Overview of the COVID-19 Datasets on Harvard Dataverse

Tao Hu, Postdoc Fellow
The Center for Geographic Analysis (CGA)
1/17/2021

Dataverse Project

The Dataverse® Project

Open source research data repository software

Researchers

Enjoy full control over your data. Receive *web visibility, academic credit,* and *increased citation counts.* A personal Dataverse collection is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. Want to set up your personal Dataverse collection?

Journals

Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data.* Participate in the open data movement by using a Dataverse collection as part of your journal data policy or list of repository recommendations. Want to find out more about journal Dataverse collections?

Institutions

Establish a research data management solution for your community. Federate with a growing list of Dataverse repositories worldwide for increased discoverability of your community's data. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. Want to install a Dataverse repository?

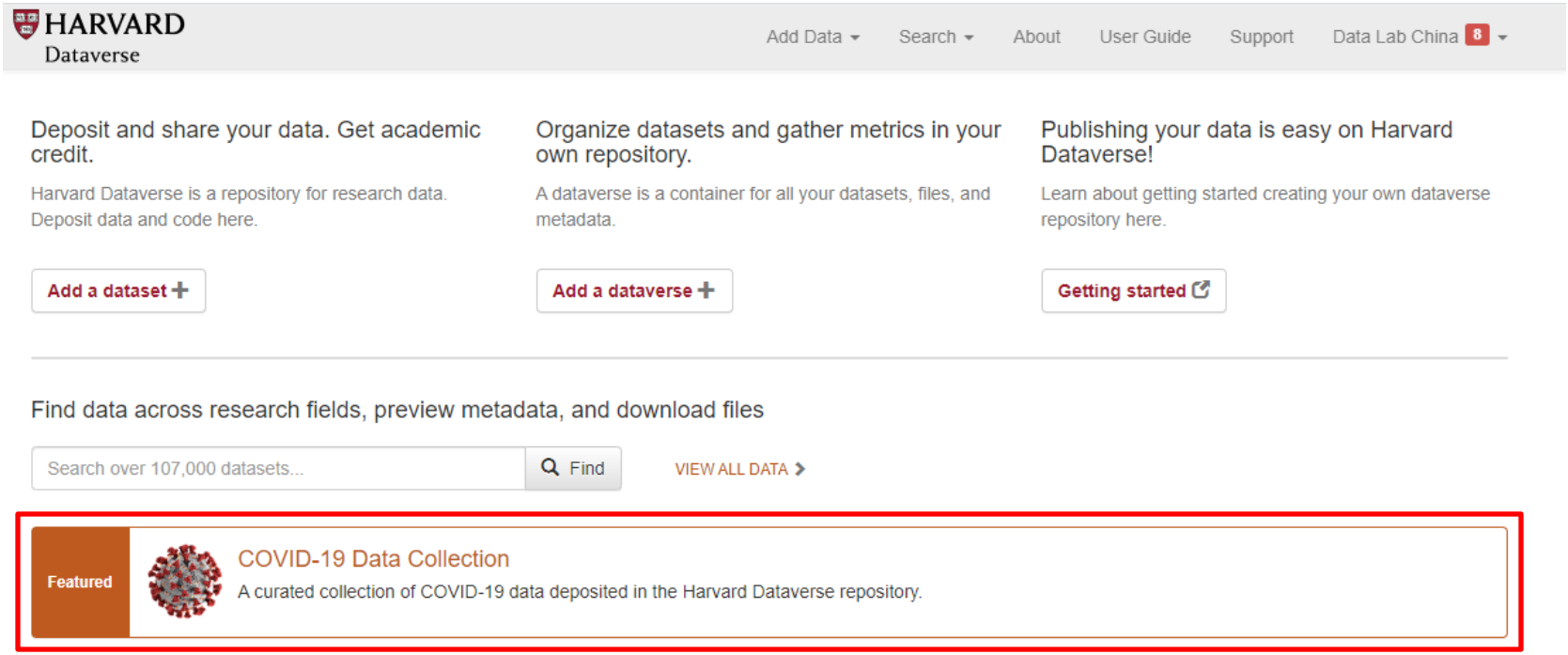
Developers

Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization* and *exploration tools,* or other research and data archival systems with the Dataverse Project. Want to contribute?

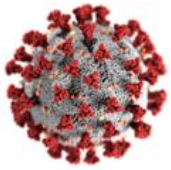
Source: https://dataverse.org

Harvard Dataverse

https://dataverse.harvard.edu



COVID-19 Data Collection



COVID-19 Data Collection

Dataverses (33) **Datasets (337)** **Files (6,961)**

Subject

Social Sciences (230)
Medicine, Health and Life Sciences (177)
Earth and Environmental Sciences (42)
Computer and Information Science (31)
Other (26)
Business and Management (12)
Engineering (7)
Agricultural Sciences (5)
Arts and Humanities (4)
Mathematical Sciences (4)
Law (3)
Chemistry (1)

Author Name

China Data Lab (26)
Abbott, Sam (20)
Allen, Jamie (20)
Funk, Sebastian (20)
Hickson, Joe (20)
Sherratt, Katharine (20)
Collins, Elliott (9)
NSF Spatiotemporal Innovation Center (9)
Parkerson, Doug (9)
Warren, Shana S. (9)
Ngo, Thoai D. (8)
Ahmar, Ansari Saleh (5)
Mu, Ren (4)
Nokes, David James (4)
Ogundokun, Roseline Oluwaseun (4)

Author Affiliation

China Data Lab (30)
Centre for the Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK (20)
Met Office, Exeter, UK (20)
Harvard University (14)
Population Council (12)
NSF Spatiotemporal Innovation Center (10)
Innovations for Poverty Action (9)
Landmark University Omu Aran, Kwara State (6)
Cornell University (5)
Universitas Negeri Makassar (5)

https://dataverse.harvard.edu/dataverse/covid19

Background

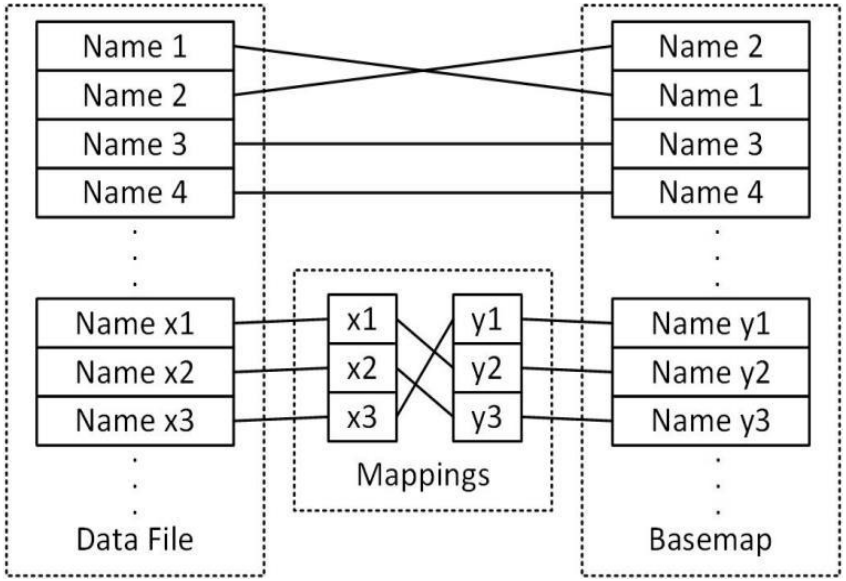
- China Data Lab: NSF Industry-University Cooperative Research Centers (I/UCRC) Program (Harvard)
- NSF's Rapid Response Research (RAPID) Program (Harvard + GMU)

  - Integrating data from different sources
  - Standardizing different data units and formats
  - Matching data with geographic entities for spatial analysis
  - Publishing data in permanent collections
  - Supporting collaborative research on COVID-19 studies

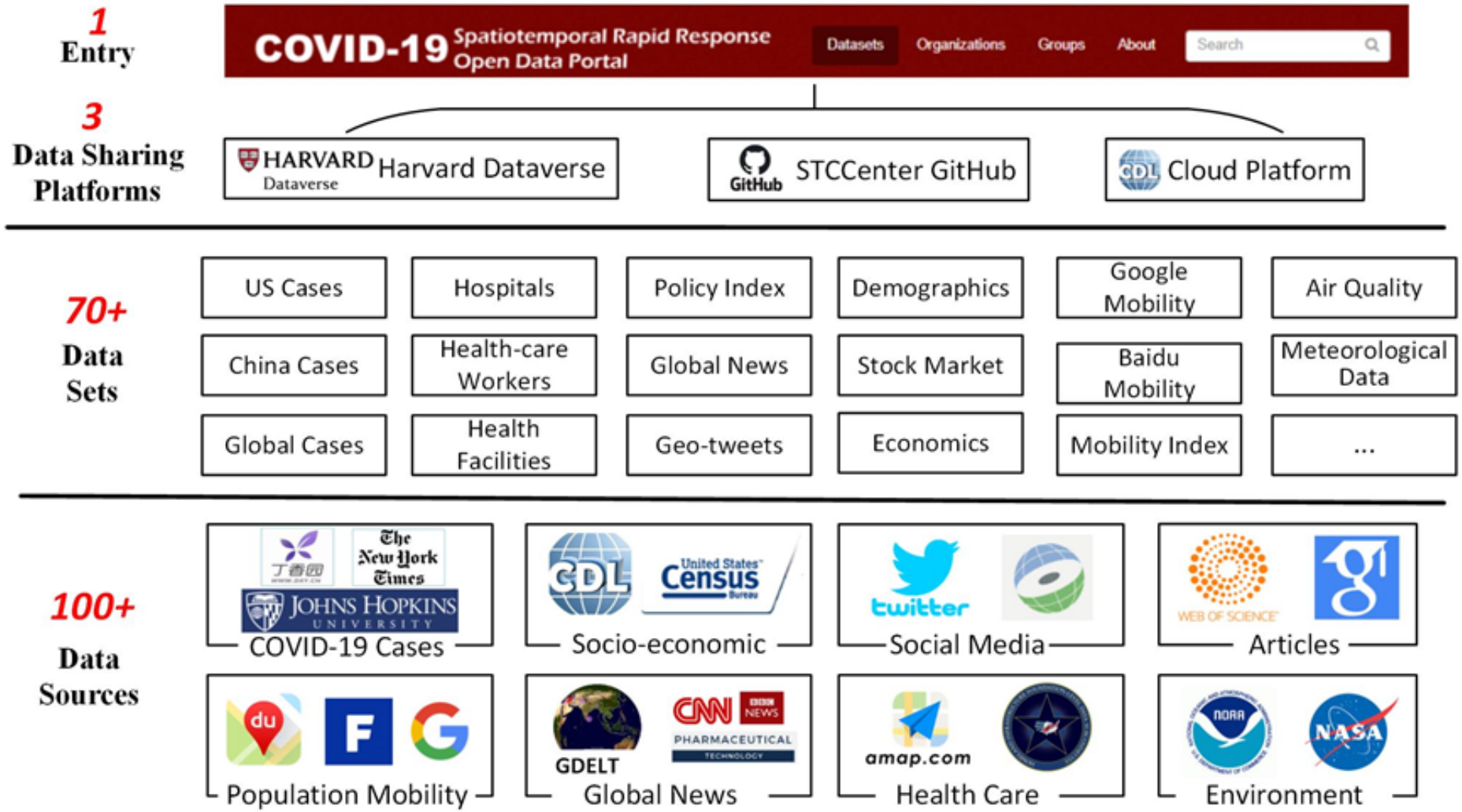
Data Integration with Basemap

- ❑ **Data standardization and association**
- ❑ **Data Spatialization**

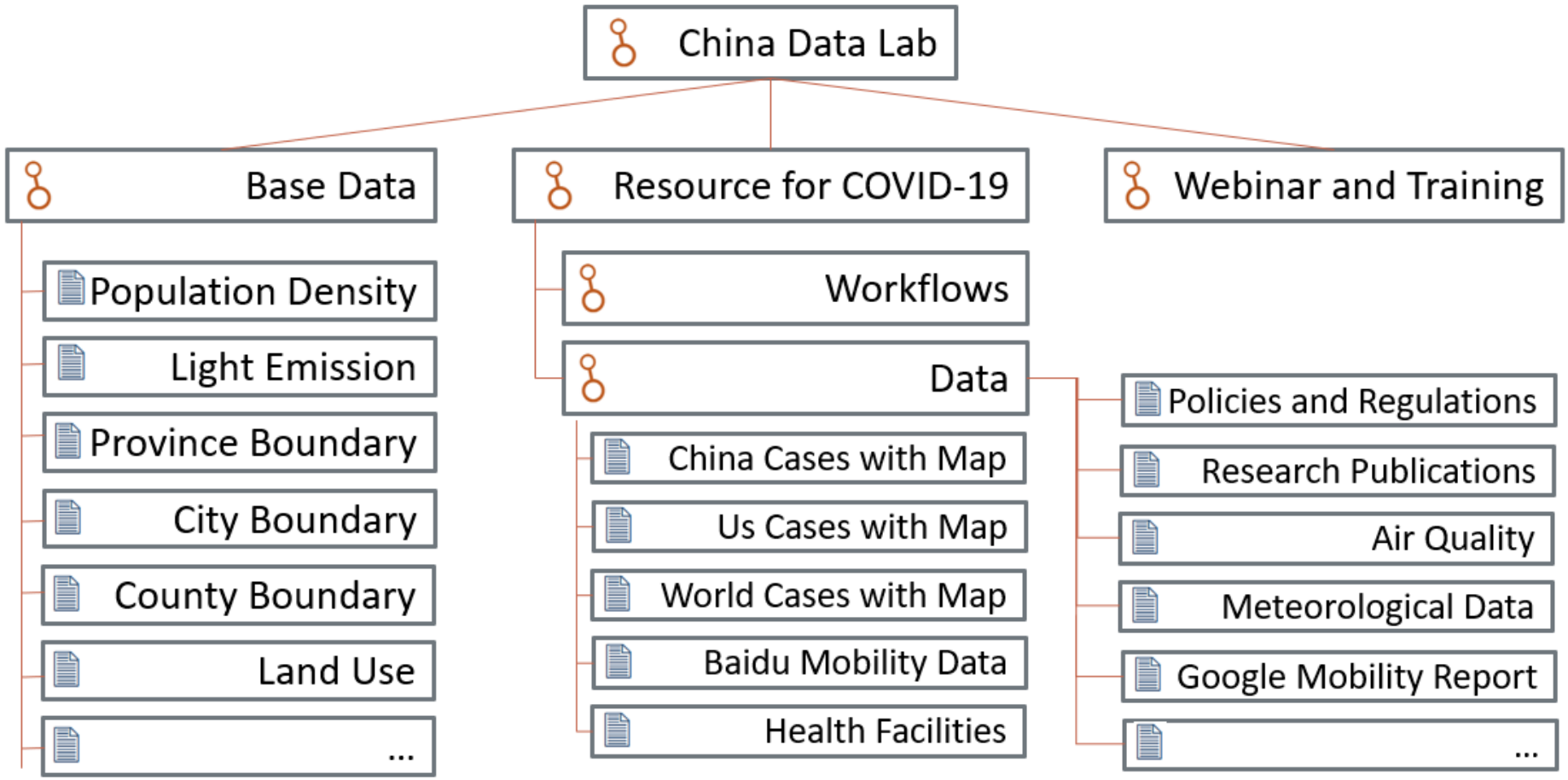
Administrative Regions Mapping



Data Framework



Data Sharing on Harvard Dataverse



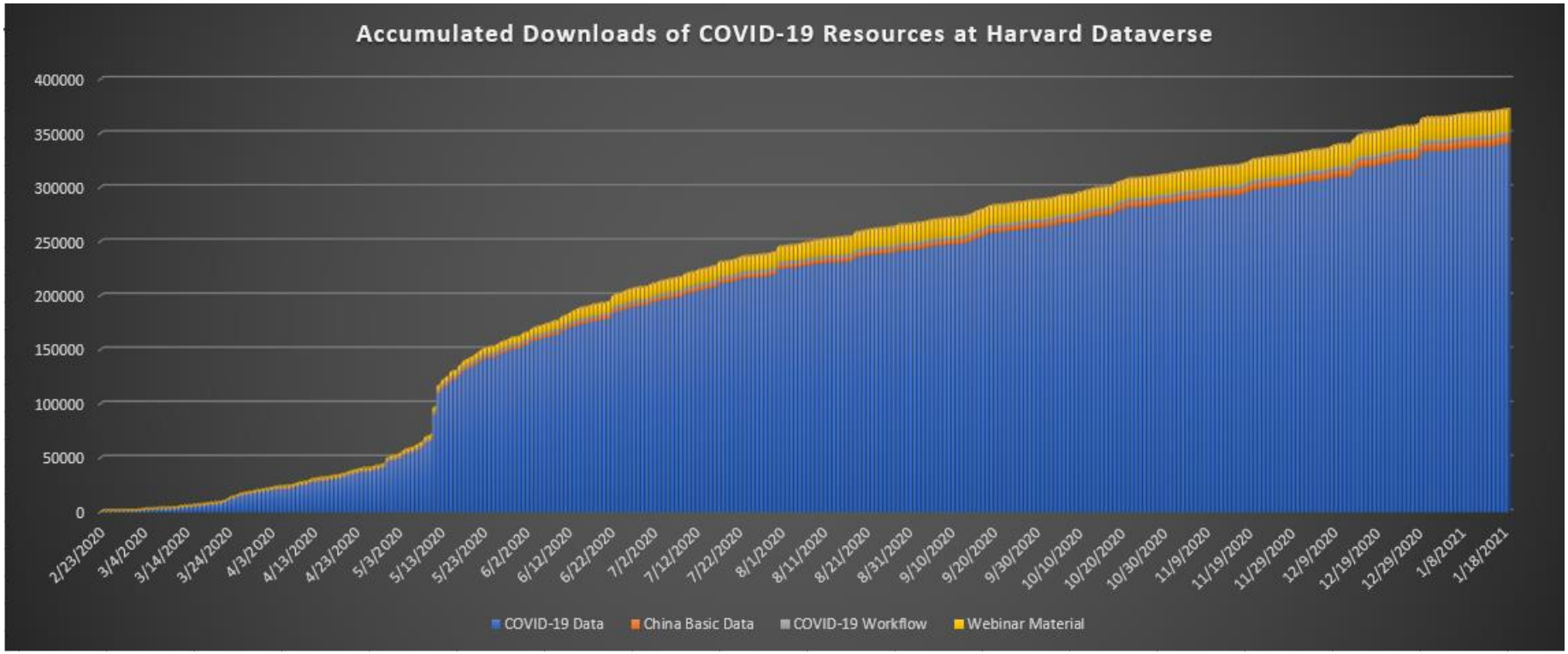
Datasets Structures of Dataverse China Data Lab

Data Publishing on Harvard Dataverse

ID	Dataset	Region	DOI	Start Date	Update
1	World COVID-19 Daily Cases with basemap	Global	https://doi.org/10.7910/DVN/L20LOT	01/22/2020	Weekly
2	US COVID-19 Daily Cases with basemap	US	https://doi.org/10.7910/DVN/HIDLTK	01/22/2020	
3	China COVID-19 Daily Cases with basemap	China	https://doi.org/10.7910/DVN/MR5IJN	01/14/2020	
4	Human Mobility Data	China, US	https://doi.org/10.7910/DVN/FAEZIO	01/01/2020	
5	Health Facilities Data	US, China	https://doi.org/10.7910/DVN/KRSGT3		
6	News and Policy Data	Global	https://doi.org/10.7910/DVN/OAM2JK	01/01/2020	Weekly
7	Scholarly Articles	Global	https://doi.org/10.7910/DVN/MHL8JC	01/01/2020	
8	Meteorological Data	China	https://doi.org/10.7910/DVN/XETLSS	12/01/2019	
9	Flight Data	Global	CDL Cloud Platform	12/01/2019	
10	Social Media Data	Global	CDL Cloud Platform	01/01/2020	
11	Global News	Global	CDL Cloud Platform	01/01/2020	

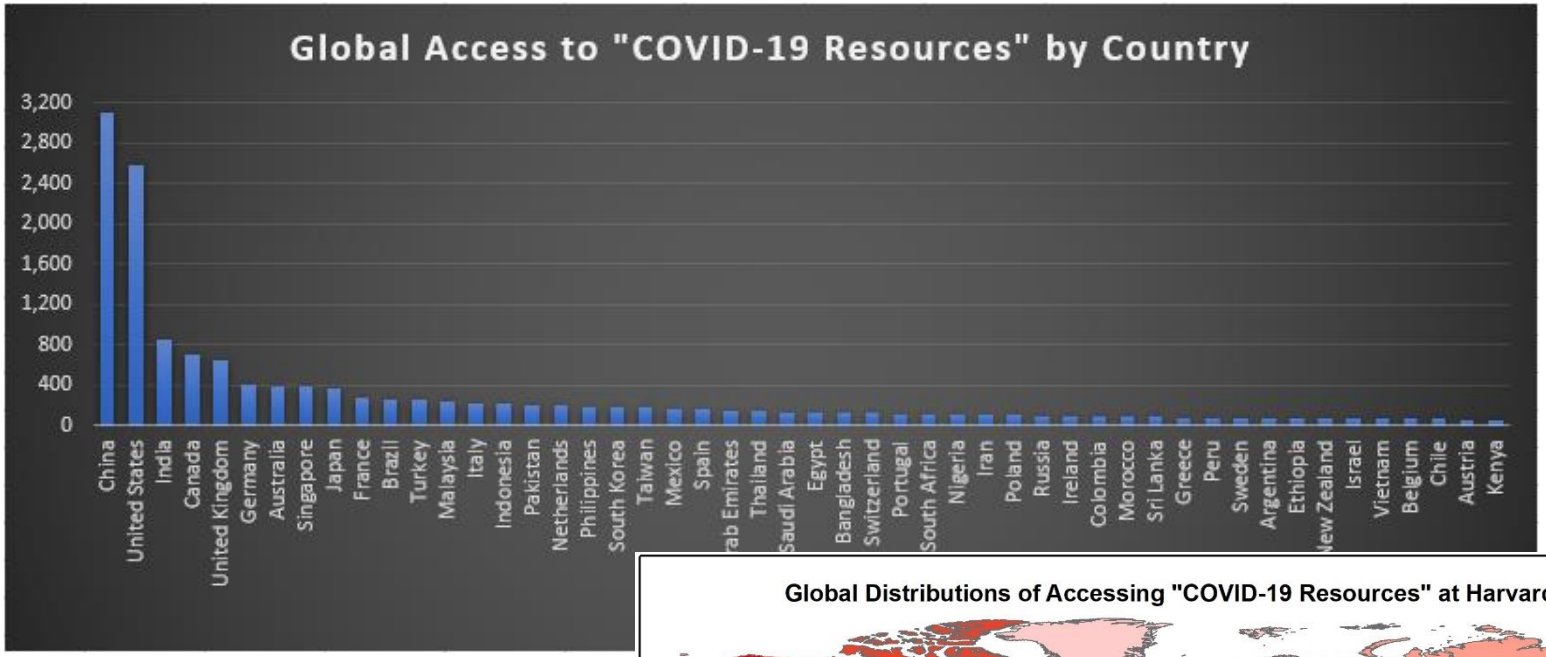
COVID-19 Data Access

Dataverses (7) Datasets (85) Files (1,335)

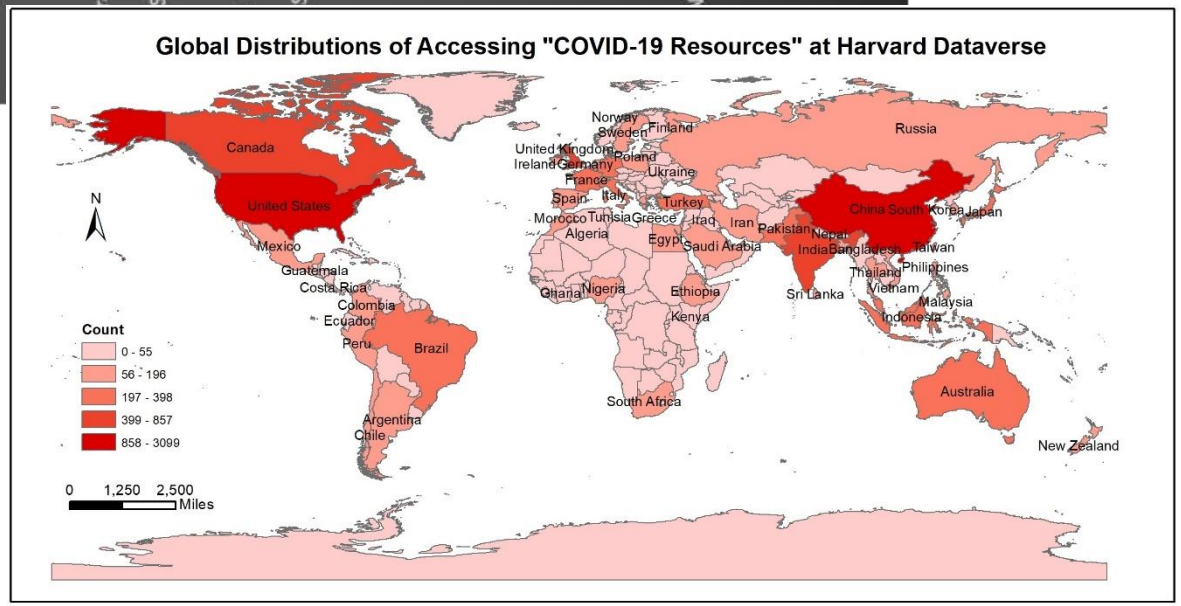


Over ***370,000*** accumulated download times as of January 18, 2021

COVID-19 Data Access



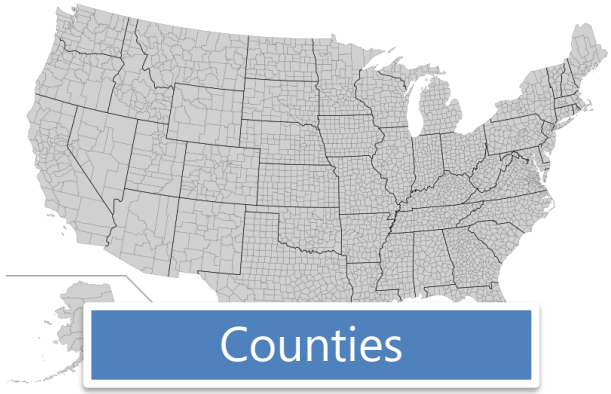
Over ***150*** countries have accessed our COVID-19 data sources as of January 18, 2021



COVID-19 Cases in the US

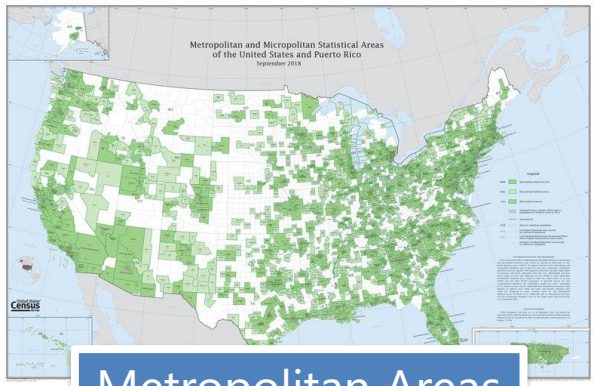
States

- **Source**: John Hopkins University
- **DOI**: https://doi.org/10.7910/DVN/HIDLTK



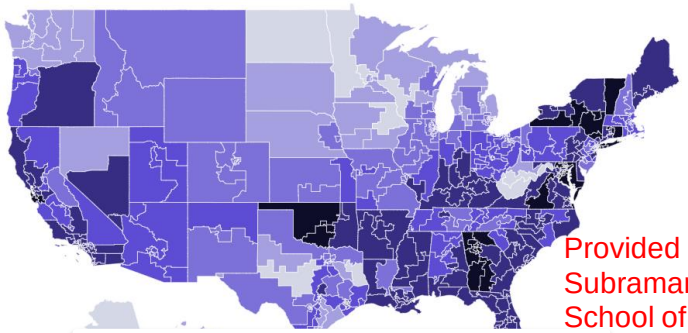
Counties

- **Source**: USAFacts
- **DOI**: https://doi.org/10.7910/DVN/HIDLTK



Metropolitan Areas

- Aggregated from county cases
- **DOI**: https://doi.org/10.7910/DVN/5B8YM8



Provided by
Subramanian, S.V.,
School of Public Health

Congressional Districts

- Census track data + County cases
- **DOI**: https://doi.org/10.7910/DVN/KPI7WW

COVID-19 Cases in the US

US COVID-19 Daily Cases with Basemap **Published** **Admin** **Contributor**

Dec 24, 2020 - Data

China Data Lab, 2020, "US COVID-19 Daily Cases with Basemap", https://doi.org/10.7910/DVN/HIDLTK, Harvard Dataverse, V47, UNF:6:4EIxrvPuiiTKBCxzWoiRpQ== [fileUNF]

Updated to December 22, 2020. It contains COVID-19 Daily Cases with US basemap, including state, county-level, and metropolitan data.

COUNTY_MAP.zip
Shapefile as ZIP Archive - 38.0 MB - Mar 13, 2020 - 1,436 Downloads
MD5: cc7f305a033c6dd0fe017a1828faf97b

STATE_MAP.zip
Shapefile as ZIP Archive - 14.8 MB - Mar 13, 2020 - 961 Downloads
MD5: 68dcec3ad873aedb1a070e28ad46c8dc

us_county_confirmed_cases.tab
Tabular Data - 3.7 MB - Dec 24, 2020 - 51 Downloads
351 Variables, 3143 Observations - UNF:6:70Ohrl1yWeiuBqlm+Zc+4Q==

us_county_deaths_cases.tab
Tabular Data - 2.7 MB - Dec 24, 2020 - 30 Downloads
351 Variables, 3143 Observations - UNF:6:/SzICmKUosSI50OGN1yBWg==

us_state_confirmed_case.tab
Tabular Data - 93.6 KB - Dec 24, 2020 - 27 Downloads
349 Variables, 51 Observations - UNF:6:Sj7J0fARdnU3bWrfHqZkQQ==

us_state_deaths_case.tab
Tabular Data - 70.5 KB - Dec 24, 2020 - 18 Downloads
349 Variables, 51 Observations - UNF:6:5RKZd0mO4XyxiDqnLB5+SA==

Human Mobility Datasets

Google Community Mobility Report

- **Source:** https://www.google.com/covid19/mobility/
- **Format:** tabular/pdf
- **Admin Level**: country, state, and county
- **DOI**: https://doi.org/10.7910/DVN/1CLYWS

Apple Mobility Trends Report

- **Source:** https://www.apple.com/covid19/mobility
- **Format:** tabular/pdf
- **Admin Level**: country, state, and city
- **DOI**: https://doi.org/10.7910/DVN/E9EIB6

Foursquare Mobility Reports

- **Source:** https://visitdata.org
- **Format:** tabular
- **Admin Level**: state, and county
- **DOI**: https://doi.org/10.7910/DVN/PFLAH4

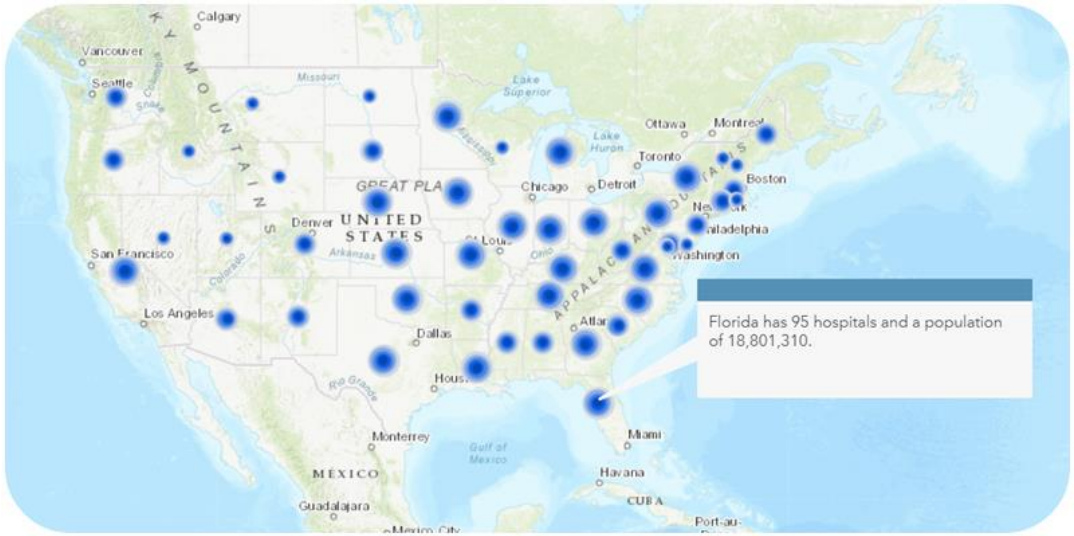
Descartes Lab Mobility Data

- **Source:** https://github.com/descarteslabs/DL-COVID-19
- **Format:** tabular/pdf
- **Admin Level**: state, county
- **DOI**: https://doi.org/10.7910/DVN/W2FT3K

Health Facilities

- **Provider:** China Data Lab
- **Data Source**： Homeland Infrastructure Foundation
- **Format**： table (csv)
- **DOI**: https://doi.org/10.7910/DVN/KRSGT3
- **Last Update**: **December 23, 2020**

The facilities include **hospitals**, **medical centers**, federally qualified **health centers**, home **health services** and **nursing homes**. Hospitals have been categorized into **children**, **chronic disease**, **critical access**, **general acute care**, **long term care**, **military**, **psychiatric**, **rehabilitation**, **special**, and **women** based on the range of the available values from the various sources after removing similarities.



Social Media

Geotweet Archive v2.0: A global social media record with high granularity for geography and time.

- **Provider:** The Center for Geographic Analysis (CGA)
- **Data Source**： Twitter
- **Format**： table (csv)
- **DOI**: https://doi.org/10.7910/DVN/3NCMB6

- Extends from 2010 to the present
- All records geotagged by GPS or user-designated place
- Estimated size 10 billion records
- Updated daily
- Stored on Harvard FASRC

Geotweets Data Visualization with Omnisci

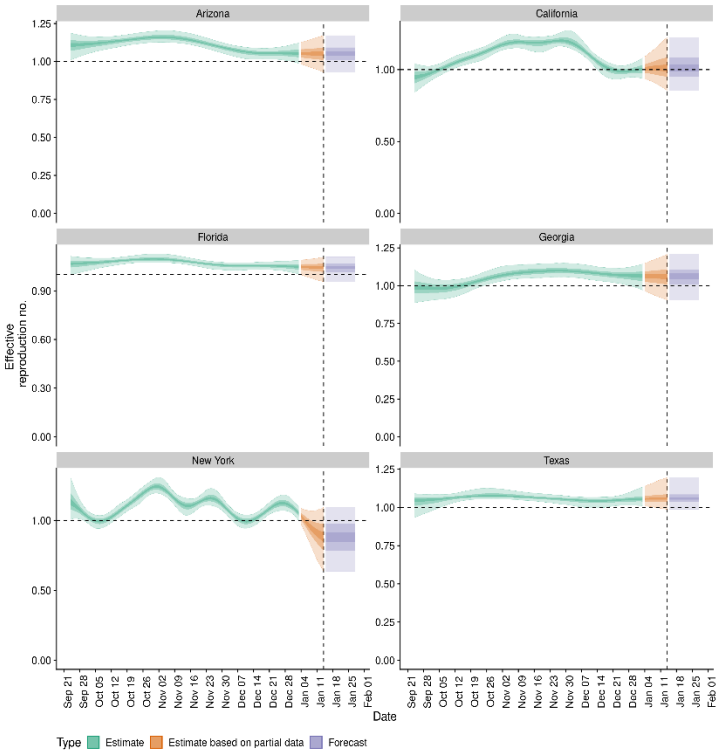
Reproduction Rate

National and Subnational Estimates of the Covid 19 Reproduction Number (R) for the United States of America Based on Test Results

cmmid centre for mathematical modelling of infectious diseases

- **Provider:** Centre for Mathematical Modeling of Infections Diseases
- **Format：** table (csv)
- **DOI**: https://doi.org/10.7910/DVN/BZ7FPH
- **Last Update： Jan 15, 2021**

Identifying changes in the **reproduction number**, **rate of spread**, and **doubling time** during the course of the COVID-19 outbreak whilst accounting for potential biases due to delays in case reporting both **nationally** and **subnationally** in the United States of America. These results are impacted by changes in testing effort, increases and decreases in testing effort will increase and decrease reproduction number estimates respectively.



Data for Publication Replication

"Replication Data" ⟹ Datasets (103)

https://dataverse.harvard.edu/dataverse/covid19?q=%22Replication+Data%22&types=datasets

Replication Data for: Confirmed and Unreported COVID-19-Like Illness Death Counts: An Assessment of Reporting Discrepancy

Sep 16, 2020 - Mazbahul Ahamad Dataverse

Ahamad, Mazbahul, 2020, "Replication Data for: Confirmed and Unreported COVID-19-Like Illness Death Counts: An Assessment of Reporting Discrepancy", https://doi.org/10.7910/DVN/H9RHWN, Harvard Dataverse, V3, UNF:6:3LcGi+Y0zFgzikGaj5N8oA== [fileUNF]

... **Replication Data** for: Confirmed and Unreported Covid-19 Death Counts: An Assessment of Reporting Discrepancy ...

Replication Data for: Markedly enhanced levels of peroxyacetyl nitrate (PAN) during COVID-19 in Beijing

Jun 30, 2020 - Harvard Dataverse

Qiu, Yulu, 2020, "Replication Data for: Markedly enhanced levels of peroxyacetyl nitrate (PAN) during COVID-19 in Beijing", https://doi.org/10.7910/DVN/LUX8JX, Harvard Dataverse, V1, UNF:6:fmffivXEC7rcXBOESSPeyw== [fileUNF]

... **Replication Data** for: Markedly enhanced levels of peroxyacetyl nitrate (PAN) during COVID-19 in Beijing ...

Replication Data for: How do Americans Want Elections to be Run During the COVID-19 Crisis?

Nov 10, 2020 - Research & Politics Dataverse

Lockhart, Mackenzie; Kousser, Thad; Hill, Seth J.; Merolla, Jennifer; Romero, Mindy, 2020, "Replication Data for: How do Americans Want Elections to be Run During the COVID-19 Crisis?", https://doi.org/10.7910/DVN/3GFZ9L, Harvard Dataverse, V1, UNF:6:TtNNDzBD+ZiFr1IqyF3ttQ== [fileUNF]

... **Replication Data** for: How do Americans Want Elections to be Run During the COVID-19 Crisis? ...

Replication data for: Coronavirus Fears and Macroeconomic Expectations

Apr 28, 2020 - Review of Economics and Statistics Dataverse

Binder, Carola, 2020, "Replication data for: Coronavirus Fears and Macroeconomic Expectations", https://doi.org/10.7910/DVN/6LPBJF, Harvard Dataverse, V1

Binder, Carola, (2020) "Coronavirus Fears and Macroeconomic Expectations." Review of Economics and Statistics 102:4, 721-730.

Data for Publication Replication

https://dataverse.harvard.edu/dataverse/covid19?q=%22Replication+Data%22&types=datasets

Data

border_orientation_2018_country_means_controls.tab
Tabular Data - 10.8 KB - Jul 20, 2020 - 22 Downloads
7 Variables, 152 Observations - UNF:6:S1n6WQdNM9eYGbFW73jozA==

border_orientation_annual_means.tab
Tabular Data - 465 B - Jul 20, 2020 - 10 Downloads
2 Variables, 19 Observations - UNF:6:+qpTTQpxdK9+ZoGUeMWehg==

Code

kenwick_simmons_io_covid.R
R Syntax - 14.2 KB - Jul 20, 2020 - 21 Downloads
MD5: dbbc6f09346719087f729553d6b96d67

Result

fig3.pdf
Adobe PDF - 4.9 KB - Jul 20, 2020 - 6 Downloads
MD5: aa220ccf702d2d90cf3a9731fcbd9593

Publication

Pandemic Response as Border Politics

Michael R. Kenwick and Beth A. Simmons

Abstract Pandemics are imbued with the politics of bordering. For centuries, border closures and restrictions on foreign travelers have been the most persistent and pervasive means by which states have responded to global health crises. The ubiquity of these policies is not driven by any clear scientific consensus about their utility in the face of myriad pandemic threats. Instead, we show they are influenced by public opinion and preexisting commitments to invest in the symbols and structures of state efforts to control their borders, a concept we call *border orientation*. Prior to the COVID-19 pandemic, border orientation was already generally on the rise worldwide. This trend has made it convenient for governments to "contain" the virus by externalizing it, rather than taking costly but ultimately more effective domestic mitigation measures. We argue that the pervasive use of external border controls in the face of the coronavirus reflects growing anxieties about border security in the modern international system. To a great extent, fears relating to border security have become a resource in domestic politics—a finding that does not bode well for designing and implementing effective public health policy.

Kenwick, Michael; Beth Simmons, 2020, "Replication Data for: Pandemic Response as Border Politics", https://doi.org/10.7910/DVN/J0PGNY, Harvard Dataverse, V1, UNF:6:c9PM5yolQ20XYZCw/GWf5w== [fileUNF]

Acknowledge

- China Data Institute
- Harvard Dataverse Team
- GMU Spatiotemporal Center
- Wuhan University

THANK YOU !

COVID19 Data Collection on Harvard Dataverse

https://dataverse.harvard.edu/dataverse/covid19

Resources for Coronavirus Study

http://chinadatalab.net