



Building a data product in a pandemic


Perspectives from the Oxford COVID-19 Government
Response Tracker

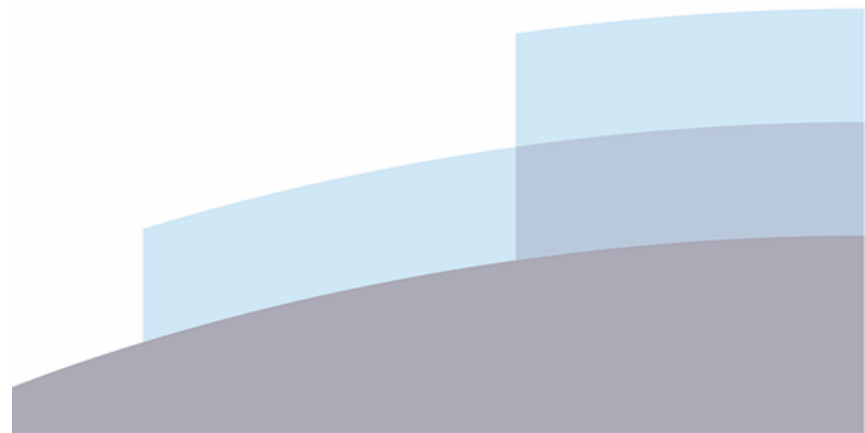
5 July 2023

Toby Phillips

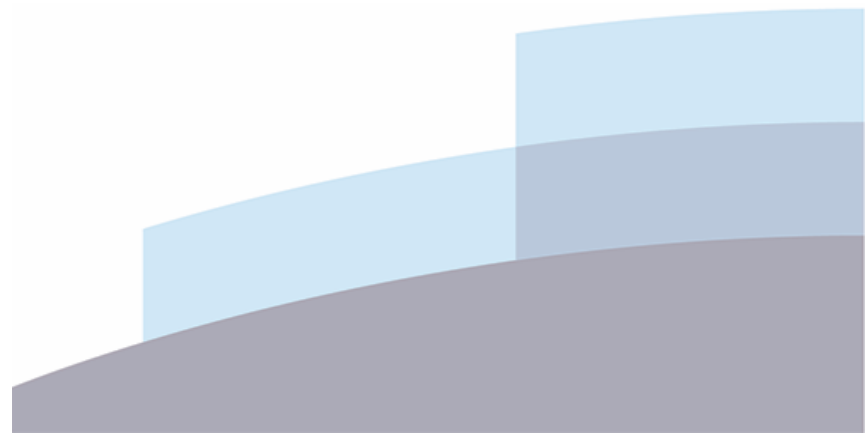
Executive Director, OxCGRT
toby.phillips@bsg.ox.ac.uk

Outline

- 
- 01 WHAT IS OxCGRT?
 - 02 DESIGNING DATA PRODUCTS
 - 03 COLLECTING DATA
 - 04 PUBLISHING DATA
 - 05 THE NEXT PANDEMIC



What we did at the OxCGRT



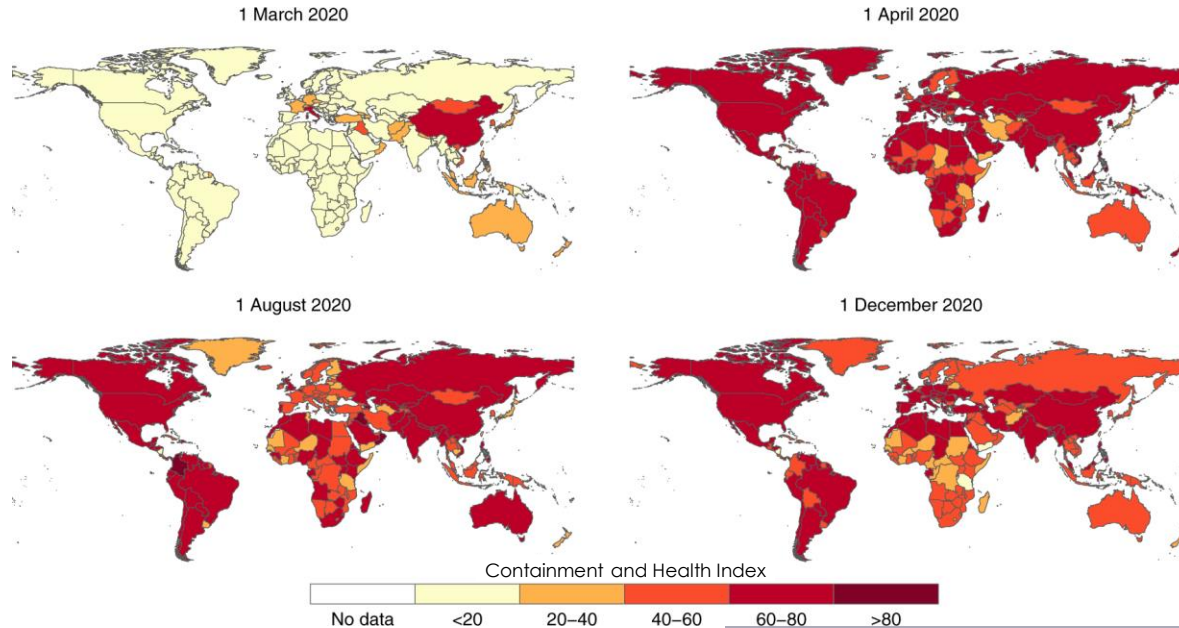
What is the Oxford COVID-19 Government Response Tracker (OxCGRT) for?



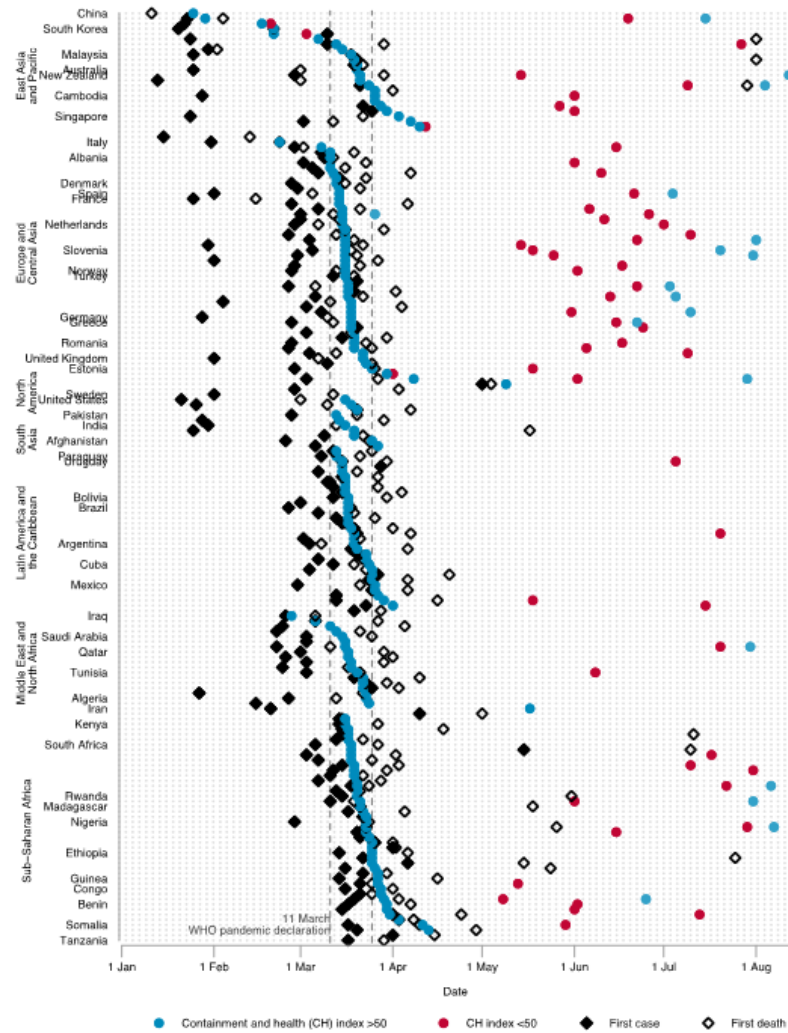
- Back in March 2020, we wanted to know what was happening and answer key research questions on these new pandemic policies.
 - The data didn't exist, so we created it.
- OxCGRT provides a systematic cross-national, cross-temporal measure of how government responses have evolved over the full period of the disease's spread.
 - approx. 25 indicators on closures, economic support, public health measures, and vaccine roll outs.
 - most indicators are recorded on an ordinal scale to capture degree of response.
 - 185+ countries
 - also states/provinces and some cities in Australia, Brazil, Canada, China, India, UK, and USA.

Supporting decision-makers

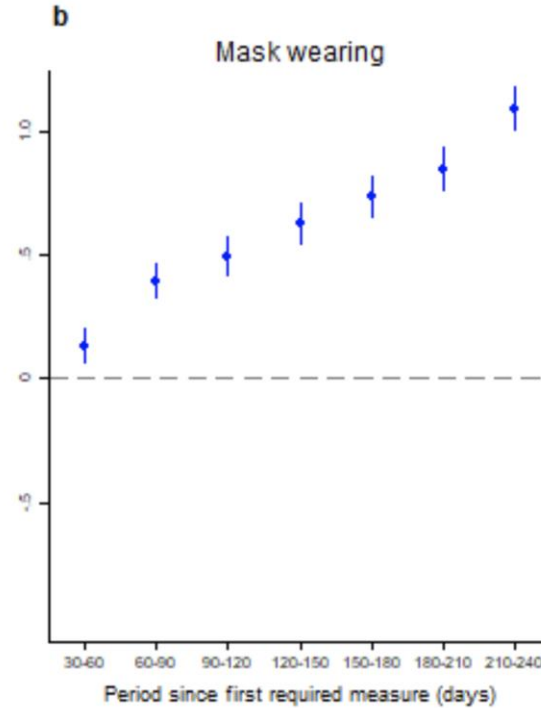
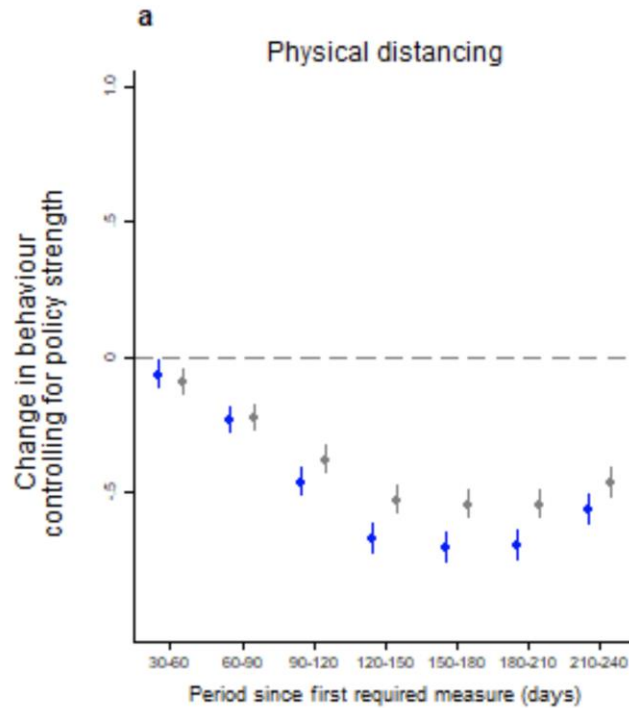
- Policymakers around the world used our data to understand what is happening, in real time.



Clear early convergence

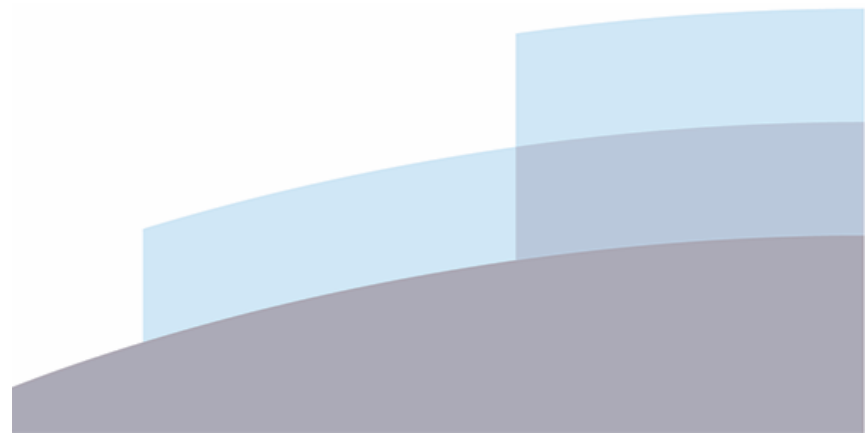


Patterns in compliance



- ◆ Avoidance of gatherings
- ◆ Avoidance of going out

Some challenges designing our system on the fly



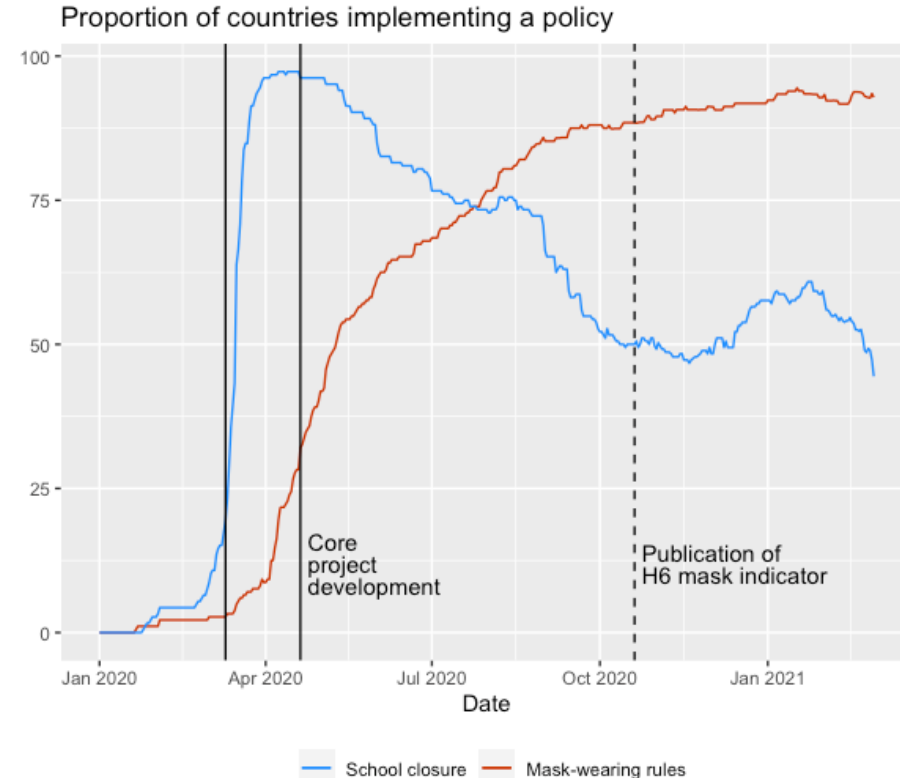
Designing systems in a crisis is hard!



- When thinking about how to make data relevant for policy, you have to ask: what data do people need? What is the right data to even describe the pandemic?
- At the start of the pandemic, we didn't have the answers to these questions – we just had to make our best guess.
- Different jurisdictions report things very differently; eg. definitions of lockdowns or definitions of covid-19 “deaths”. The creators of pandemic datasets have to make judgement calls.
- This meant we had to adapt and design as we go, and in some cases we have been locked into sub-optimal design decisions.

Common approaches change over time

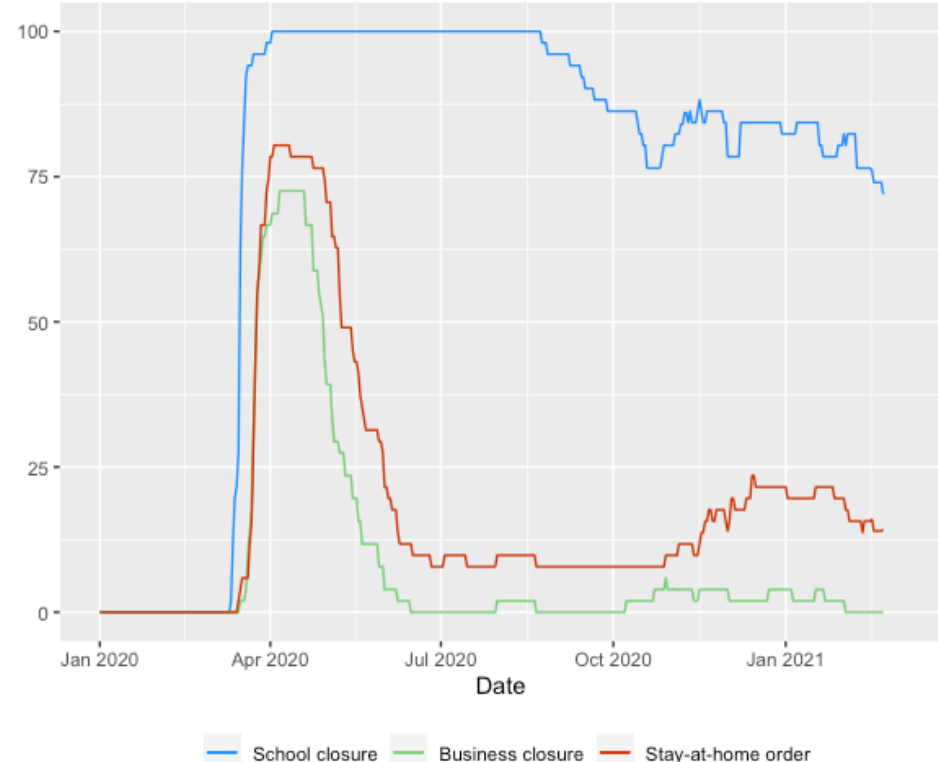
- We launched the project with 7 indicators, and quickly expanded to 11 in the first month.
 - These were based on salient policies at the time.
- It has been a constant game of catch up, adding new indicators.
- Towards the end, if we want to add a new indicator, it requires recording almost 500,000 data points.



Significant subnational variation

- We report the most stringent policy in a jurisdiction.
- For most indicators, we capture subnational variation with a binary variable:
 - 1 = policy applies country-wide
 - 0 = policy is geographically targeted
- But this doesn't tell the user how many people are affected, or *where* the relevant most-stringent policy is.

Proportion of USA states implementing policies



Collecting data requires massive
resources

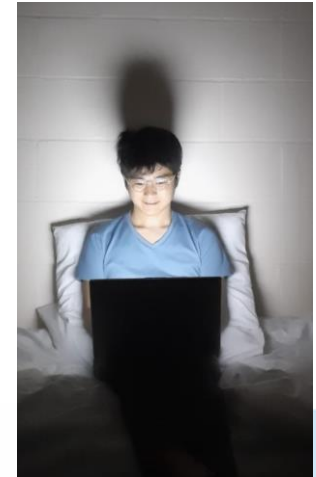
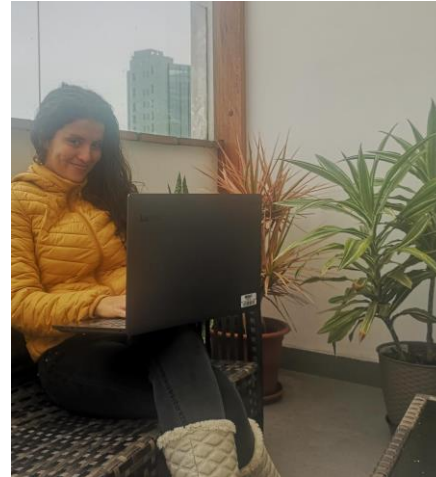
Global data is big



- For one variable over the course of the COVID-19 pandemic, a global dataset requires about 200,000 values.
 - For our policy data: when you multiply this by 20+ indicators (with 100+ variables) as well as hundreds of subnational jurisdictions – it is **tens of millions** of hand-coded datapoints.
- During COVID-19, a lot of NGOs and university groups stepped in to manually collate critical data.
 - Johns Hopkins: epidemiological case + death data
 - Our World in Data: vaccine data
- Some of this is published as machine readable administrative datasets – but not much!

Our approach – volunteers

- For us: no such thing as "machine readable" policy data out there.
- **Collaborative citizen science:** Our OxCGRT data was collected in real time by a team of global volunteers.
 - Several hundred at any given time. Almost 2000 people over the course of the project.
- This is very resource-intensive. It's hard to imagine mobilising this many people-hours in another context. **So who will do it next time?**

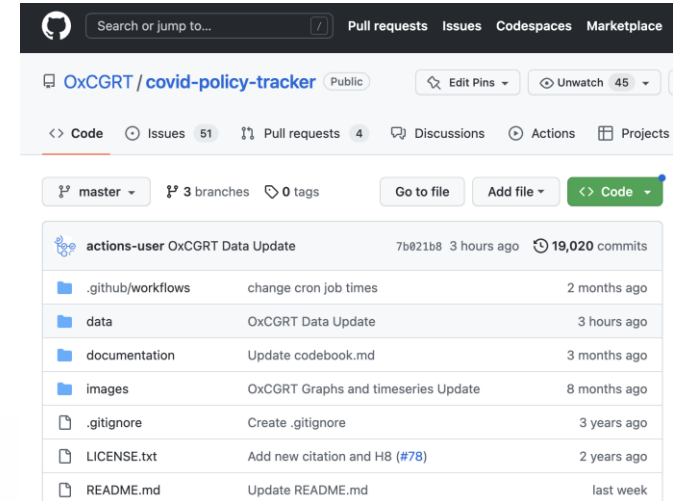


Infrastructure for publishing and using data

A covid data ecosystem emerged organically on GitHub and other platforms

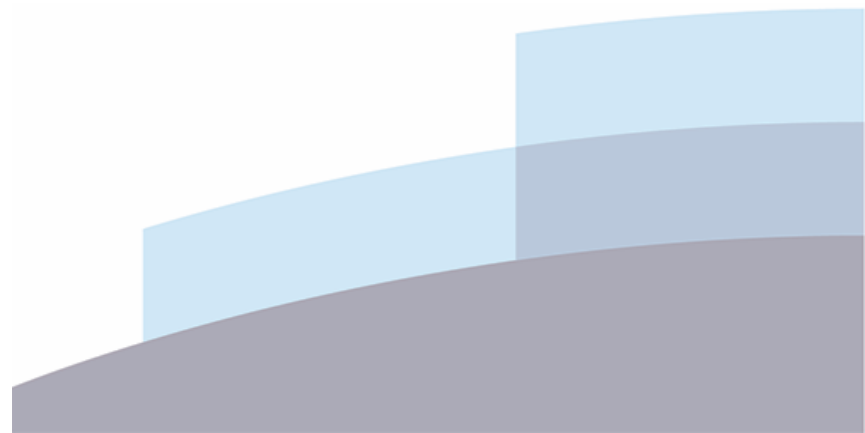


- GitHub became the de facto home of covid datasets – but GitHub has a lot of limitations as a data store!
- Our World in Data became the de facto curated way to explore the data.
- It's great that this data ecosystem emerged, but it could be better.
 - When we worked on the X-PRIZE, the team had to create a custom architecture for pooling and modelling pandemic data – nothing like that exists already.
 - A common infrastructure could lower the barrier to usage – eg. consistent metadata, data structures, and APIs.



Our World
in Data

Preparing for future pandemics



What's next?

- At the Blavatnik School, we have been thinking about data system preparedness for pandemics.
- In an ideal world, we won't need to rely on an ad-hoc collection of actors next time.
 - Data taxonomies
 - Common infrastructure and APIs
 - Global governance of health data
 - Quality assurance



What would a data framework for policy responses to pandemic diseases look like?

January 2023 working paper

<https://doi.org/10.35489/BSG-WP-2023/050>



Thank you.