

Global.health: a data science initiative for the curation of detailed public health data

ABSTRACT

To track and respond to infectious disease outbreaks like COVID-19, we need epidemiological data that is:

- Detailed for each individual case, to high spatial and temporal resolution.
- Updated in **real-time**, to reflect a rapidly evolving situation.
- Standardized across countries, for cross-country and global comparisons.
- Open-access to health researchers regardless of geography.

As the COVID-19 pandemic emerged, standardized 'line-list' data was absent. **Global.health build** technology to collate disparate linelist datasets from across the world into a single online platform, which allows both visualisation and easy access to real-time epidemiological data.





Curator Portal used for source tagging, manually adding cases and data verification. Automated Python parsers run on weekly schedule, powered by AWS Batch. Code on <u>GitHub</u>.

APPLICATIONS OF OUR DATA



Tracking key dates and estimating incubation periods²

Mapping outbreak events²

FUTURE WORK & RESEARCH

- Integrating variant, vaccination, clinical and policy data
- Open access data whilst maintaining privacy
- Decentralised data storage and local governance of data (partnerships with local public health agencies)
- Automated regular generation of epidemiological delay distributions
- Estimating transmission dynamics across spatial scales

1 Xu et al (2020), Nature Scientific Data; 2 Kraemer et al (2020), Science; 3 Hellewel et al (2020), Lancet Global Health



Estimating effects of interventions such as quarantines and contact tracing³