China Data Institute

Center for Geographic Analysis
Harvard University

社会地理计算联合研究中心
Geocomputation Center for Social Science

Introduction to "Resources for COVID-19"

Tao Hu, Post-doctoral Fellow, Harvard University

2020-03-24

Outline

① COVID-19 Open Datasets

② Framework of “Resources for COVID-19”

③ “Resources for COVID-19” at Harvard Dataverse

④ “Resource for COVID-19” at CDL Cloud Platform

COVID-19 Open Datasets

ID	Name	Author/Provider	Region	Update Frequency	Data Type	Data Source
1	COVID-19 Cases	JHU CSSE (John Hopkins University)	Global	Hours	csv	• World Health Organization (WHO) • DXY.cn. Pneumonia. 2020. • BNO News • National Health Commission of the People's Republic of China (NHC): • China CDC (CCDC) • Hong Kong Department of Health • Macau Government • Taiwan CDC • US CDC • Government of Canada • Australia Government Department of Health • European Centre for Disease Prevention and Control (ECDC) • Ministry of Health Singapore (MOH) • Italy Ministry of Health
2	Geographic distribution of COVID-19 cases worldwide	European Centre for Disease Prevention and Control	Global	Daily	xlsx/R	• European Centre for Disease Prevention and Control
3	COVID-19/2019-nCoV Time Series Infection Data Warehouse	Isaac Lin	China	Hours	csv/json	• DXY.cn. Pneumonia. 2020.
4	nCov2019: An R package with real-time data, historical data and Shiny app	Guangchuang Yu, Xijin Ge, et al.	China, South Korea, US, Japan Iran, Italy, UK and Germany	Hours	R	• Tencent SARS-COV-2 website • DXY.cn. Pneumonia. 2020. • Wuhan-2019-nCoV GitHub repository • National Health Commission of the People's Republic of China • Jinritoutiao
5	Dati COVID-19 Italia	Presidency of the Council of Ministers-Department of Civil Protection	Italy	Days	csv	• Presidency of the Council of Ministers-Department of Civil Protection

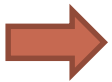
COVID-19 Open Datasets

ID	Name	Author/Provider	Region	Update Frequency	Data Type	Data Source
6	COVID19-Korea	Sang Woo Park https://github.com/parksw3/COVID19-Korea	South Korea	Days	xlsx	• KCDC press release • Official Website of the Seoul Metropolitan City Government • Official Website of the Busan Metropolitan City Government • Official Blog of Daegu Metropolitan City Government • Official Website of the Gwangju Metropolitan City Government • Official Website of the Ulsan Metropolitan City Government • Official Website of Gyeonggi Province Government
7	Dati COVID-19 Italia	Presidency of the Council of Ministers-Department of Civil Protection	Italy	Days	csv	• Presidency of the Council of Ministers-Department of Civil Protection
8	OpenCOVID19-fr	Antoine Augusti, Colin Maudry, et al. https://github.com/opencovid19-fr/data	France	Hours	csv/json	• Public health France • Regional Health Agencies • Videos of the Ministry of Solidarity and Health
9	Europe Situation Updates	midas-network	UK	Days	csv	N/A

COVID-19 Open Datasets

❑ Data Limitations

- Data are scattered at many sources
- Data are not comparable, lack of standard
- Data are not permanent, may be available for a short time only
- Data are not rich in rich spatial information



Data Management and Sharing Platform

❑ Analysis Limitations

- Analyses are not connected with data
- Analyses are not easily replicable
- Analytical tools are hard to maintain
- Computing environment are hard to set up
- Research cycle is time consuming if to start from scratch

Integrated Data Analysis and Results Sharing Platform

Outline

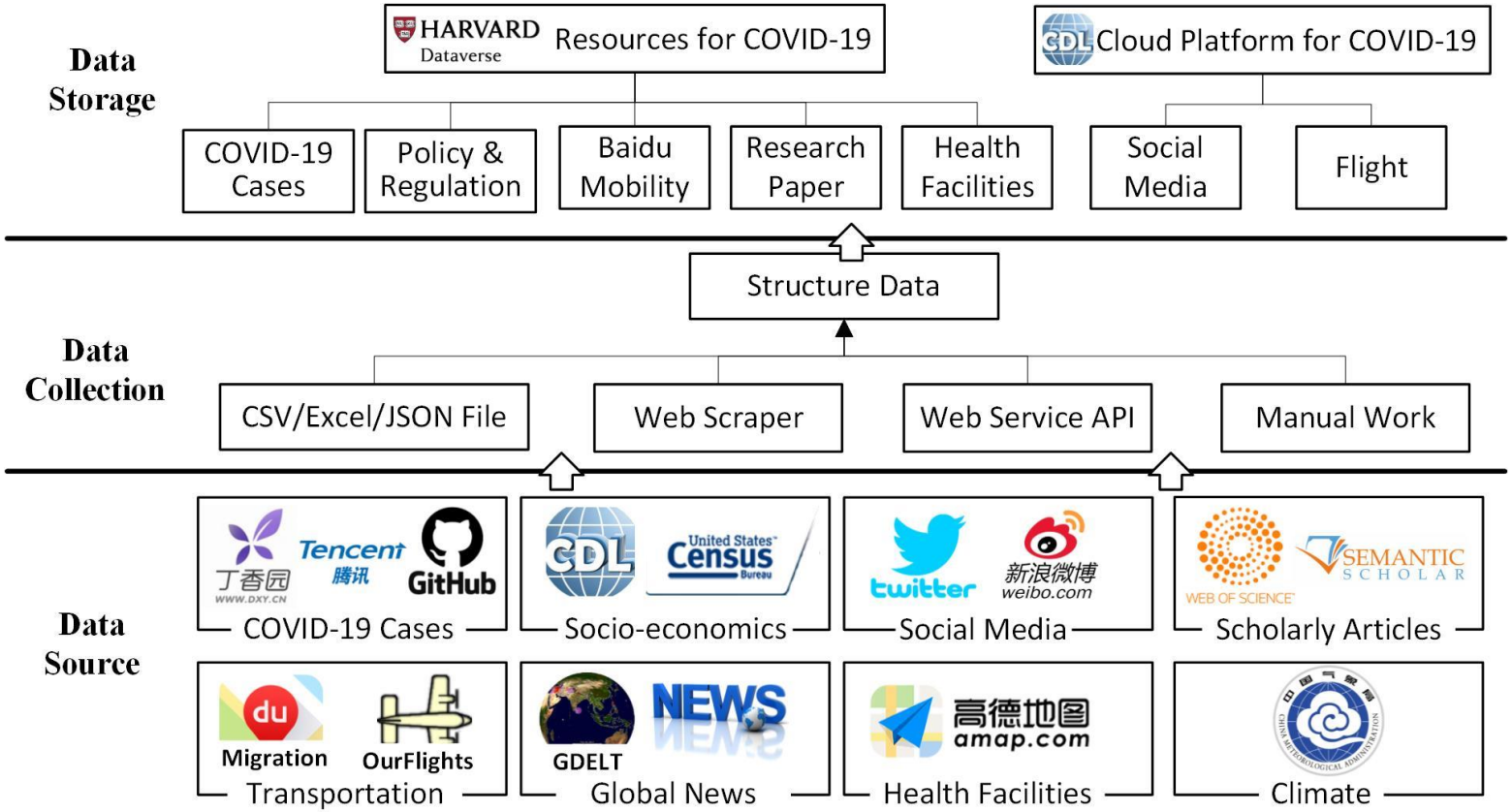
① COVID-19 Open Datasets

② Framework of “Resources for COVID-19”

③ “Resources for COVID-19” at Harvard Dataverse

④ “Resource for COVID-19” at CDL Cloud Platform

Framework of “Resources for COVID-19”



Outline

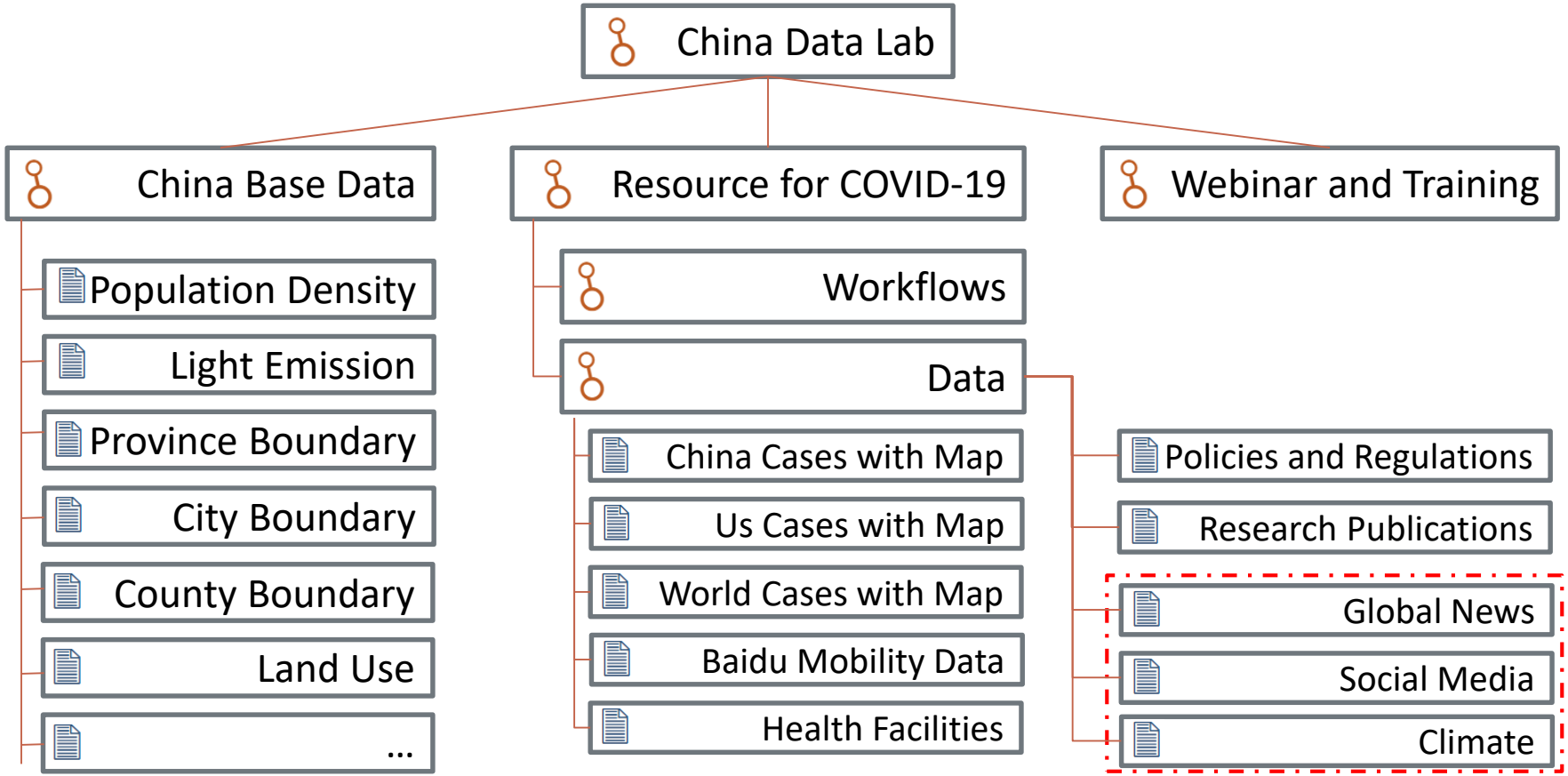
① COVID-19 Open Datasets

② Framework of “Resources for COVID-19”

③ “Resources for COVID-19” at Harvard Dataverse

④ “Resource for COVID-19” at CDL Cloud Platform

“Resources for COVID-19” at Harvard Dataverse



“Resources for COVID-19” at Harvard Dataverse

❑ Updated Data List

数据名称	数据描述	数据类型	文件类型	起始时间	更新时间	负责人员
US COVID-19 Daily Cases with Basemap	It covers US COVID-19 daily cases in :	COVID-19 Cases	csv; shapefile	1/22/2020	3/15/2020	胡涛
China COVID-19 Daily Cases with Basemap		COVID-19 Cases	csv; shapefile	1/14/2020		刘老师；李志斌
World COVID-19 Daily Cases with Basemap		COVID-19 Cases	csv; shapefile	1/22/2020	3/13/2020	周怡；刘洪强
Baidu Migration Data		Transportation	csv;json	1/1/2020	3/15/2020	唐奕城
Flight		Transportation	csv	2/7/2020	3/13/2020	王老师；邵老师
Hospital POI	Wuhan, Hubei, China Hospital POI	Public Health	csv			邵老师；王佳磊；周桓
China Events Timeline		Policies and Regulations				杜晶
Global Events Timeline		Policies and Regulations	.xlsx	12/1/2019	3/15/2020	周怡；杜晶
Community Cases from Tecent		COVID-19 Cases			3/15/2020	刘洪强
Trajactory Data from Tecent		COVID-19 Cases		1/20/2020	3/16/2020	刘洪强
GDELT		Social Media		1/1/2020	3/15/2020	胡涛
Twitter(eng)		Social Media		1/1/2020	3/10/2020	胡涛
temperature		Climate				张老师
air quality		Climate				张老师

Outline

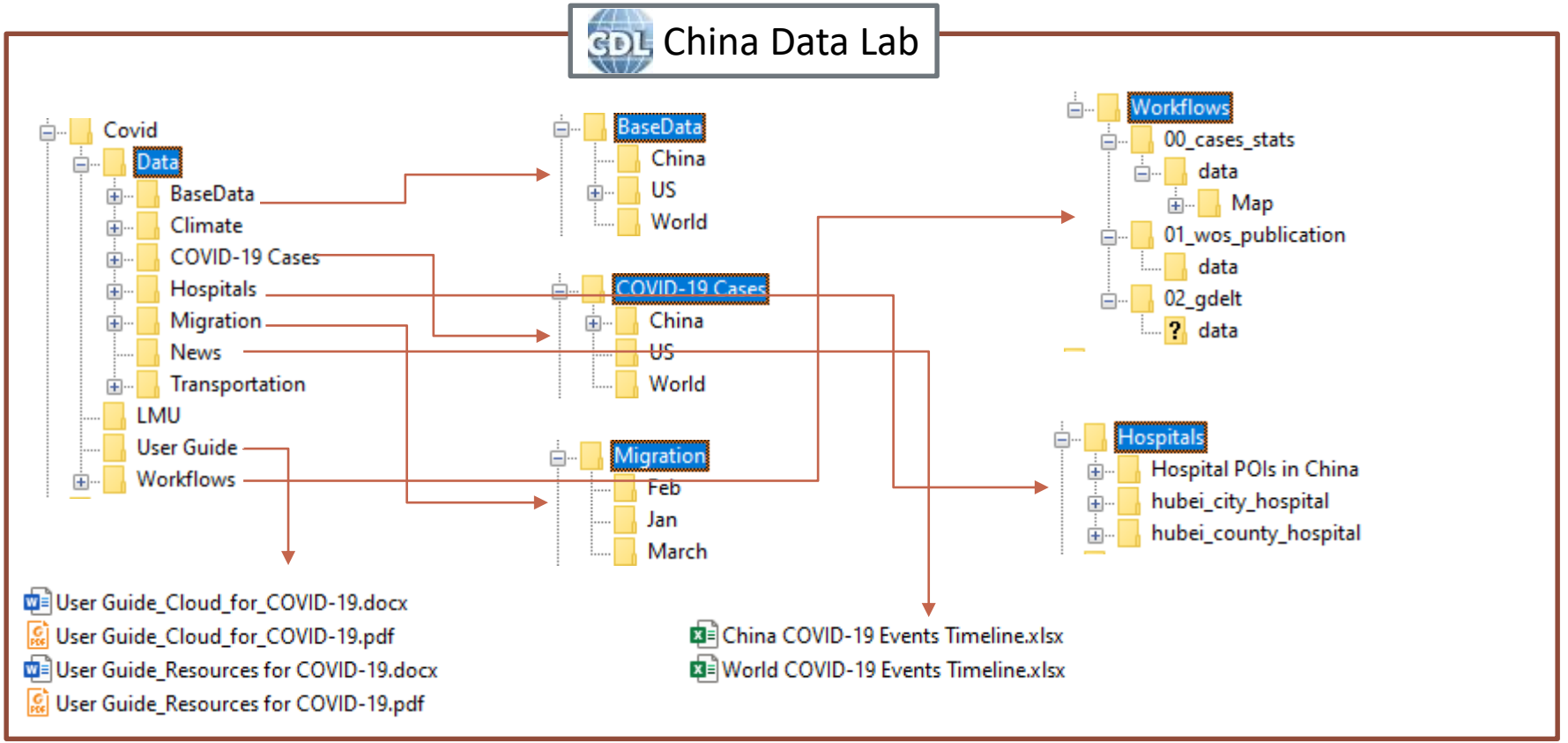
① COVID-19 Open Datasets

② Framework of “Resources for COVID-19”

③ “Resources for COVID-19” at Harvard Dataverse

④ “Resource for COVID-19” at CDL Cloud Platform

“Resources for COVID-19” at CDL Platform



“Resources for COVID-19” at CDL Platform

❑ Data Mapping between Harvard Dataverse and CDL platform

	Data Category		Data Files on http://chinadatalab.org	Datasets on DataVerse
2	Base Data	Base Maps	/Covid/Data/BaseData	https://dataverse.harvard.edu/dataverse/chinadatainstit
3		China	/Covid/Data/COVID-19 Cases/China	https://doi.org/10.7910/DVN/MR5IJN
4		US	/Covid/Data/COVID-19 Cases/US	https://doi.org/10.7910/DVN/HIDLTK
5	Virus Data	World	/Covid/Data/COVID-19 Cases/World	https://doi.org/10.7910/DVN/L20LOT
6		Patient Trajectories in China	/Covid/Data/COVID-19 Cases/China/communities	
7		Community Cases in China	/Covid/Data/COVID-19 Cases/China/trajectories	
8	Mobility Data	Baidu Mobility Data	/Covid/Data/Migration	https://doi.org/10.7910/DVN/FAEZIO
9	Transportation	Flight	/Covid/data/Transportation/flight	
10	Hospital		/Covid/Data/Hospitals	https://doi.org/10.7910/DVN/KRSGT3
11	Chronicle		/Covid/Data/News	https://doi.org/10.7910/DVN/OAM2JK
12	Climate	weather	/Covid/Data/Climate/weather	

Index | Base Data | Covid-19 Cases | Climate | Hospitals | Migration | Transportation | Chronicle | Workflows

❑ Data List at CDL platform

	Country	Data Title	File Name	File Type	File Path
2	China	City map with virus cases	City Map 0115-0308.csv	CSV file	public\Covid\Data\COVID-19 Cases\China
3	China	Province map with virus cases	Prov Map 0115-0308.csv	CSV file	public\Covid\Data\COVID-19 Cases\China
4	China	Individual virus cases	Community Cases.xlsx	Excel File	public\Covid\Data\COVID-19 Cases\China\Communities
5	China	Tracks of individual patients	Patient Trajectoies.xlsx	Excel File	public\Covid\Data\COVID-19 Cases\China\trajectories
6	US	Confirmed cases by US county	US County Confirm Cases 0311.csv	CSV file	public\Covid\Data\COVID-19 Cases\US
7	US	Recovered cases by US county	US County Recover Cases 0311.csv	CSV file	public\Covid\Data\COVID-19 Cases\US
8	US	Death cases by US county	US County Death Cases 0311.csv	CSV file	public\Covid\Data\COVID-19 Cases\US
9	US	Confirmed cases by US states	US State Confirm Cases 0311.csv	CSV file	public\Covid\Data\COVID-19 Cases\US
10	US	Recovered cases by US statss	US State Recover Cases 0311.csv	CSV file	public\Covid\Data\COVID-19 Cases\US
11	US	Death cases by US states	US State Death Cases 0311.csv	CSV file	public\Covid\Data\COVID-19 Cases\US
12	World	Virus by Countries0302.xlsx	Virus Cases by Countries	Excel File	public\Covid\Data\COVID-19 Cases\World

Index | Base Data | Covid-19 Cases | Climate | Hospitals | Migration | Transportation | Chronicle | Workflows

Base Data

Data Title	File Name	File Type	File Path
Prefecture County Census Data with 2000-2010 GIS Maps	2000-2010CountyCensus.shp	Shape File	public\Covid\Data\BaseData\China
Prefecture City Census Data with 2000-2010 GIS Maps	2000-2010PrefectureCensus.shp	Shape File	public\Covid\Data\BaseData\China
Prefecture Province Census Data with 2000-2010 GIS Maps	2000-2010ProvCensus.shp	Shape File	public\Covid\Data\BaseData\China
Base map, County	County0010.shp	Shape File	public\Covid\Data\BaseData\China
Base map, Prefecture City	City0010.shp	Shape File	public\Covid\Data\BaseData\China
Base map, Province	Prov0010.shp	Shape File	public\Covid\Data\BaseData\China
Base map, Country Boundary	chinaln.shp	Shape File	public\Covid\Data\BaseData\China
Index File for Province Census Data with 2000-2010 GIS Maps	List Maps three categories Prov 00-10.xlsx	Excel File	public\Covid\Data\BaseData\China
Index File for Prefecture Census Data with 2000-2010 GIS Maps	List Maps three categories City 00-10.xlsx	Excel File	public\Covid\Data\BaseData\China
Index File for County Census Data with 2000-2010 GIS Maps	List Maps three categories County 00-10.xlsx	Excel File	public\Covid\Data\BaseData\China
Base map, US States	ST CARTO.shp	Shape File	public\Covid\Data\COVID-19 Cases\US
Base map, US Counties	CO CARTO.shp	Shape File	public\Covid\Data\COVID-19 Cases\US
Base map, Countries	Countries.shp	Shape File	public\Covid\Data\COVID-19 Cases\World

Base maps

- China: 2000-2010 comparable census maps (province, city and county/urban district)
- USA: 1970-2010 comparable census maps (state, county)

Virus Data

Country	Data Title	File Name	File Type	File Path
China	City map with virus cases	City Map 0115-0308.csv	CSV file	public\Covid\Data\COVID-19 Cases\China
China	Province map with virus cases	Prov Map 0115-0308.csv	CSV file	public\Covid\Data\COVID-19 Cases\China
China	Individual virus cases	腾讯看点-疫情小区0316V1.1.xlsx	Excel File	public\Covid\Data\COVID-19 Cases\China\Communities
China	Tracks of individual patients	轨迹数据0316V1.0.xlsx	Excel File	public\Covid\Data\COVID-19 Cases\China\trajectories
US	Confirmed cases by US county	US County Confirm Cases 0311.csv	CSV file	public\Covid\Data\COVID-19 Cases\US
US	Recovered cases by US county	US County Recover Cases 0311.csv	CSV file	public\Covid\Data\COVID-19 Cases\US
US	Death cases by US county	US County Death Cases 0311.csv	CSV file	public\Covid\Data\COVID-19 Cases\US
US	Confirmed cases by US states	US State Confirm Cases 0311.csv	CSV file	public\Covid\Data\COVID-19 Cases\US
US	Recovered cases by US statss	US State Recover Cases 0311.csv	CSV file	public\Covid\Data\COVID-19 Cases\US
US	Death cases by US states	US State Death Cases 0311.csv	CSV file	public\Covid\Data\COVID-19 Cases\US
World	Virus by Countries0302.xlsx	Virus Cases by Countries	Excel File	public\Covid\Data\COVID-19 Cases\World

Note: The virus data aggregated by city in China has been adjusted to the 2000-2010 comparable city map

Hospital Data

Country/Region	Data Title	File Name	File Type	File Path
China	Hospitals by province	XXX.csv	csv	public/Covid/Data/Hospitals/Hospital POIs in China/Hospitals by Province/
	Hospital POI metadata	Metadata.xlsx	xlsx	public/Covid/Data/Hospitals/Hospital POIs in China
Hubei, China	County hospitals	hubei_city_hospital.shp	shapefile	public/Covid/Data/Hospitals/hubei_city_hospital/hubei_city_hospital
Hubei, China	City hospitals	hubei_county_hospital.shp	shapefile	public/Covid/Data/Hospitals/hubei_city_hospital/hubei_county_hospital

Note:

1. The hospital data aggregated by city in China has been adjusted to the 2000-2010 comparable city map
2. Original hospital data were from AutoNavi and filtered by excluding duplicated records

Mobile Population Data

Country	Data Title	File Name	File Type	File Path
China	Baidu Mobility Data	baidu_in	csv	public/Covid/Data/Migration/
China	Baidu Mobility Data	baidu_out	csv	public/Covid/Data/Migration/

Note:

1. The original mobile population data were from Baidu and has been adjusted to the 2000-2010 comparable city list

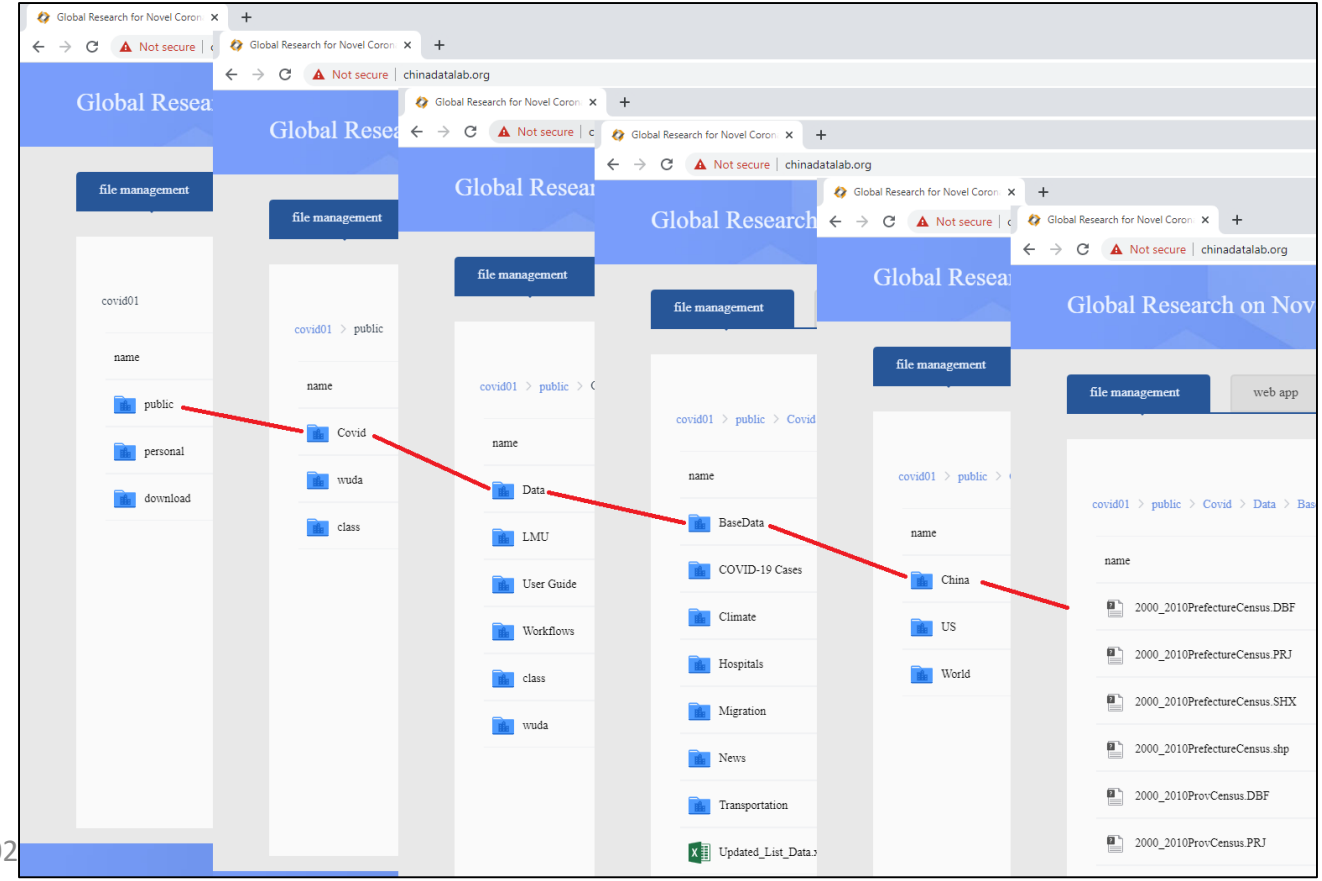
Flight Data

Country	Data Title	File Name	File Type	File Path
Global	Global Flight Data	airport-frequencies	csv	public/Covid/Data/Transportaion/flight/
		airports	csv	public/Covid/Data/Transportaion/flight/
		countries	csv	public/Covid/Data/Transportaion/flight/
		navaids	csv	public/Covid/Data/Transportaion/flight/
		regions	csv	public/Covid/Data/Transportaion/flight/
		runway	csv	public/Covid/Data/Transportaion/flight/

Workflows

Title	workflow_name	workflow_path
1. COVID19_cases_statistics	covid19_stats_confirmed.knwf	/Covid/Workflows/00_cases_stats
	covid19_stats_death.knwf	/Covid/Workflows/00_cases_stats
	covid19_stats_recovered.knwf	/Covid/Workflows/00_cases_stats
2. COVID19 Publication Analysis	covid19_wos_pub.knwf	/Covid/Workflows/01_wos_publication

The Hierarchical Data Structure



Acknowledgement

Wuhan University
Xinyan Zhu, Yuanzhen Shao, Lingbo Liu, Yandong Wang, Kun Qin, Jing Du, Yanlin Li, Zhiwei Fu, Yicheng Tang, Hongqiang Liu, Huan Zhou, An Li, Lian Duan, Jialei Wang, Xuesong Yu

Peking University
Jie Huang, Yi Zhou

Harvard University
Wendy Guan, Tao Hu, Wei Luo, Zhibin Li, Danny Brooke

George Mason University
Phil Yang, Dexuan Sha

East China Normal University
Luyao Zhang

Vesystem Inc.
Chong Gao, Ran Wei

Websites and Contact

Resources for Coronavirus Study

http://chinadatalab.net

The Cloud for COVID-19 Study

Http://chinadatalab.org

covid-2019@umich.edu