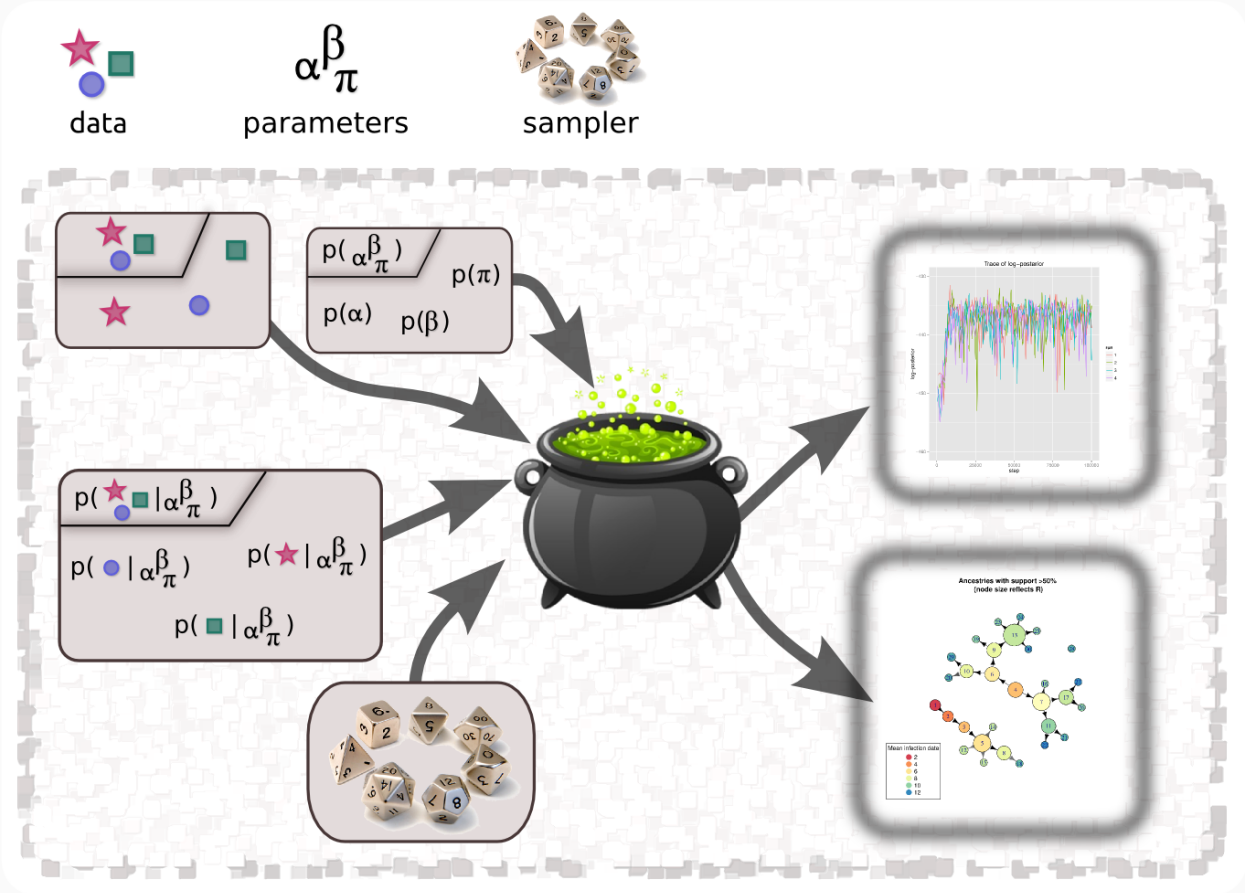
outbreaker2: a general cauldron for cooking methods

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



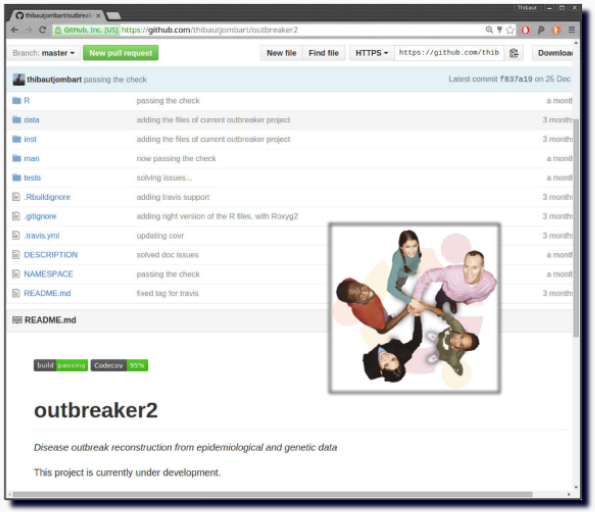
outbreaker2: a general cauldron for cooking methods

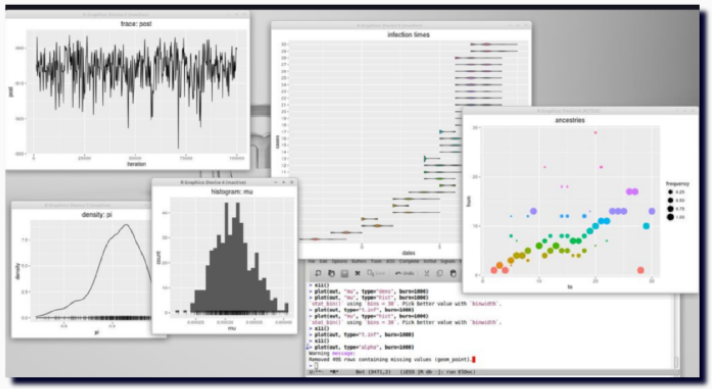
Use-your-own: data type, likelihood, prior, MCMC.



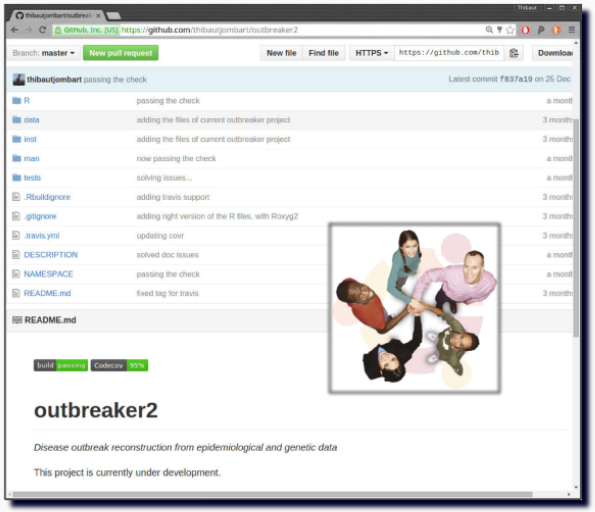
Modularity is key to generalising approaches

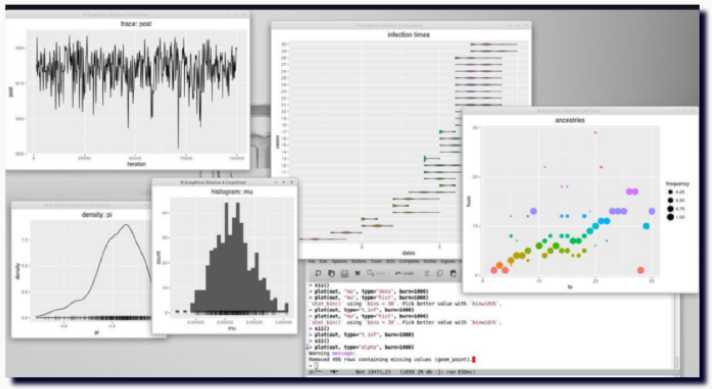
outbreaker2: a general tool for outbreak reconstruction





outbreaker2: a general tool for outbreak reconstruction

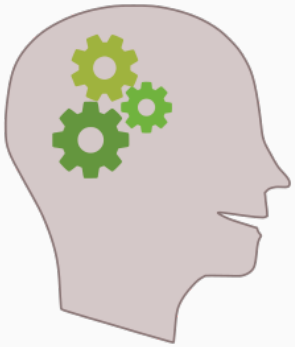




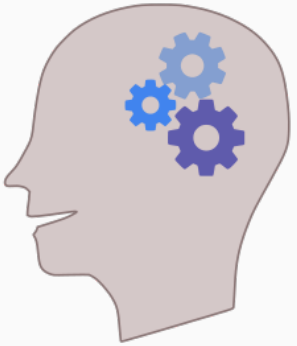
- **modularity**: customise model and sampler, optional Rcpp
- **reliability**: continuous integration, extensive unit testing (aiming for 100% coverage)
- **prettier**: plot methods using *ggplot2*, interactive networks visualisation
- enable **contributions from the community**
- release planned for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

Methodological dialogue

Methodological development relies on an interdisciplinary dialogue

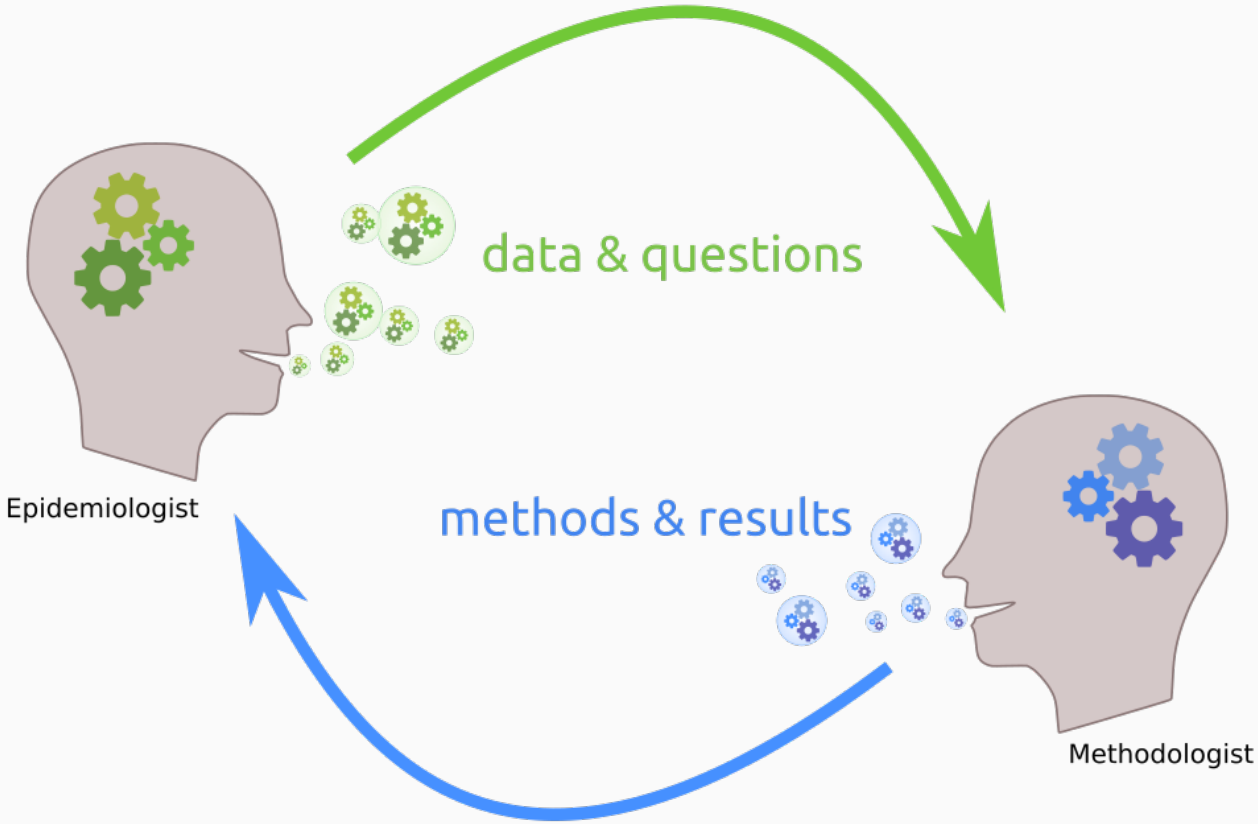


Epidemiologist

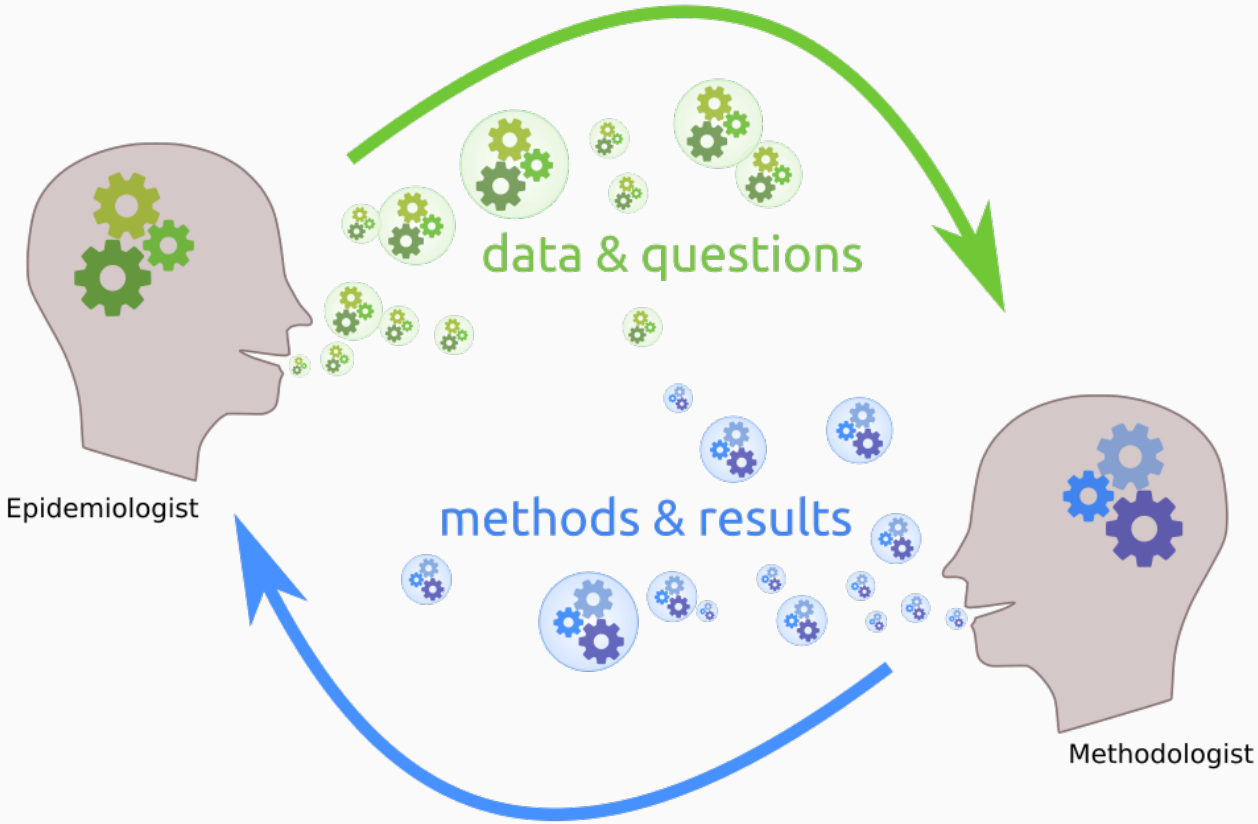


Methodologist

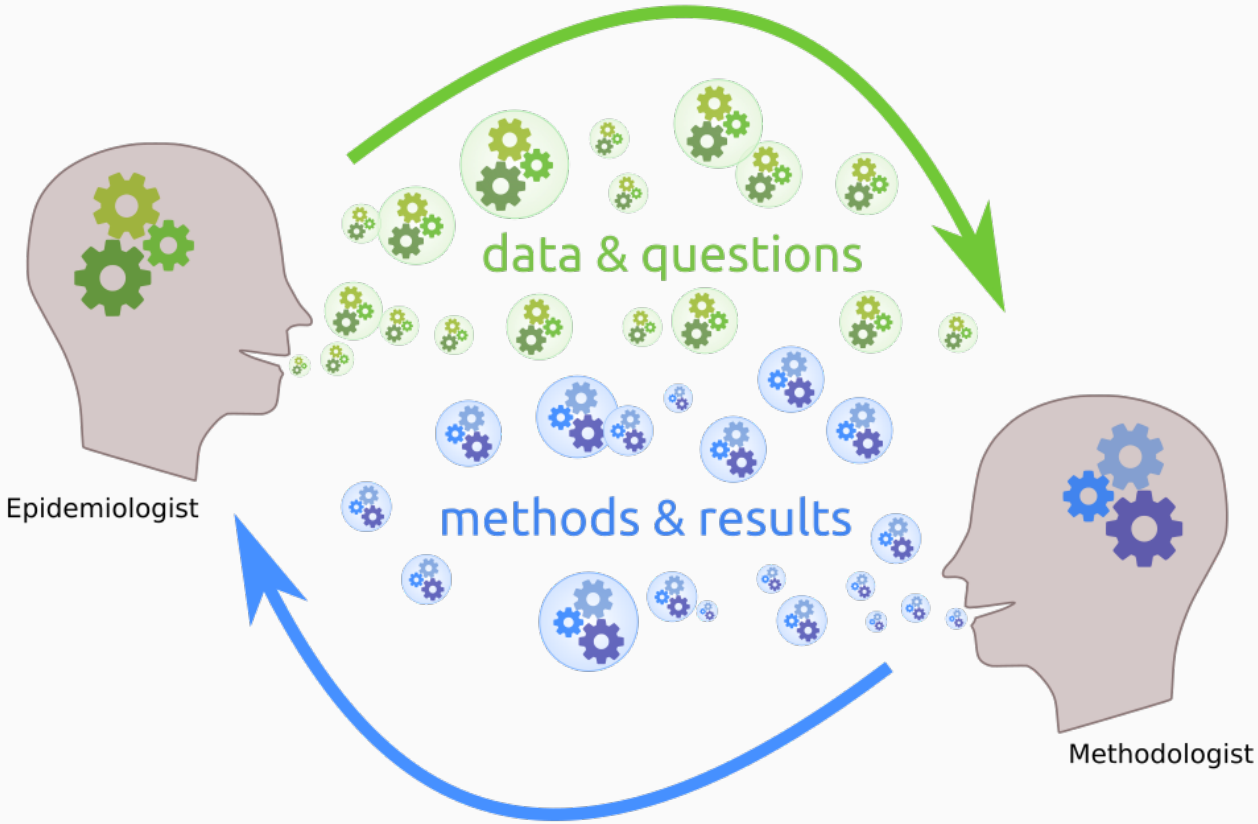
Methodological development relies on an interdisciplinary dialogue



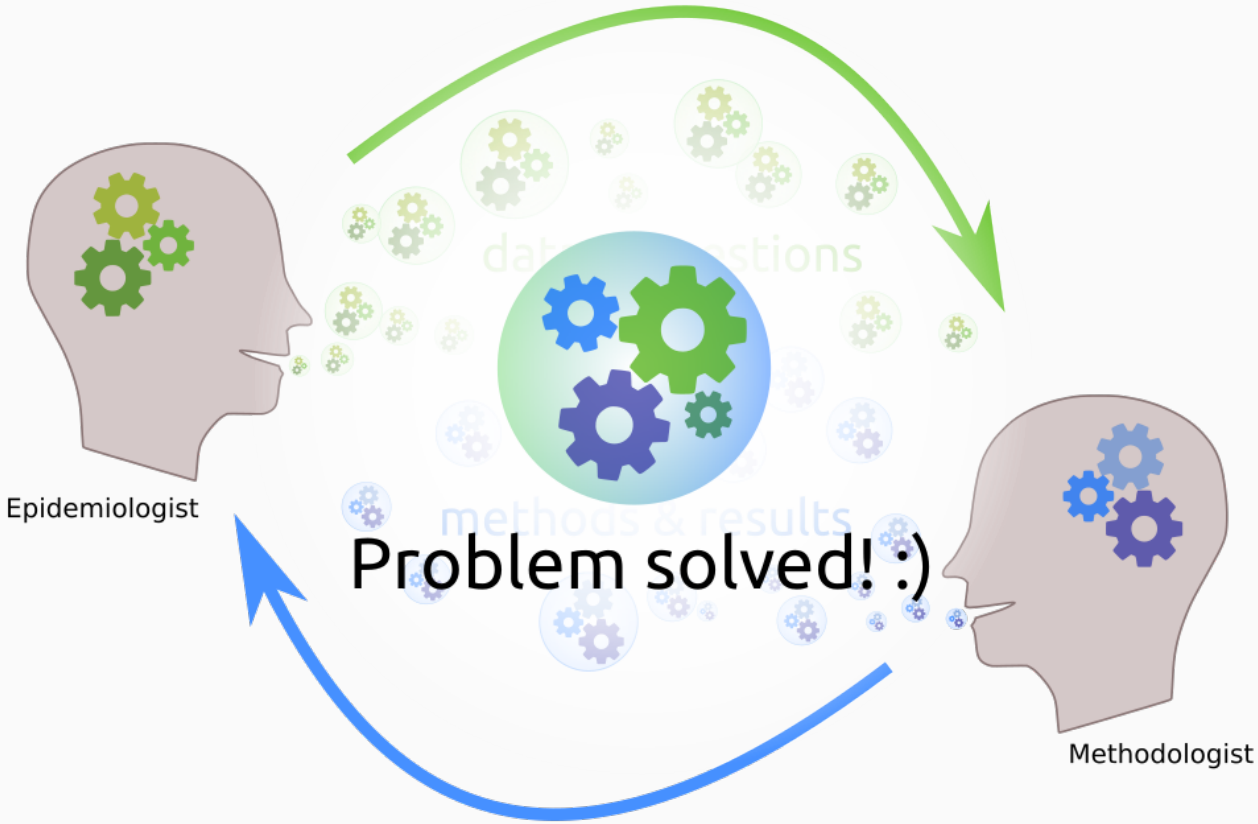
Methodological development relies on an interdisciplinary dialogue



Methodological development relies on an interdisciplinary dialogue



Methodological development relies on an interdisciplinary dialogue



Outbreak response context creates distance and delays

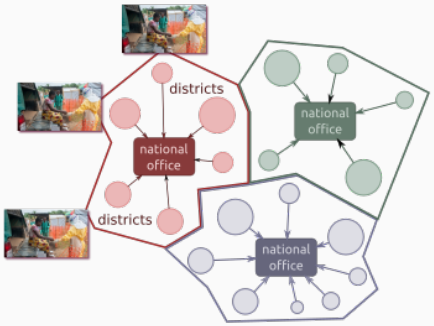
Affected countries

data collection

Outbreak response context creates distance and delays

Affected countries

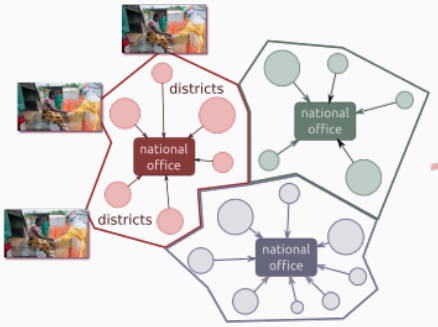
Outbreak response context creates distance and delays



Affected countries

data collection

Outbreak response context creates distance and delays

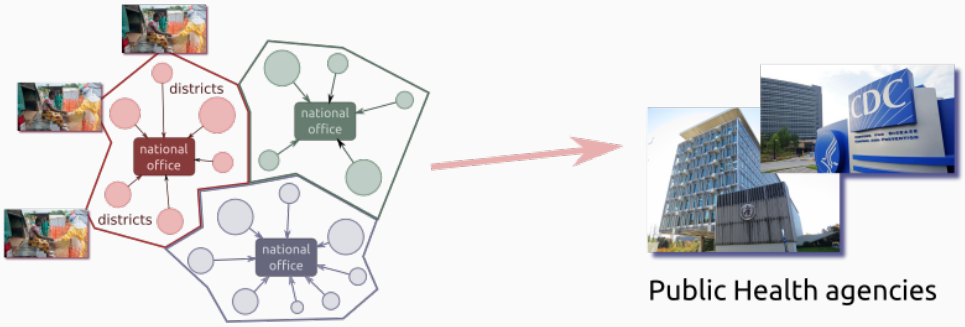


Affected countries

Public Health agencies

data collection

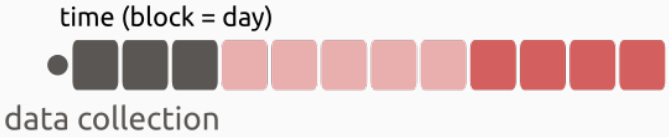
Outbreak response context creates distance and delays



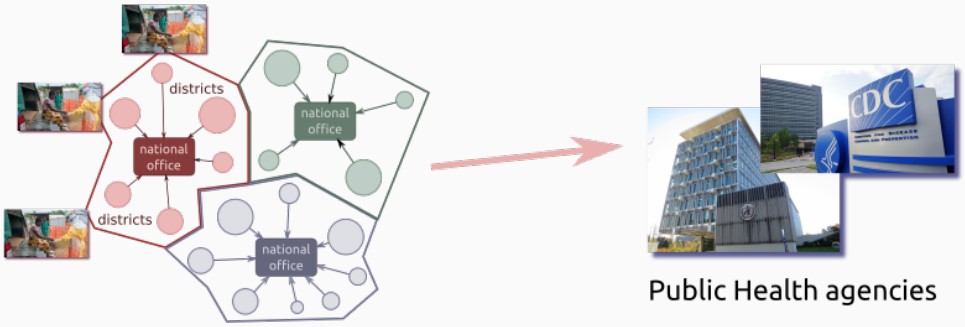
Affected countries

Public Health agencies

Modelling groups



Outbreak response context creates distance and delays



Affected countries

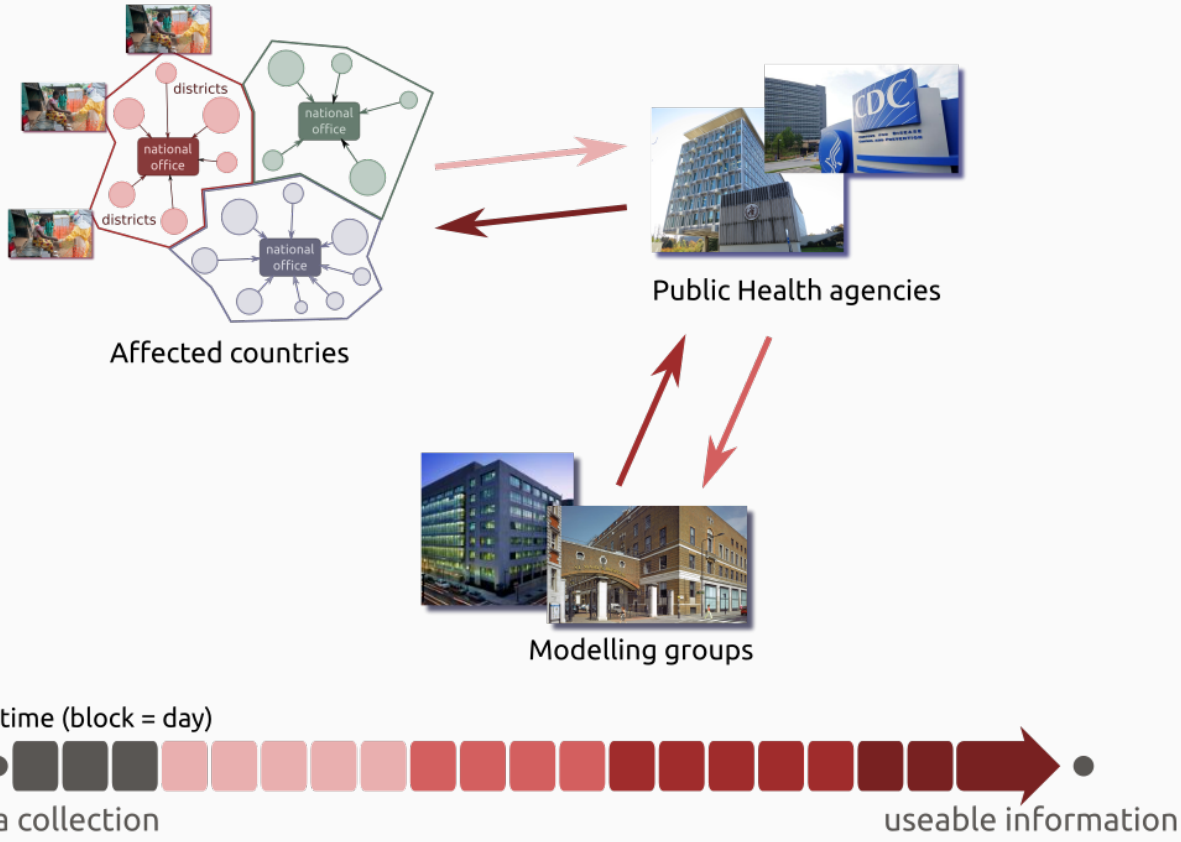
Public Health agencies

Modelling groups

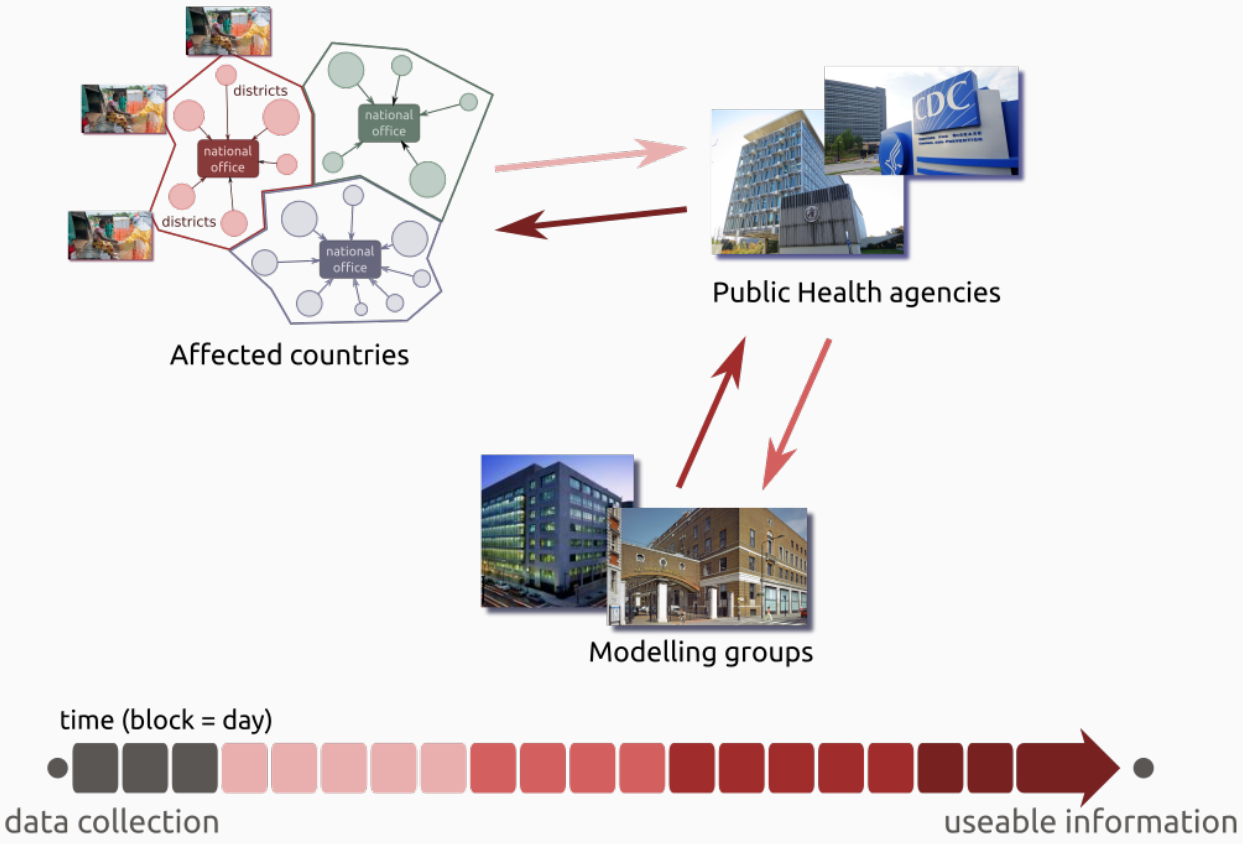
time (block = day)

data collection

Outbreak response context creates distance and delays

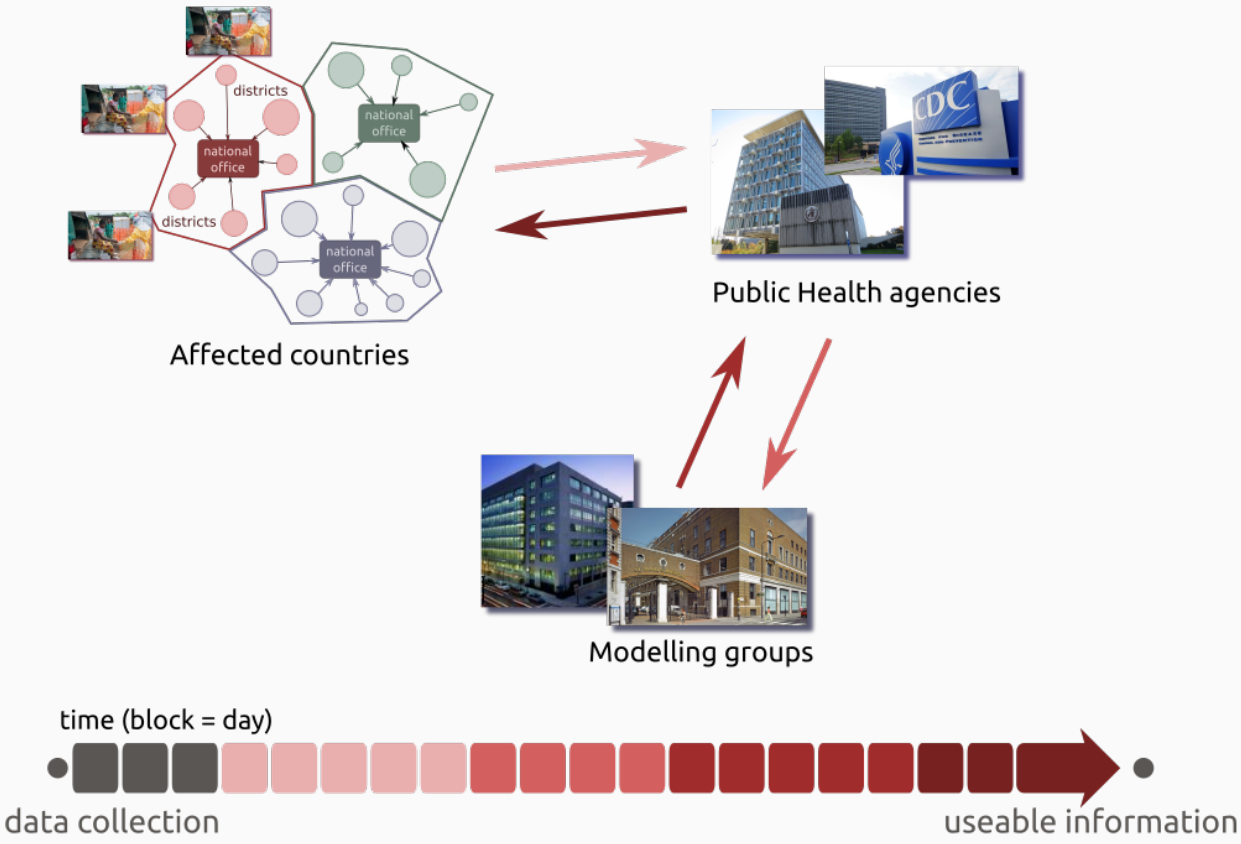


Outbreak response context creates distance and delays



- 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?