



CONSULATE GENERAL
OF DENMARK
Shanghai

INNOVATION
CENTRE
DENMARK



REPORT

Chinese Open Data in ***PUBLIC HEALTH***

DECEMBER 2018

Table of Contents

1.	Introduction: Chinese open data in the area of public health	2
2.	The Data Center of China Public Health Center.....	3
3.	Biologic Medicine Information Center, National Scientific Data Sharing Platform for Population and Health	4
4.	Population and Reproductive Health Science Data Center, National Scientific Data Sharing Platform for Population and Health	4
5.	Public Health Knowledge Integrated Database	5
6.	Zhejiang Provincial Government Service Network	6
7.	Anhui Provincial Government Affairs Open Network.....	6
8.	Shanghai Municipal Government Data Service Center – Area of Hygiene and Health	7
9.	Other city-level platforms: Nanjing, Ningbo, and Wuxi	8
10.	Private data providers: Juhe and Tianyancha	8

1. Introduction: Chinese open data in the area of public health

This report examines Chinese open data resources available to international researchers in the area of public health.

The trigger of open health data in China was the SARS epidemic in 2003, China's worst public health crisis in the reform era. The government was criticized for its late response and a lack of transparency, with hundreds of thousands of people left disabled or dead. Data collection and monitoring to detect and deal with public health treats thus became a top priority, triggering the build-up of the National Health Information System.

The system was implemented in 2006 with the aim of systematizing and standardizing data collection; improving transparency of health and medical services; spur research and innovation; and enhance the ability to counter public health threats in a timely manner.

The data is collected from local hospitals, clinics, community healthcare centers, and other care providers, as well as pharmacies and food and drug production facilities. Statistics of patient demographics, diseases, health conditions, and risk factors are also collected through surveys and various research programs. A portion of the data is made available to the public.

The data collection is typically coordinated by local government agencies, most commonly city or township Health and Family Planning Commissions. Other local government agencies involved in data collection include Municipal Human Resources and Social Security Bureaus and Municipal Food and Drug Administration. The data is then transferred to provincial level government agencies, typically the provincial Health Departments. They in turn aggregate the data and transfer it to national-level institutions, such as the Ministry of Health, for final compilation and storage.

Rather than one unified system, China's open data initiative lacks central coordination and spans a number of different databases. Some local governments are more progressive than others when it comes to data openness. The municipal governments in Beijing and Shanghai are leading in terms of the number of open datasets. Other local governments with open data initiatives tend to incorporate a small number of basic datasets into their public health services portals.

This report describes 12 databases in the area of public health. Four of them are national-level databases; two are province-level databases; four are municipal-level databases; and two databases are provided by private enterprises. The geographical focus of the report is the lower Yangtze River Delta.

Overall, most of the databases contain aggregated data, rather than original data on the level of individual observations. For example, the data could describe the incidence of different diseases in different segments of the population, rather than patient-level information. The exception is datasets where care providers are the unit of observation, where detailed information is available on the level of individual care providers. Most of the sources are in Chinese only. Some datasets are updated frequently, in an extreme case even daily, while others have not been updated for more than a decade. In some datasets, the date and updating frequency is not mentioned. Most of the datasets are available for research use, though some require authorization.

In terms of data quality, most of the datasets build on information that has been collected manually, leading to a risk of errors and inconsistencies. Despite standardization attempts, different datasets use different formats. For example, the data platforms of both Shanghai and Nanjing provide information about hospitals, but the two datasets contain different variables. Since most of the data collection takes place on the level of care providers, double counting could potentially be an issue, in cases where patients visit more than one care provider.

2. The Data Center of China Public Health Center

URI: <http://www.phsciencedata.cn/Share/>

The website is under the management of the Center of Disease Control and Prevention and covers six categories: infectious diseases, chronic non-infectious diseases, health risk factors, information about deceased, general population demographics, and “others”. The database mainly records nationwide demographic data for each category.

Main content

- Infectious diseases and Chronic non-infectious diseases
 - Include aggregate patient data on the level of diseases, including number of reported cases, number of deaths, and incidence by geographic region, gender, age, and occupation.
- Health risk factors
 - Datasets related to nutrition and risk factors, with incidence in different demographics including by geographic region, gender, age, occupation, and eating habits.
- Information about deceased
 - Aggregate statistics about deceased, including death cause, region of residence, gender, age, occupation, social factors, and health condition.
- General population demographics
 - Nationwide population statistics including region, gender, age, occupation, and disease related information, which can be used as reference points for the other datasets in the database
- Others
 - Nationwide demographic data related to parasitic diseases, rabies vaccination, viral hepatitis, and other special diseases.

Main data providers

- The Center of Disease Control and Prevention (CDCP)
- Provincial CDCP branches

Data usage and limitations

- Available in English.
- Many of the datasets have not been updated in the last decade.
- Registration is required before being able to access the database.
- The data can be used, disseminated, and shared freely, but cannot be used commercially.

3. Biologic Medicine Information Center, National Scientific Data Sharing Platform for Population and Health

URL: <http://www.bmicc.cn/web/share/home>

The platform was established as part of a medical and scientific data management and sharing service system. The data center gathers data sources related to biomedicine, with the overall theme of health and diseases. It provides dataset portals in four different categories, of which one, “The Human Database”, is related to public health (other categories relate to cellular and molecular data). The datasets are mainly based on scientific research projects and experiments.

Main content

- Physiological Reference Database of Chinese
 - Provides reference data of physique and health levels of different segments of the population, including information about the physical development of Chinese children, functions of vital organ, and disease trends based on national surveys.
- Psychological Reference Database of Chinese
 - Contains information on about 70 psychology traits for different segments of the population, including those related to intelligence and interpersonal communication, based on national surveys and research.
- Maternal and Child Nutrition Reference Database of China
 - Contains indicators reflecting the nutritional status and health conditions of women and children based on nationwide investigation.

Main data providers

- Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences
- Chinese Center for Disease Control and Prevention
- Capital Institute of Pediatrics

Data usage and limitations

- Available in English.
- The age group 7-80 years old are covered in the data
- The databases can be used as references for research.
- Some databases (e.g. Physiological Reference Database of Chinese People) require application for full access.

4. Population and Reproductive Health Science Data Center, National Scientific Data Sharing Platform for Population and Health

URL: <http://www.e-healthdata.cn/healthdata/>

This database gathers and shares data sources in the field of population and reproductive health. Other materials such as training courses, publications, and standard specifications are also provided.

Main content

- Neonatal medicine

- Contains data on incidence of neonatal genetic and metabolic diseases based on survey data collected from 12 provinces.
- Dataset with statistics on lifestyle and dietary factors among pregnant women and how they affect health conditions of newborn, covering 31 provinces.
- Child diseases
 - Aggregate data on the family history of children with different critical and rare diseases, based on data collected in Sichuan, Henan, and Hubei Provinces.
- Reproductive system diseases
 - Data on female reproductive health and how different conditions are affected by lifestyle factors, based on survey data collected in 31 provinces.

Main data providers

- National Institute of Science and Technology of Health and Family Planning
- Commissions of Health and Family Planning (municipal level) and Health Departments (provincial level)

Data usage and limitations

- While most datasets are from 2017, some have not been updated since 2010-2014
- The databases can be used as references for research and analysis.
- Full data of some databases require application to get access.

5. Public Health Knowledge Integrated Database

URL: <http://qgwszk.cnki.net/>

The database is the outcome of the National Knowledge Infrastructure (CNKI) project and is part of the national knowledge resources repository, which has developed a collection of journals, doctoral theses, master theses, conference papers, newspapers, reference books, yearbooks, patents, standards, and overseas literature resources. The database covers four areas: preventive medicine and hygiene, medical technology, biology, and information technology.

5

Main content

- Preventive medicine and hygiene
 - Publications in areas including personal hygiene, birth control, epidemiology, epidemic prevention, nutrition, and child health.
- Medical technology
 - Publications in areas spanning medical regulation, clinical medicine, Chinese materia medica, and technology related to the functioning of human organs.
- Biology
 - Publications in areas such as physiology, biochemistry, molecular biology, genetic engineering, and cytology.
- Information technology
 - Publications covering the application of computer technology in biological sciences, medicine and hygiene

Main data providers

- More than 5400 Chinese journals, newspapers and industry associations.

Data usage and limitations

- Not available in English.
- Registration is required for being able to use the search function.
- The user needs to pay to download documents.

6. Zhejiang Provincial Government Service Network

URL: <http://www.zjzfw.gov.cn>

The website is an online government platform providing multiple services, including access to open data. The open data spans eight areas, of which one is related to public health.

Main content

- Medical healthcare
 - Basic information, such as contact information and services covered by different insurances, of healthcare providers in Zhejiang Province.
 - Catalogs of medicines covered by healthcare insurances in Zhejiang Province.
- Food safety
 - Standards of food safety for enterprises in Zhejiang Province.
 - Information of businesses with catering service licenses in Zhejiang Province.

Main data providers

- Zhejiang Provincial Commission of Health and Family Planning
- Zhejiang Provincial Office of Human Resources and Social Security

Data usage and limitations

- The website and databases are not available in English.
- The data can be downloaded freely, but data use requires authorization.

7. Anhui Provincial Government Affairs Open Network

URL: <http://www.ahzwgk.gov.cn>

What data is available

The website is an online government service platform to disclose information regarding working system, progress, and results. There is a special column for information disclosure of important areas, including public service & people's livelihood and public regulation, where medical & healthcare and food & drug safety are covered.

Main content

- Medical and healthcare
 - Statistics of services and operations of medical institutions in Anhui Province. E.g. total number of patients, number of emergency patients, number of admissions and discharges of each institution.
- Food and drug safety
 - Supervision and inspection information of food, medicine, medical instrument, etc. in Anhui.
 - Exposure of illegal behaviors and notifications of penalty results.

Main data providers

- Anhui Provincial Commission of Health and Family Planning
- Anhui Provincial Food and Drug Administration

Data usage and limitations

- Not available in English.
- The data can be downloaded freely and can be used as reference for research and analysis.
- The website not only includes datasets but also news and notifications, so researchers need to search and select based on requirements.

8. Shanghai Municipal Government Data Service Center – Area of Hygiene and Health

URL: <http://www.datashanghai.gov.cn>

The website is a government data service portal covering 12 different areas including Hygiene and Health, where data resources related to public health are provided. The database contains statistics of medical conditions in Shanghai and an overview of authorized medical institutions and health-related enterprises in Shanghai.

Main content

- Healthcare
 - Data on operations of medical institutions, e.g. total number of patients, number of emergency patients, and number of admissions and discharges of each institution.
 - Statistics on population, health, and diseases, e.g. death causes, screening of common diseases, and maternal and children health conditions in the whole city.
- Medical and sanitary management
 - Basic information (e.g. name, address, and contact information) of medical institutions, healthcare centers, health examination institutions, and other care providers.
 - Basic information (e.g. name, address, and license number) of enterprises engaged in food management, medical equipment production, drug handling, and cosmetics production.

Main data providers

- Shanghai Municipal Commission of Health and Family Planning
- Shanghai Municipal Human Resources and Social Security Bureau
- Shanghai Municipal Food and Drug Administration
- Shanghai Municipal Development & Reform Commission

Data usage and limitations

- Most of the datasets are updated yearly, some are updated quarterly.
- Not available in English.
- The datasets are mainly intended to inform citizens and to ensure public transparency.
- The data can be used, disseminated, and shared freely.

9. Other city-level platforms: Nanjing, Ningbo, and Wuxi

Nanjing Municipal Government Data Resources (<http://data.nanjing.gov.cn>)

Ningbo Municipal Government Data Service Network (<http://www.datanb.gov.cn>)

Wuxi Municipal Government Data Service Network (<http://etc.wuxi.gov.cn/opendata/>)

The websites are local government service platforms, typically providing multiple services, of which open data access is one. The open datasets contain catalogues with information about care providers, insurance coverage, pharmacy operations, and nursing institutions.

Main content

- Care providers and pharmacies
 - Lists of hospitals, specialist clinics, and other care providers covered by medical insurance in each city.
- Pharmacies
 - Lists of licensed pharmacies, covering basic information
- People's livelihood services
 - Lists of community health service centers and family planning service centers, covering basic information
 - Lists of nursing institutions and community centers for the elderly, covering basic information
 - Lists of Infant care providers, covering basic information

Main data providers

- Municipal Food and Drug Administrations
- Municipal Commission of Health and Family Planning
- Municipal Human Resources and Social Security Bureau
- Some datasets are derived from Zhejiang Provincial Government Service Network

Data usage and limitations

- The websites and databases are not available in English.
- Some of the databases require registration and usage requires authorization.

10. Private data providers: Juhe and Tianyancha

Juhe (<http://www.juhe.cn>)

Tianyancha (<http://www.tianyancha.com>)

Two private data providers Juhe and Tianyancha, both with large databases in a wide range of fields, provide detailed information about Chinese business in public health, on the level of individual businesses. The information can be accessed through a free or premium version search engine (Tianyancha) or through an application programming interface (API) with a fee per request (Juhe and Tianyancha).

Main content

- Tianyancha provides detailed company information about individual business in different public health categories, including registered capital, ownership and investments, business scope, patents, recruitment ads posted by the company, credit score, and legal cases that the business has been involved in. Basic search is free of charge, while advanced search costs 10,000 yuan/year. The database updates daily.

Innovation Centre Denmark, Shanghai

- Juhe provides hospital information in Beijing and Shanghai. First 1000 requests free of charge, then 99 yuan per 10,000 requests.

Main data providers

The information comes from a wide range of data providers, including the Commerce Bureau, various recruitment websites, the National Intellectual Property Administration, and the open data platforms of the local governments of Beijing and Shanghai.

Data usage and limitations

- API access brings flexibility and the possibility to monitor data over time
- All data is in Chinese
- Commercial use is allowed, but advanced search options and large-scale downloading are not free of charge.
- The data is updated with very high frequency



Innovation Centre Denmark

It is the aim of Innovation Centre Denmark to strengthen research, innovation and higher education in Denmark and to promote Denmark as a leading knowledge society. The Danish innovation centres can provide you with access to knowledge, technology and networks in selected global leading research, education and innovation environments. Innovation Centre Denmark in Boston, Munich, New Delhi/Bangalore, Sao Paolo, Seoul, Shanghai and Silicon Valley are established in cooperation between the Ministry of Foreign Affairs and the Ministry of Higher Education and science. More information: www.icdk.dk.