

# Supporting modeling hubs across the globe

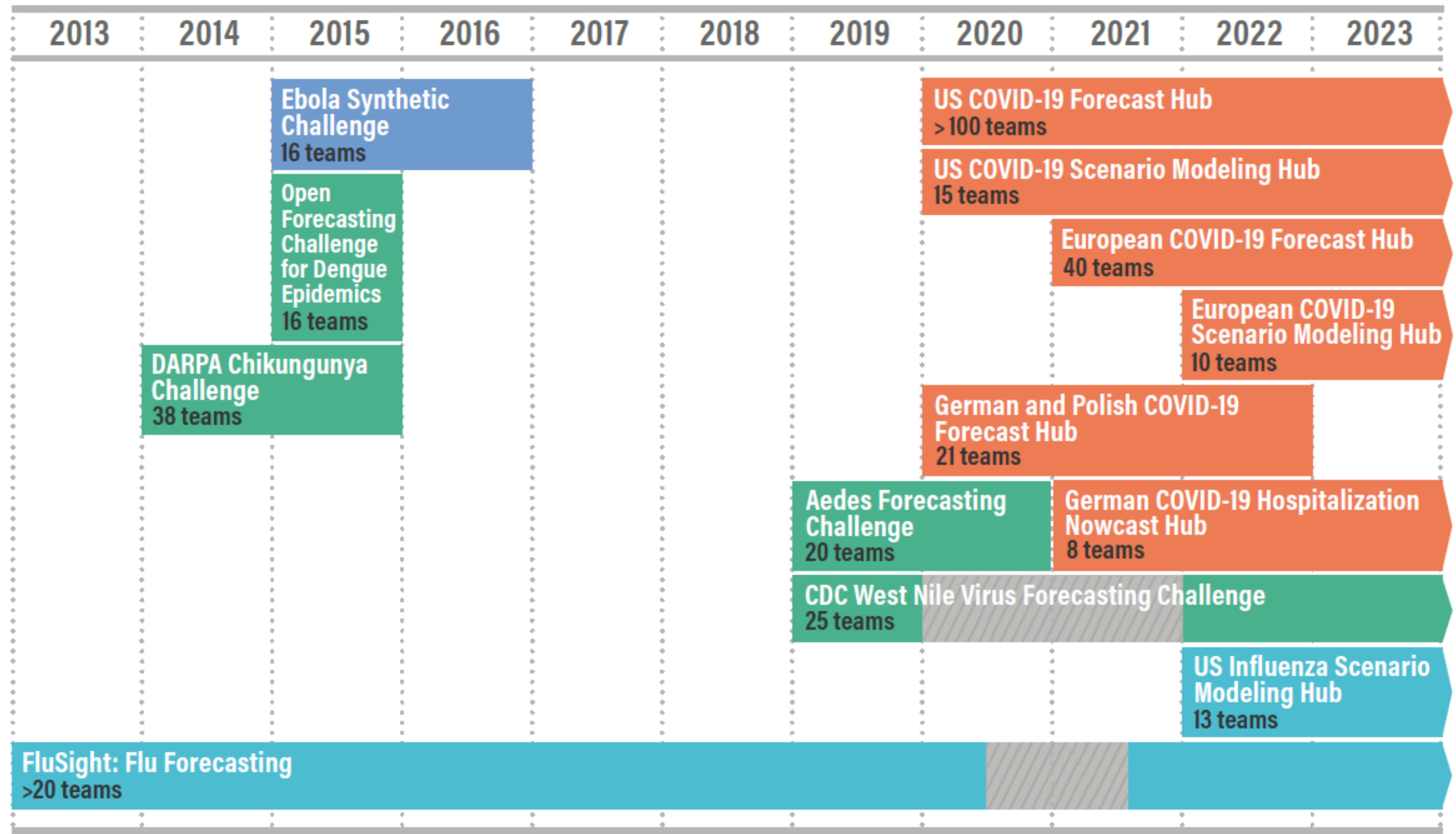
Nicholas G. Reich

ECDC RespiCast Hub Launch  
20 November 2023

# The State of Forecasting, early 2010s

*“Comparing the accuracy of forecasting applications is difficult because forecasting methods, forecast outcomes, and reported validation metrics varied widely.”*  
Chretien et al., PLOS ONE, 2014

- Landscape of epidemic forecast models
  - Developing sense of the key challenges (e.g., data revisions, exponential growth, horizons of predictability)
  - No clear standards for model evaluation
  - Little to no model coordination or synthesis
- Linkage to public health
  - What should we forecast? Daily/weekly/monthly incidence, epidemic duration, time of peak, ...
  - How can the forecasts be used to support decision-making?



● Influenza

● Vector-borne disease

● Ebola

● COVID-19

Figure credits: Alex Vespignani and Nicole Samay

# The State of Forecasting, early 2020s

- Landscape of epidemic forecast models
  - Developing sense of the key challenges (e.g., data revisions, exponential growth)
    - ➔ lots of research has been done to understand these challenges
    - ➔ taxonomies of forecasting models (e.g., statistical to mechanistic)
    - ➔ improved understanding of what models work, and at what horizon
  - No clear standards for model evaluation
    - ➔ EPIFORGE 2020 guidelines (Pollett et al, 2021, PLOS Med)
    - ➔ open-source software tools (scoringutils, etc...)
  - Little to no model coordination or synthesis
    - ➔ coordination and ensemble models are the norm

# Model coordination is key

- A combination of individual forecasts is pragmatic: it reduces dependency on a single model or team, and incorporates multiple perspectives.
- One consistent finding across all efforts:

***Combining models into an "ensemble" provides more consistent forecasts than any single model.***

Flu: Reich et al. 2019, *PLOS Comp Bio*. <https://doi.org/10.1371/journal.pcbi.1007486>  
Flu: McGowan et al. 2019, *Sci Rep*. <https://doi.org/10.1038/s41598-018-36361-9>  
Dengue: Johansson et al. 2019, *PNAS*.  
Ebola: Viboud et al. 2018, *Epidemics*.  
COVID-19: Cramer et al. 2022, *PNAS*.  
COVID-19: Ray et al. 2022, *Int'l J Forecasting*.  
COVID-19: Sherratt et al. 2023, *eLife*.

# The State of Forecasting, early 2020s

- Linkage to public health (**lots of open questions still**)
  - What should we forecast? Daily/weekly/monthly incidence, epidemic duration, time of peak, ...
    - ➔ Still we have variation in data sources and forecast targets
    - ➔ Questions remain about spatial and temporal granularity
  - How can the forecasts be used to support decision-making?
    - ➔ Lots of documented use for improving “situational awareness”
    - ➔ But, examples of use in decision-making are few and far between
    - ➔ Need for close collaboration to tailor models and evaluations to support specific decisions

Since late 2021, a collaborative team made up of people who worked on building and maintaining hubs during the pandemic have met regularly to develop the tools we wished we had during COVID!



June 2023, Amherst MA, USA



## Overview

- An open-source toolkit to support collaborative modeling efforts in public health.\*
- Framework for forecasts, nowcasts, scenario modeling, parameter estimates, ...
- Customizable modeling targets in standard formats to minimize duplicative tool development.
- Support for multiple representations of probabilistic model outputs (e.g., samples, quantiles, ...).

More info: [hubdocs.readthedocs.io](https://hubdocs.readthedocs.io)

\* Note that the tools can also be used to facilitate model development by individual researchers.



# hubverse

tools in production at



+ ECDC RespiCast Hub

with planned back-compatibility for



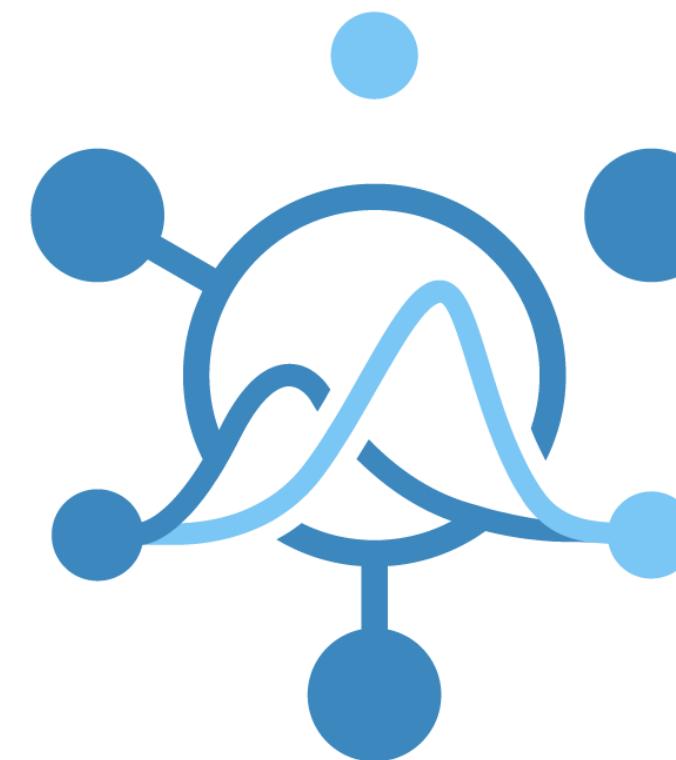
+ hopefully many others!



## Progress on hubverse goals

- Clear definition of a data format for model output
- Structural requirements for a hub file/directory system
- Open-source tools to access and work with data
  - hubUtils
  - hubVis
  - hubEnsembles
  - hubValidations
  - hubCI
- Documentation of concepts and tools
- (...) Template hubs and continuous integration workflows
- (...) Model containerization for interfacing with hubs
- (...) Dashboards and interactive real-time visualizations

# Progress on hubverse goals



# hubverse

- ✓ Clear definition of a data format for model output
- ✓ Structural requirements for a hub file/directory system
- (✓) Open-source tools to access and work with data

- hubUtils
- hubVis
- hubEnsembles
- hubValidations
- hubCI

```
## extract some data  
> library(hubUtils)  
> hub_con <- connect_hub('path/to/my/hub')  
> model_outputs <- hub_con %>%  
  filter(output_type == "quantile", location == "DE") %>%  
  collect()
```

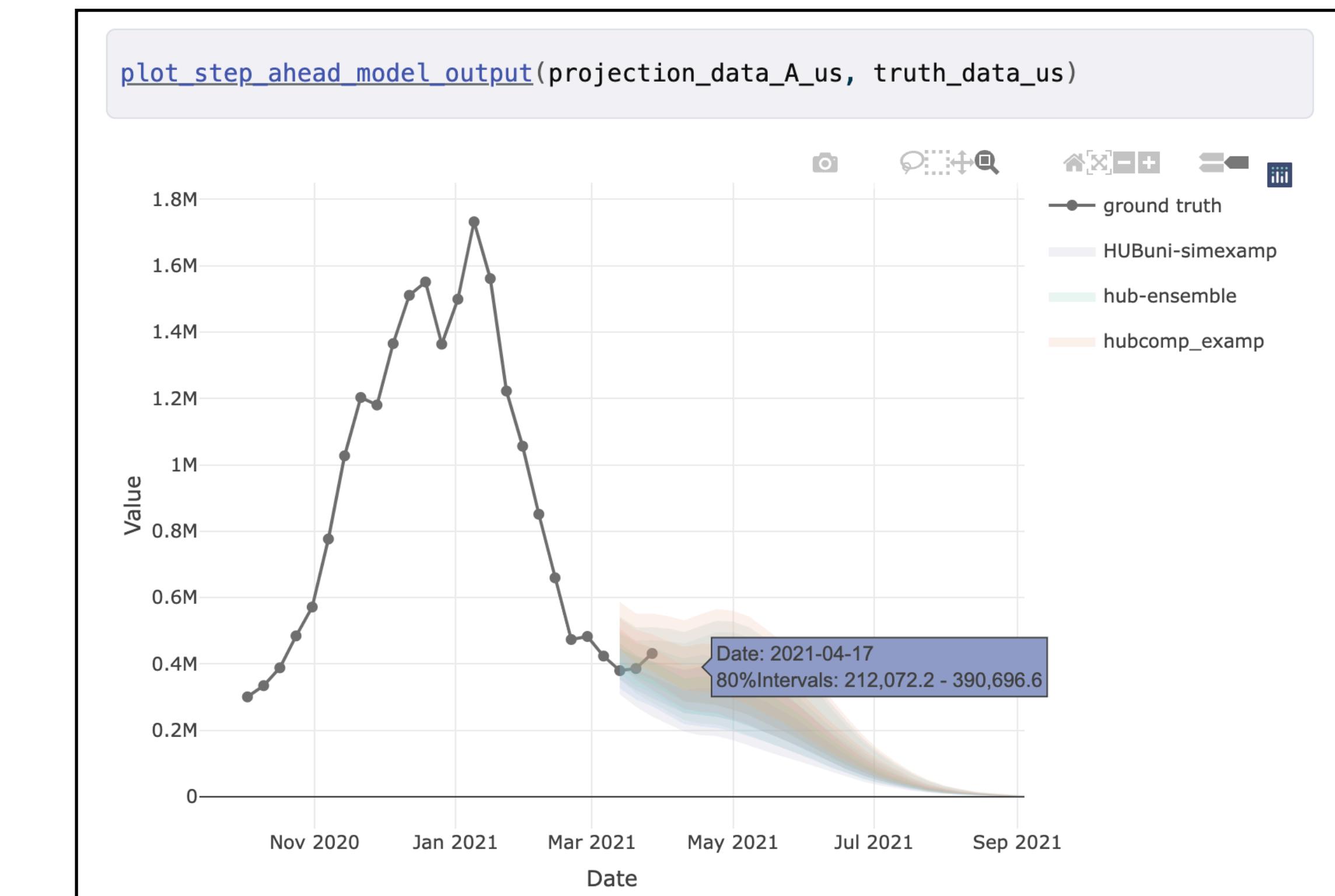
- (✓) Documentation of concepts and tools
- (...) Template hubs and continuous integration workflows
- (...) Model containerization for interfacing with hubs
- (...) Dashboards and interactive real-time visualizations

# Progress on hubverse goals

- ✓ Clear definition of a data format for model output
- ✓ Structural requirements for a hub file/directory system
- (✓) Open-source tools to access and work with data
  - hubUtils
  - hubVis
  - hubEnsembles
  - hubValidations
  - hubCI
- (✓) Documentation of concepts and tools
- (...) Template hubs and continuous integration workflows
- (...) Model containerization for interfacing with hubs
- (...) Dashboards and interactive real-time visualizations



# hubverse



## Progress on hubverse goals

- ✓ Clear definition of a data format for model output
- ✓ Structural requirements for a hub file/directory system
- (✓) Open-source tools to access and work with data

- hubUtils
- hubVis
- hubEnsembles
- hubValidations
- hubCI



# hubverse

- (✓) Documentation of concepts and tools
- (...) Template hubs and continuous integration workflows
- (...) Model containerization for interfacing with hubs
- (...) Dashboards and interactive real-time visualizations

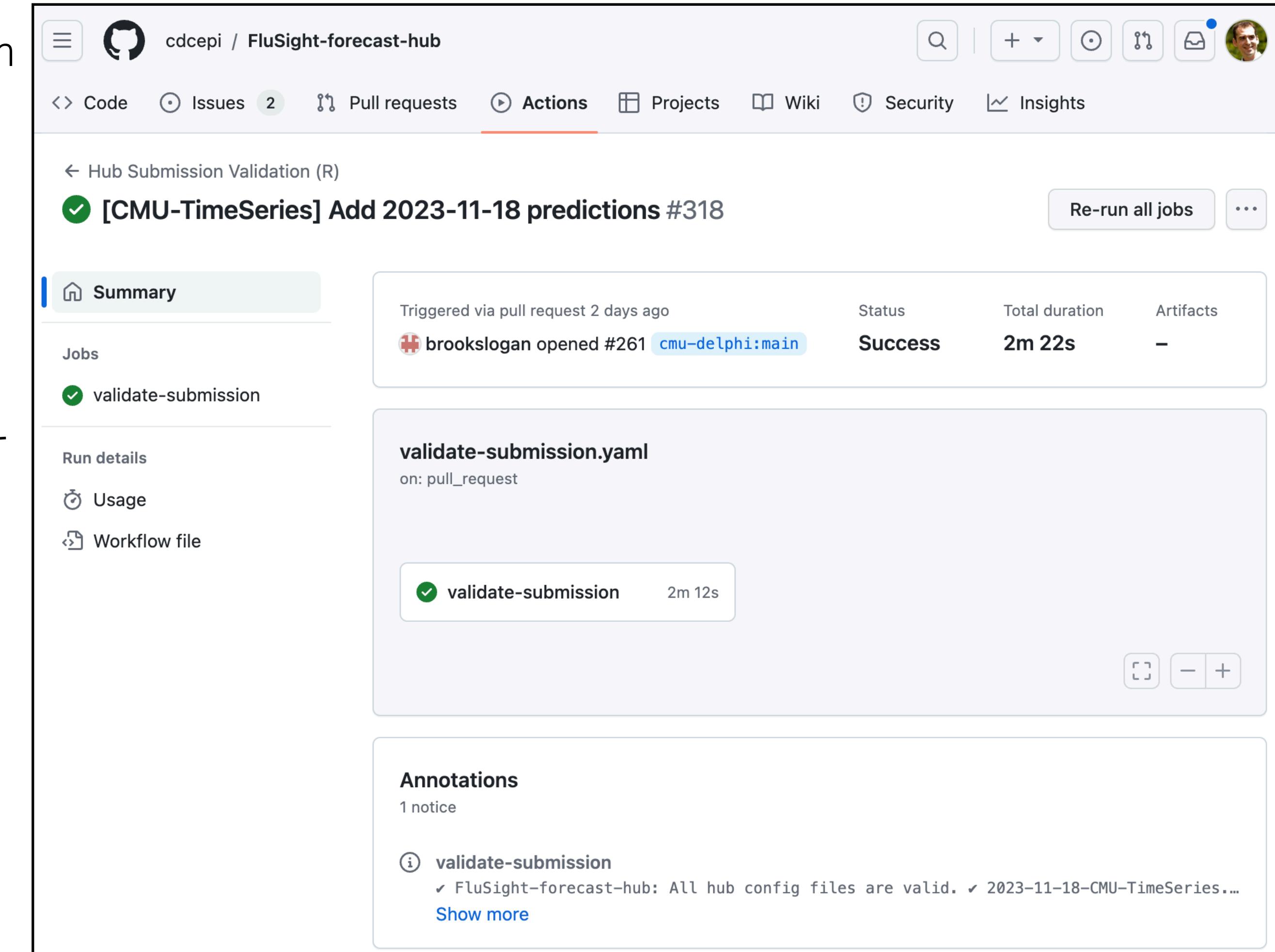
```
## build some ensembles  
> library(hubEnsembles)  
> mean_ens <- simple_ensemble(model_outputs)  
  
> median_ens <- simple_ensemble(model_outputs, agg_fun = "median")
```

# Progress on hubverse goals

- Clear definition of a data format for model output
- Structural requirements for a hub file/directory system
- Open-source tools to access and work with data
  - hubUtils
  - hubVis
  - hubEnsembles
  - hubValidations
  - hubCI
- Documentation of concepts and tools
- (...) Template hubs and continuous integration workflows
- (...) Model containerization for interfacing with hubs
- (...) Dashboards and interactive real-time visualizations



# hubverse



The screenshot shows a GitHub Actions pipeline for the repository `cdcepi / FluSight-forecast-hub`. The pipeline has triggered via a pull request 2 days ago. A job named `validate-submission` was run by user `brookslogan` on branch `#261 cmu-delphi:main`. The status is **Success** with a total duration of **2m 22s**. The workflow file is `validate-submission.yaml`, which runs on `pull_request`. The annotations section shows 1 notice for the `validate-submission` job, stating that all hub config files are valid.



# Thank you!

With acknowledgments for ideas and slide contributions to  
collaborators at UMass-Amherst,  
all contributors to the Hubverse project, CDC collaborators  
(Matt Biggerstaff, Michael Johansson, Rebecca Borcherding, and others),  
and all contributors to modeling hubs.

Funding from NIGMS, US CDC Influenza Division and US CDC CFA.