

COVID-19

Data extraction process

Data collection team

We need your help for....



Province folder

- **html files**

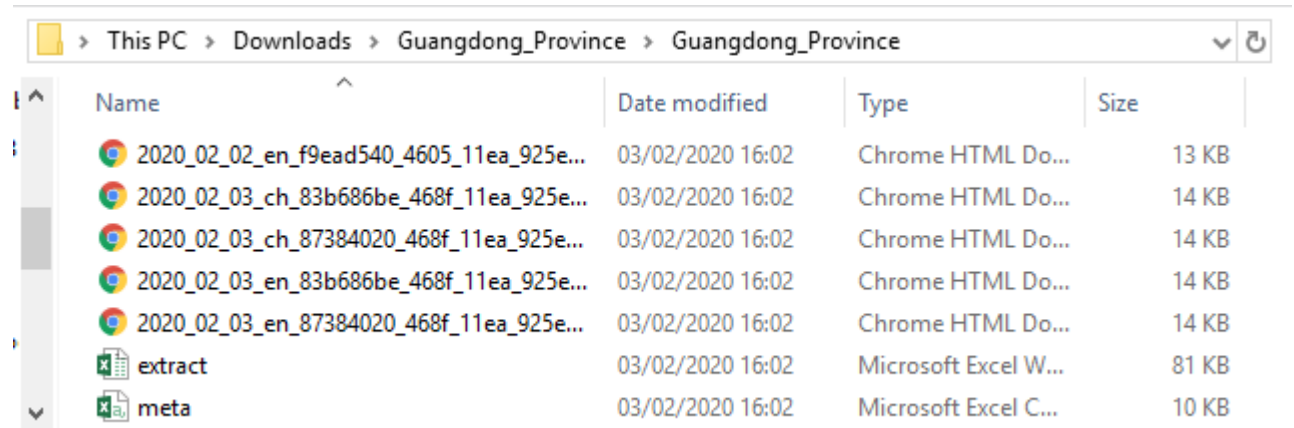
- en - English reports
- ch - Chinese reports
- Files with the same unique ids have the same contents.

- **extract.xlsx**

- Extraction table

- **meta.csv**

- Summary of scraped reports
- URLs of scraped reports



Name	Date modified	Type	Size
2020_02_02_en_f9ead540_4605_11ea_925e...	03/02/2020 16:02	Chrome HTML Do...	13 KB
2020_02_03_ch_83b686be_468f_11ea_925e...	03/02/2020 16:02	Chrome HTML Do...	14 KB
2020_02_03_ch_87384020_468f_11ea_925e...	03/02/2020 16:02	Chrome HTML Do...	14 KB
2020_02_03_en_83b686be_468f_11ea_925e...	03/02/2020 16:02	Chrome HTML Do...	14 KB
2020_02_03_en_87384020_468f_11ea_925e...	03/02/2020 16:02	Chrome HTML Do...	14 KB
extract	03/02/2020 16:02	Microsoft Excel W...	81 KB
meta	03/02/2020 16:02	Microsoft Excel C...	10 KB

② Filter and archive

- **Keep reports with**
 - ✓ Case numbers
 - ✓ Death numbers
 - ✓ Contact numbers
 - ✓ Individual case/death reports
- **Delete reports on**
 - ✓ List of hospitals
 - ✓ Public health notice of COVID-19

✓ **KEEP & ARCHIVE**

Pneumonia situation of new coronavirus infection in Guangdong Province at 12:00 on January 30, 2020

Time: 2020-01-30 15:47:42 source: this website

As of 12:00 on January 30, the province has reported a total of 354 confirmed cases of pneumonia due to new coronavirus infection. At 02:00 on the 30th, 43 new cases were confirmed and 2 were discharged from the province. Among the newly confirmed cases, 15 were in Guangzhou, 12 in Shenzhen, 8 in Zhuhai, 4 in Dongguan, 2 in Zhongshan, and 2 in Zhanjiang. 1,514 close contacts are under medical observation.

Of the confirmed cases, 94 were in Guangzhou, 98 in Shenzhen, 26 in Zhuhai, 12 in Shantou, 25 in Foshan, 4 in Shaoguan, 1 in Heyuan, 5 in Meizhou, 17 in Huizhou, and Shanwei. There were 1 case in Shenzhen, 11 cases in Dongguan, 18 cases in Zhongshan City, 1 case in Jiangmen City, 10 cases in Yangjiang City, 11 cases in Zhanjiang City, 3 cases in Maoming City, 5 cases in Zhaoqing City, 6 cases in Qingyuan City, and 6 cases in Jieyang City. There were 176 males and 178 females, aged between 11 months and 85 years.

At present, there are 37 cases of severe illness, 13 cases of critical illness, no deaths, and a total of 8 patients have been discharged.

② Filter and archive

✓ **KEEP & ARCHIVE**
(individual case reports)

As of 22:00 on January 30, one new case of pneumonia confirmed by a new coronavirus infection has been added to the number of confirmed cases in Tianjin

Date: 2020-01-30 23:02

source: Commission office

As of 22:00 on January 30, Tianjin has newly added one confirmed case of pneumonia infected with a new type of coronavirus, and the number of confirmed cases has increased to 31.

The 31st patient, female, 29 years old, lived in the Sino-Singapore Ecological City in Tianjin Binhai New District. The patient and his family of four drove to his hometown in Suizhou City, Hubei Province on January 21. He returned to Tianjin by self-driving on January 24. He suffered from pharyngeal discomfort on the 26th and fever on the 29th. On January 30, the 31st case of Tianjin was confirmed by the expert group of the city, which was a common case. He has been transferred to Haihe Hospital for treatment.

An epidemiological investigation revealed that the patient's main close contacts were his 3 family members. At present, 3 people have been concentrated at the observation point for isolated medical observation, and the CDC of Binhai New Area has terminally disinfected their dwellings.

Up to now, 31 cases of pneumonia of new coronavirus infection have been found in Tianjin, including 23 males and 8 females; 13 severe cases and 18 common cases; and no critical cases and deaths.

Tianjin Municipal Health Committee

January 30, 2020

② Filter and archive

✕ DELETE

List of third-party testing institutions for new coronavirus nucleic acid testing in Shaanxi

source: Medical administration

Release time: 2020-02-11 15:34

Font: 【 [Tai Zhong small](#) 】

In order to effectively do a good job of preventing and controlling the epidemic situation and accelerating the nucleic acid testing of New Coronary Pneumonia, the provincial health committee organized experts to evaluate the qualifications of third-party testing institutions for conducting nucleic acid testing. The list of qualified third-party testing institutions is now included (the first batch) It will be announced for reference by local governments in epidemic prevention and control.

1. Name of institution: Xi'an Jinyu Medical Laboratory

Address: 8989 Shangjing Road, Economic Development Zone, Xi'an

Phone: 029-82683888

2. Institution name: Xi'an Tianbo Medical Laboratory

Address: Building 2, No. 1155, Caotan Ten Road, Xi'an

Phone: 029-86554262

3. Institution name: Xi'an Friendship Medical Laboratory

Address: 4th Floor, Block B, Xijing Company, Electronic Third Road, Xi'an High-tech Zone

Phone: 029-88412556

② Filter and archive

Note: both google translates and original documents are needed

{}	2020_02_01_ch_cc443e3c_4511_11ea_925e_5f121fde4e57.html	✓
{}	2020_02_01_ch_cf4923e0_4511_11ea_925e_5f121fde4e57.html	✓
{}	2020_02_01_en_cc443e3c_4511_11ea_925e_5f121fde4e57.html	✓
{}	2020_02_01_en_cf4923e0_4511_11ea_925e_5f121fde4e57.html	✓
{}	2020_02_02_ch_f6c5a7dc_4605_11ea_925e_5f121fde4e57.html	✗
{}	2020_02_02_ch_f9ead540_4605_11ea_925e_5f121fde4e57.html	✗
{}	2020_02_02_en_f6c5a7dc_4605_11ea_925e_5f121fde4e57.html	✗
{}	2020_02_02_en_f9ead540_4605_11ea_925e_5f121fde4e57.html	✗



{}	2020-02-01-GUANGDONG-HC-1-CH.html
{}	2020-02-01-GUANGDONG-HC-1-EN.html
{}	2020-02-01-GUANGDONG-HC-2-CH.html
{}	2020-02-01-GUANGDONG-HC-2-EN.html

② Filter and archive

{}	2020_02_01_ch_cc443e3c_4511_11ea_925e_5f121fde4e57.html	✓
{}	2020_02_01_ch_cf4923e0_4511_11ea_925e_5f121fde4e57.html	✓
{}	2020_02_01_en_cc443e3c_4511_11ea_925e_5f121fde4e57.html	✓
{}	2020_02_01_en_cf4923e0_4511_11ea_925e_5f121fde4e57.html	✓
{}	2020_02_02_ch_f6c5a7dc_4605_11ea_925e_5f121fde4e57.html	
{}	2020_02_02_ch_f9ead540_4605_11ea_925e_5f121fde4e57.html	
{}	2020_02_02_en_f6c5a7dc_4605_11ea_925e_5f121fde4e57.html	
{}	2020_02_02_en_f9ead540_4605_11ea_925e_5f121fde4e57.html	



{}	<u>2020-02-01-GUANGDONG-HC-1-CH.html</u>
{}	2020-02-01-GUANGDONG-HC-1-EN.html
{}	2020-02-01-GUANGDONG-HC-2-CH.html
{}	2020-02-01-GUANGDONG-HC-2-EN.html

Format of files:

YYYY-MM-DD-Province-Source-No.-Language.html

Date of release

*Serial number in
the same date*

③ Extract data

- See *Dictionary* sheet
- Information for collection
 - Province
 - Dates and hours
 - Cumulative cases
 - New cases
 - Current/prevalent cases
 - Data source

DICTIONARY - PROVINCES WITHOUT CLINICALLY CONFIRMED CASES					
	Column name	Definition and format			
A	Record number	Order of entry			
B	Province	Province/Autonomous region/City of the health authorities that publish the report			
C	Date release	Date of the report/press release. Format: YYYY-MM-DD			
D	Date reporting start (notification)	Start date of reporting new confirmed cases. Format: YYYY-MM-DD			
E	Hour reporting start (notification)	Start hour of reporting new confirmed cases. Format: 24 hour. Leave blank if only date is available			
F	Date reporting end (notification)	End date of reporting new confirmed cases. Format: YYYY-MM-DD			
G	Hour reporting end (notification)	End hour of reporting new confirmed cases. Format: 24 hour. Leave blank if only date is available			
H	Cumulative confirmed	Number of total confirmed cases till the end time of reporting (notification)			
I	Cumulative critical	Number of total confirmed critical cases till the end time of reporting (notification), 危重症			
J	Cumulative severe	Number of total confirmed severe cases till the end time of reporting (notification), 重症			
K	Cumulative common	Number of total confirmed common cases till the end time of reporting (notification), 普通症			
L	Cumulative mild	Number of total confirmed mild cases till the end time of reporting (notification)			
M	Cumulative suspect	Number of total suspected cases till the end time of reporting (notification)			
N	Cumulative cured	Number of total cured cases till the end time of reporting (notification)			
O	Cumulative deaths	Number of total deaths till the end time of reporting (notification)			
P	Cumulative close contacts	Number of total close contacts till the end time of reporting (notification)			
Q	Cumulative close contacts removed (cleared)	Number of total close contacts removed from observation till the end time of reporting (notification)			
R	New confirmed	Number of newly confirmed cases over the reporting time (notification)			
S	New critical	Number of newly confirmed critical cases over the reporting time (notification)			
T	New severe	Number of newly confirmed severe cases over the reporting time (notification)			
U	New common	Number of newly confirmed common cases over the reporting time (notification)			
V	New mild	Number of newly confirmed mild cases over the reporting time (notification)			
W	New suspect	Number of newly confirmed suspected cases over the reporting time (notification)			
X	New cured	Number of newly cured cases over the reporting time (notification)			
Y	New deaths	Number of newly deaths over the reporting time (notification)			
Z	New close contacts	Number of close contacts newly included over the reporting time (notification)			
AA	New close contacts removed (cleared)	Number of close contacts newly removed from observation over the reporting time (notification)			
AB	Current suspect	Number of suspect cases currently observed			
AC	Under treatment	Number of total cases currently under treatment			
AD	Under treatment critical	Number of critical cases currently under treatment			
AE	Under treatment severe	Number of severe cases currently under treatment			
AF	Under treatment common	Number of common cases currently under treatment			
AG	Under treatment mild	Number of mild cases currently under treatment			
AH	Under contact observation	Number of contacts currently under medical observation			
AI	Other	Other information/notes			
AJ	URL1	URL of the source report. Not required for files saved in Dropbox.			
AK	Dropbox file 1	Filename saved in the dropbox for URL1			
AL	URL2	URL of additional source report if available. Not required for files saved in Dropbox.			

③ Extract data

This should be: From 0:00 to 24:00 of January 29, 2020

*Note: google translate often has problem with 'time' & 'hour'.
Please note down when you see potential errors.*

Epidemic situation of new coronavirus infection in Tianjin on January 30, 2020

Date: 2020-01-30 09:38

source: Commission office

At 04:00 on January 29, 2020, Tianjin reported 3 new confirmed cases of pneumonitis with new coronavirus infection, 0 new cases of severe cases, 0 new deaths, and 0 new discharged cases. among them:

Among the newly confirmed cases in each district, there was 1 case in Heping District and 2 cases in Tianjin.

As of 24:00 on January 29, Tianjin has reported 27 cases of pneumonia confirmed by new coronavirus infection, 13 cases of severe cases, 0 deaths, and 0 discharged patients. among them:

Among the confirmed cases, there were 7 cases in Hebei District, 5 cases in Hedong District, 3 cases in Hexi District, 2 cases in Nankai District, 2 cases in Xiqing District, 2 cases in Heping District, 1 case in Hongqiao District, 1 case in Binhai New District, and 4 places in Laijin. example.

At present, 210 close contacts have been traced, 2 people have been released from medical observation, and 201 people are still receiving medical observation.

Tianjin Municipal Health Committee

January 30, 2020

③ Extract data

Pneumonia situation of new coronavirus infection in Guangdong Province at 12:00 on January 30, 2020

Date of release

Time: 2020-01-30 15:47:42 source: this website

Date and hour of reporting

Cumulative cases

New cases

New cured

As of 12:00 on January 30, the province has reported a total of 354 confirmed cases of pneumonia due to new coronavirus infection. At 02:00 on the 30th, 43 new cases were confirmed and 2 were discharged from the province. Among the newly confirmed cases, 15 were in Guangzhou, 12 in Shenzhen, 8 in Zhuhai, 4 in Dongguan, 2 in Zhongshan, and 2 in Zhanjiang. 1,514 close contacts are under medical observation.

Of the confirmed cases, 94 were in Guangzhou, 98 in Shenzhen, 26 in Zhuhai, 12 in Shantou, 25 in Foshan, 4 in Shaoguan, 1 in Heyuan, 5 in Meizhou, 17 in Huizhou, and Shanwei. There were 1 case in Shenzhen, 11 cases in Dongguan, 18 cases in Zhongshan City, 1 case in Jiangmen City, 10 cases in Yangjiang City, 11 cases in Zhanjiang City, 3 cases in Maoming City, 5 cases in Zhaoqing City, 6 cases in Qingyuan City, and 6 cases in Jieyang City. There were 176 males and 178 females, aged between 11 months and 85 years. **'other'**

At present, there are 37 cases of severe illness, 13 cases of critical illness, no deaths, and a total of 8 patients have been discharged.

Current critical cases, severe cases, deaths

Cumulative cured (discharged)

③ Extract data

- Extract information in 'extract.xlsx'

2019-01-30-Official reports China Mainland - Excel

Record number	Province	City	Date release	Date reporting	Hour reporting	Date reporting	Hour reporting	Cumulative con	Cumulative crit	Cumulative sev	Cumulative mo	Cumulative mil	Cumulative sus	Cumulative cur	Cumulative dea	Cumulative clo	Cumulative clo	New
1	Guangdong	-	2020-01-20			2020-01-19		1										
2	Guangdong	-	2020-01-21			2020-01-21	18	17	2	5			4		0			
3	Guangdong	-	2020-01-22			2020-01-21	24	26	3	7			1		0			
4	Guangdong	-	2020-01-23	2020-01-22	0	2020-01-22	24	32	3	12					0	212		
5	Guangdong	-	2020-01-24	2020-01-23	0	2020-01-23	24	53	3	12				2	0	258		
6	Guangdong	-	2020-01-25	2020-01-24		2020-01-24	24	78	4	15				2	0	346		
7	Guangdong	-	2020-01-26	2020-01-26	0	2020-01-26	16	111								562		
8	Guangdong	-	2020-01-26	2020-01-25		2020-01-25	24	98	5	18				2	0	505		
9	Guangdong	-	2020-01-27	2020-01-27	0	2020-01-27	12	151								804		
10	Guangdong	-	2020-01-27	2020-01-26		2020-01-26	24	146	6	20				2	0	708		
11	Guangdong	-	2020-01-28	2020-01-28	0	2020-01-28	12	207	6	20					0	1027		
12	Guangdong	-	2020-01-28	2020-01-27		2020-01-27	24	188	6	23				4	0	1019		
13	Guangdong	-	2020-01-29	2020-01-29	0	2020-01-29	12	272								1286		
14	Guangdong	-	2020-01-29	2020-01-28	0	2020-01-28	24	241	9	22				5	0	1342		

③ Extract data

- **URL 1**

- See 'link' in **meta.csv**



meta - Excel										
File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do										
L6										
	A	B	C	D	E	F	G	H	I	J
1	date	link	text_ch	text_en	uid					
2	20/01/2020	http://wsjkw.gd.gov.cn/zw<U+56FD><National	2020-01-20-fedf700c-43a7-11ea-9696-97d6e81c3666							
3	15/01/2020	http://wsjkw.gd.gov.cn/zw<U+5E7F><Guangdon	2020-01-15-ffb5b284-43a7-11ea-9696-97d6e81c3666							
4	15/01/2020	http://wsjkw.gd.gov.cn/zw<U+5E7F><Guangdon	2020-01-15-0317ee7e-43a8-11ea-9696-97d6e81c3666							

- **Dropbox file 1**

- Fill in the archived English filename:

YYYY-MM-DD-Province-Source-No.-EN.html

Date of release

Serial number in the same date

extract - Excel							
Formulas Data Review View Help Search							
AH	AI	AJ	AK	AL	AM	AN	AO
Under contact observation	Other	URL1	Dropbox file 1	URL2	Dropbox file 2	Extractor	Check
148	gender, age, ons	http://www.zjw.gov.cn/zw<U+56FD><National	2020-01-21-HENAN-Health Commission of Henan P(X				AB
148		http://www.zjw.gov.cn/zw<U+5E7F><Guangdon	2020-01-22-HENAN-Health Commission of Henan P(X				AB

③ Extract data

- **Other**

- ‘individual case/death report included’
- Additional information (when you cannot find a corresponding column)
 - Number of visits to fever clinics
 - Distribution of age and job among patients

Note: when there are only individual cases report but NO summary numbers of cases, NO need to extract

- **Extractor**

- enter your initials after completing the extraction

extract - Excel							
ulas Data Review View Help Search							
AH	AI	AJ	AK	AL	AM	AN	AO
Under contact observation	Other	URL1	Dropbox file 1	URL2	Dropbox file 2	Extractor	Check
148	gender, age, ons	http://www.zjwjw	2020-01-21-HENAN-Health Commission of Henan P	XX			AB
148		http://www.zjwjw	2020-01-22-HENAN-Health Commission of Henan P	XX			AB

Task completed!

- Please submit the whole folder through email
 - [html files](#)
 - Archived with 'standard' dropbox file names
 - [extract.xlsx](#)
 - Table filled with extracted information
- Please write down the following notes in the email....
 - Filenames with individual case reports
 - Any questions/potential errors you found



Thank you very much. We really appreciate your help!

