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

“Resources for COVID-19”

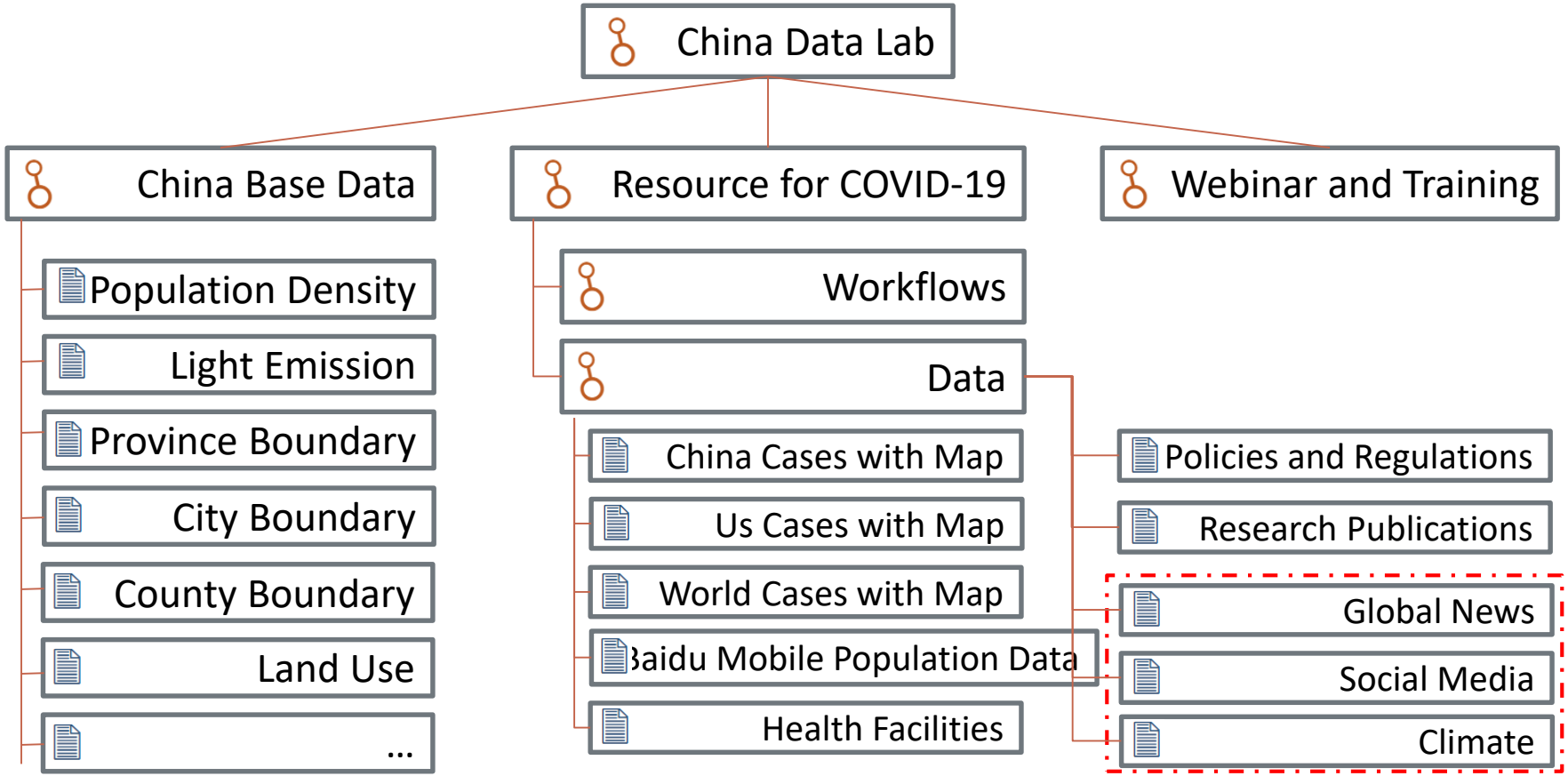
Introduction to Hospital POI
Data collection, processing and analysis

Tao Hu, Harvard University
Huan Zhou and Jialei Wang, Wuhan University

Outline

- ❑ Procedure of Data Crawling
- ❑ Data Processing
- ❑ Data Analysis
- ❑ Data Depository
- ❑ Data Cloud
- ❑ Discussion

“Resources for COVID-19” at Harvard Dataverse



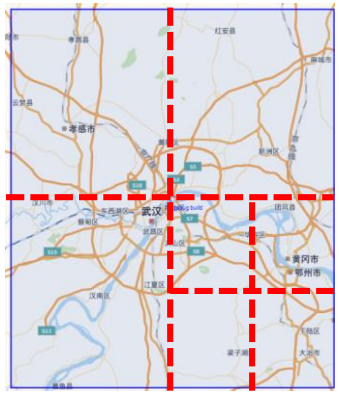
Procedure of Data Crawling

Obtain HTTP interface of Gaode POI data, and determine the type and location of POI

Extract the rectangle of administrative divisions and retrieve the number of pois

Q：When retrieving a POI of a certain category, if the number of POIs of this category exceeds 890, the excess part will be lost and cannot be obtained

Rectangle segmentation beyond data limit based on quarter method



Extract POI data for each rectangle

Data processing

Data Processing

Source: https://restapi.amap.com/v3/place/text?parameters

type_CH
生活服务;生活服务场所;生活服务场所
商务住宅;住宅区;住宅区
生活服务;美容美发店;美容美发店
生活服务;美容美发店;美容美发店
生活服务;美容美发店;美容美发店

大类	中类	Sub Category	小类	Mid Category	Big Category
生活服务	生活服务场所	Daily Life Service Place	生活服务场所	Daily Life Service Place	Daily Life Service
商务住宅	住宅区	Residential Area	住宅区	Residential Area	Commercial House
生活服务	美容美发店	Beauty and Hairdressing Store	美容美发店	Beauty and Hairdressing Store	Daily Life Service
生活服务	美容美发店	Beauty and Hairdressing Store	美容美发店	Beauty and Hairdressing Store	Daily Life Service
生活服务	美容美发店	Beauty and Hairdressing Store	美容美发店	Beauty and Hairdressing Store	Daily Life Service

type_CH
购物服务;专卖店;儿童用品店|购物服务;服装鞋帽皮具店;品牌鞋店|购物服务;体育用品店;户外用品

type_CH_1	type_CH_2	type_CH_3
儿童用品店	品牌鞋店	户外用品

Data Processing

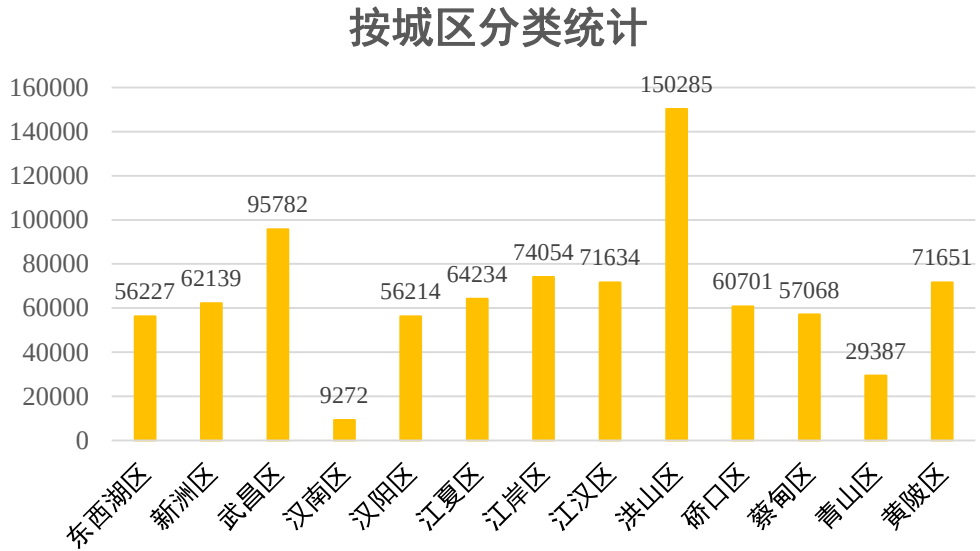
name	POI名称
type_CH	POI中文类别
address	语义地址
province	所属省份
city	所属城市
citycode	所属城市编码
district	所属区级
street	所属街道
adcode	所属区编码
typecode	所属类别编码
WGS_84_X	WGS-84坐标
WGS_84_Y	WGS-84坐标
type_CH_1	类别1
type_CH_2	类别2
type_CH_3	类别3
大类	大类(高德定义)
中类	中类(高德定义)
小类	小类(高德定义)
Big Category	大类EN(高德定义)
Mid Category	中类EN(高德定义)
Sub Category	小类EN(高德定义)

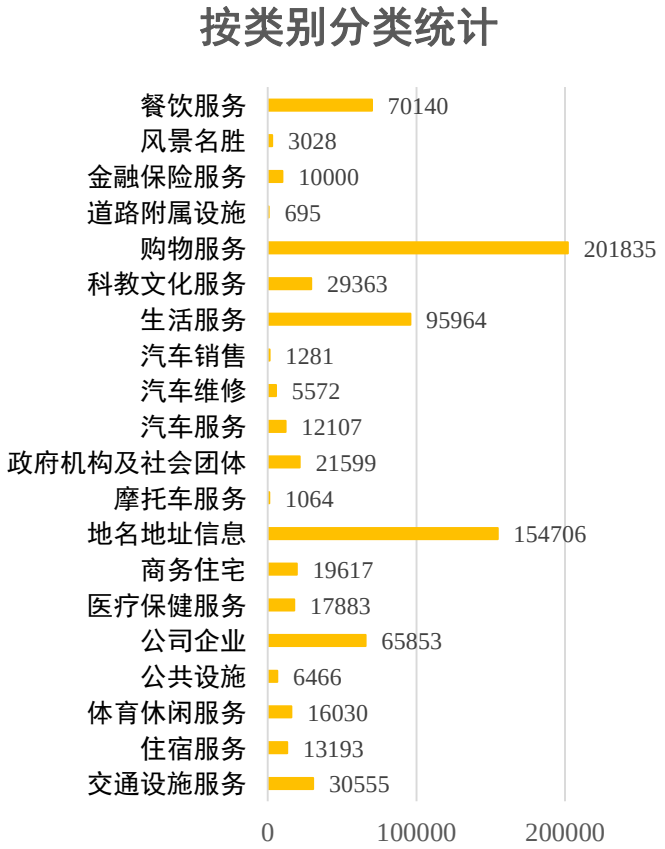
city	type_CH	name	street	citycode	province	typecode	address	district	adcode
武汉市	生活服务;生活服务场所;生活服务场所	尚亿佳房产	和平大道	27	湖北省	070000	湖北省武汉市武昌区杨园街街道武昌区杨园街国棉社区	武昌区	420106
武汉市	商务住宅;住宅区;住宅区	欧景苑-蓝俊庭	才汇巷	27	湖北省	120300	湖北省武汉市武昌区杨园街街道欧景苑	武昌区	420106
武汉市	生活服务;美容美发店;美容美发店	美丽造型(和平大道)	和平大道	27	湖北省	071100	湖北省武汉市武昌区杨园街街道武昌区杨园街国棉社区	武昌区	420106
武汉市	生活服务;美容美发店;美容美发店	美乐理发	和平大道	27	湖北省	071100	湖北省武汉市武昌区杨园街街道武昌区杨园街国棉社区	武昌区	420106
武汉市	生活服务;美容美发店;美容美发店	名媛国际造型	铁机路	27	湖北省	071100	湖北省武汉市武昌区杨园街街道武汉新江南医院住院部橡树湾	武昌区	420106

大类	中类	Sub Category	type_CH_3	小类	type_CH_2	WGS_84_Y	Mid Category	Big Category	type_CH_1	WGS_84_X
生活服务	生活服务场所	Daily Life Service Place	无	生活服务场所	无	30.611261	Daily Life Service Place	Daily Life Service	生活服务场所	114.347568
商务住宅	住宅区	Residential Area	无	住宅区	无	30.618145	Residential Area	Commercial House	住宅区	114.349590
生活服务	美容美发店	Beauty and Hairdressing Store	无	美容美发店	无	30.609689	Beauty and Hairdressing Store	Daily Life Service	美容美发店	114.344769
生活服务	美容美发店	Beauty and Hairdressing Store	无	美容美发店	无	30.610117	Beauty and Hairdressing Store	Daily Life Service	美容美发店	114.345615
生活服务	美容美发店	Beauty and Hairdressing Store	无	美容美发店	无	30.611618	Beauty and Hairdressing Store	Daily Life Service	美容美发店	114.344295

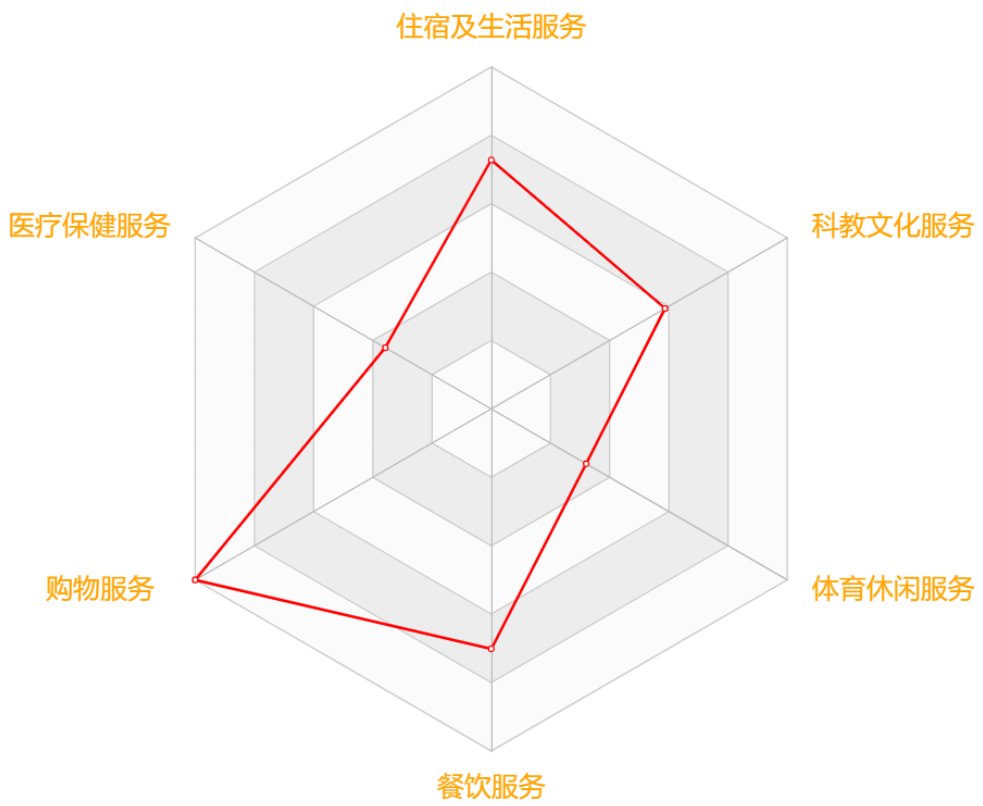
Data Analysis

21 大类　　204 中类　　600 小类

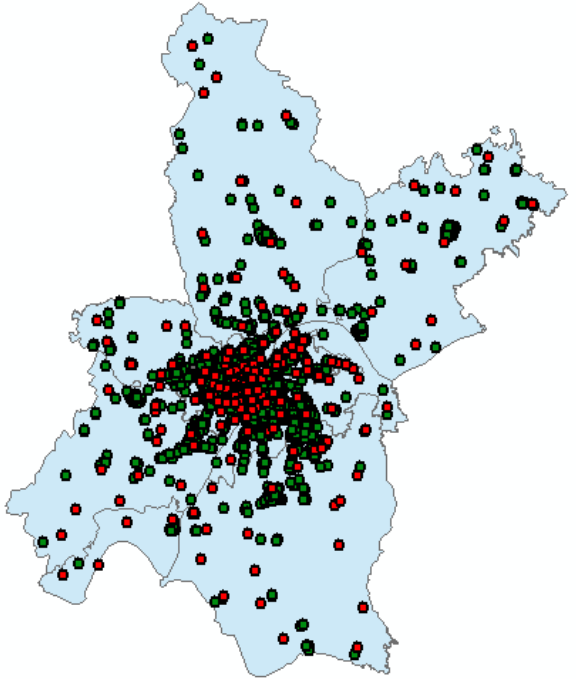




Data Analysis



武汉高风险社区和餐饮服务POI



Health Facilities Data Updated

Q1: Duplicate hospital POI

115.617036	34.396407	商丘市第一人民医院综合病房楼	凯旋南路292号	河南省	商丘市	NaN	三级甲等医院	NaN	B0172004TQ	B0FFFQFDJA
115.618448	34.397418	商丘市第一人民医院门诊部	凯旋南路292号	河南省	商丘市	NaN	三级甲等医院	NaN	B0172004TQ	B01720OWQ4
115.618007	34.396327	第一人民医院住院楼	凯旋南路292号	河南省	商丘市	NaN	三级甲等医院	NaN	B0172004TQ	B01720OVTK

Q2: Questionable clinical data

诊所	23099

诊所	16549

Hebei Province

Q： About the amount of data?

search results based on Administrative Region Coding and quartering polygon

70990 73556

1. The number of clinics

2. Pharmacy data labels are not sorted

type	
三级甲等医院	395
专科医院	893
传染病医院	90
医疗保健服务场所	12073
卫生院	3755
口腔医院	4725
妇科医院	283
急救中心	185
整形美容	333
疾病预防	518
眼科医院	628
精神病医院	63
综合医院	3732
耳鼻喉医院	141
肿瘤医院	65
胸科医院	39
脑科医院	73
药房	21912
诊所	23099
骨科医院	554

AMAP_TYPES	
三级甲等医院	1241
专科医院	72
传染病医院	153
医疗保健服务场所	17490
卫生院	13146
口腔医院	8726
妇科医院	504
急救中心	616
整形美容	655
疾病预防	1045
眼科医院	1187
精神病医院	105
综合医院	7858
耳鼻喉医院	262
肿瘤医院	135
胸科医院	67
脑科医院	109
诊所	16549
骨科医院	1070

Data Depository: dataverse.harvard.edu

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 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?

Journals

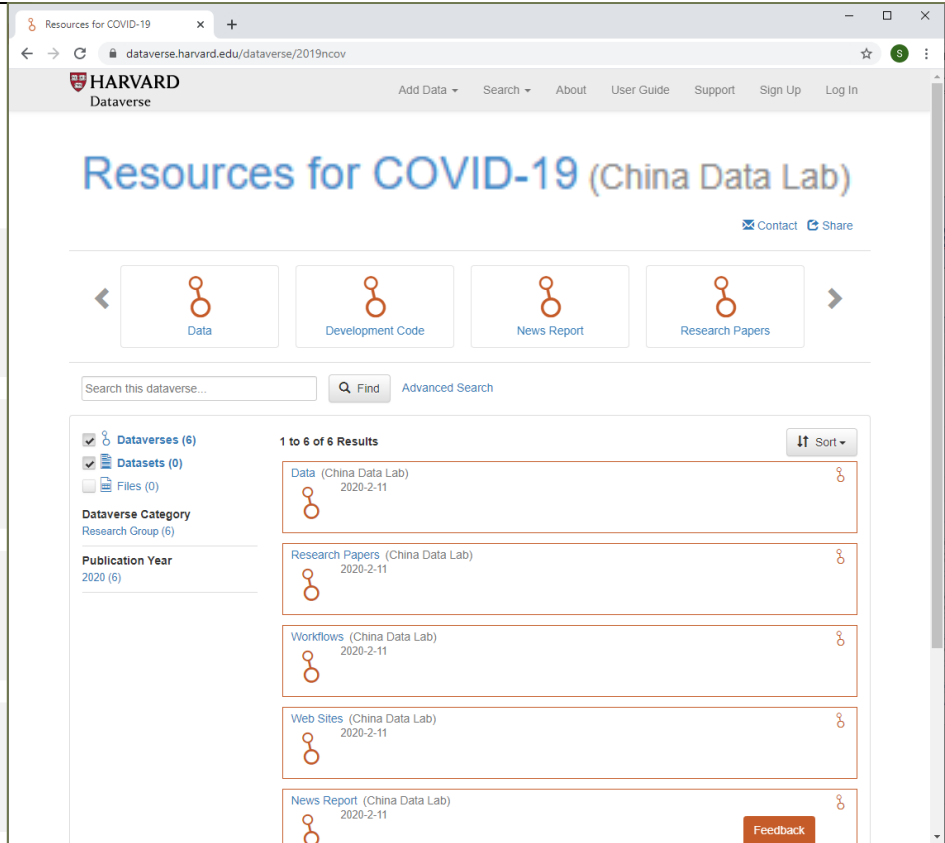
Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data.* Participate in the open data movement by using Dataverse as part of your journal data policy or list of repository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

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?

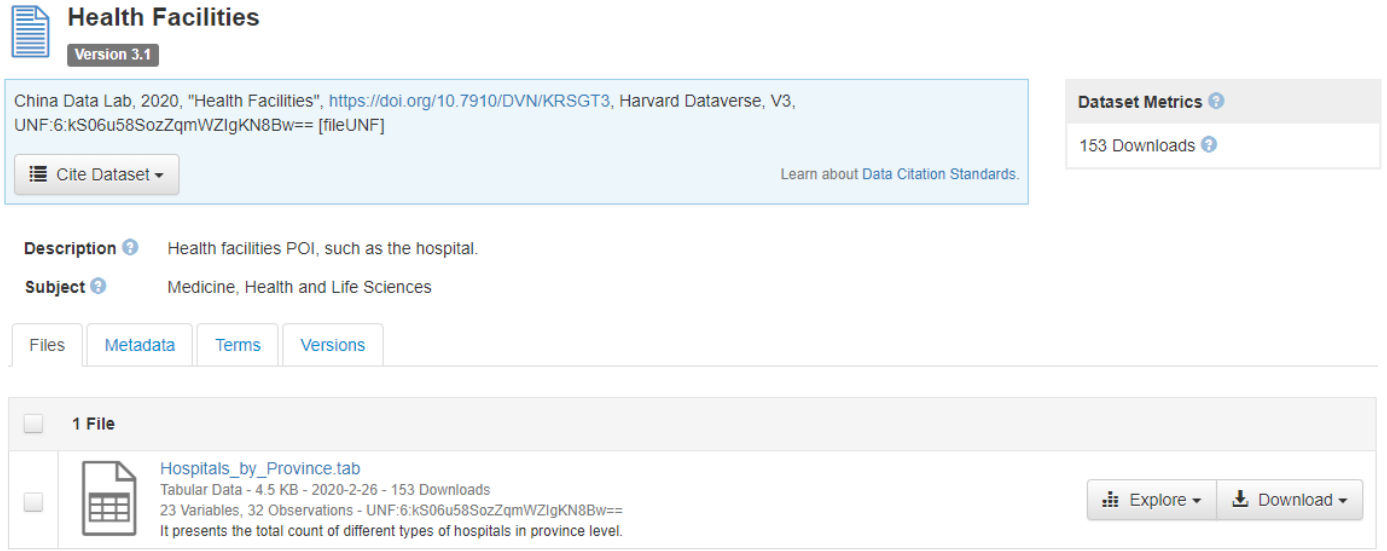
Developers

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 Dataverse. Want to contribute?



Health Facilities Data at Harvard Dataverse

Dataset for downloading: https://doi.org/10.7910/DVN/KRSGT3



Mobile Population Data on the Cloud

Website: http://chinadatalab.org
File Path: Covid\Data\Hospitals\

Websites and Contact

Resources for Coronavirus Study

http://chinadatalab.net

The Cloud for COVID-19 Study

Http://chinadatalab.org

Webinars on Virus Study

https://dataverse.harvard.edu/dataverse/cdl_training

covid-2019@umich.edu