CAICT 中国信通院

Research Report on Data and Smart Applications for Epidemic Prevention and Control

(Version 1.0)

China Academy of Information and Communications Technology (CAICT)

March 2020

Copyright Notice

Foreword

At the twelfth meeting of the Central Committee for Deepening Overall Reform on February 14, 2020, General Secretary Xi Jinping put an emphasis on “encouraging the use of digital technologies such as big data, artificial intelligence (AI), and cloud computing to better support epidemic monitoring and analysis, virus attribution, epidemic prevention, control and treatment, and resource allocation". In the fight against the epidemic caused by novel coronavirus pneumonia (Covid-19), governments, medical institutions, scientific research institutes, and tech companies have all responded quickly and adopted big data and other smart technologies in epidemic monitoring and analysis, personnel management, medical treatment, resumption of work and production, and other aspects, which have played a significant role in supporting the epidemic prevention and control (EPC) work.

In order to present a panoramic view of the applications of big data and AI for EPC, Cloud Computing and Big Data Research Institute, CAICT, as a lead institution, in collaboration with many other partners, collected and analyzed more than 200 cases, sorted out how different sectors adopted big data and data intelligence technologies for EPC, analyzed the key links of data-driven EPC, discussed the problems and challenges in this regard, and conducted preliminary discussions on using next-generation information technologies such as big data for the next-step EPC work.

Due to time constraints, the cases collected by the authoring team may not be comprehensive enough and the analysis of the current results and existing problems may not be thorough enough. In addition, the EPC work has not yet ended and the effects of some applications need to be further observed. Many applications are still being launched, and it is too early to make solid conclusions. In view of this, we may supplement and improve the report as needed in the future. You are welcome to contact the authoring team (liuhan@caict.ac.cn) and provide your valuable feedbacks. You are also welcome to continue to actively provide cases where big data applications have brought concrete benefits for EPC and resumption of work and production.

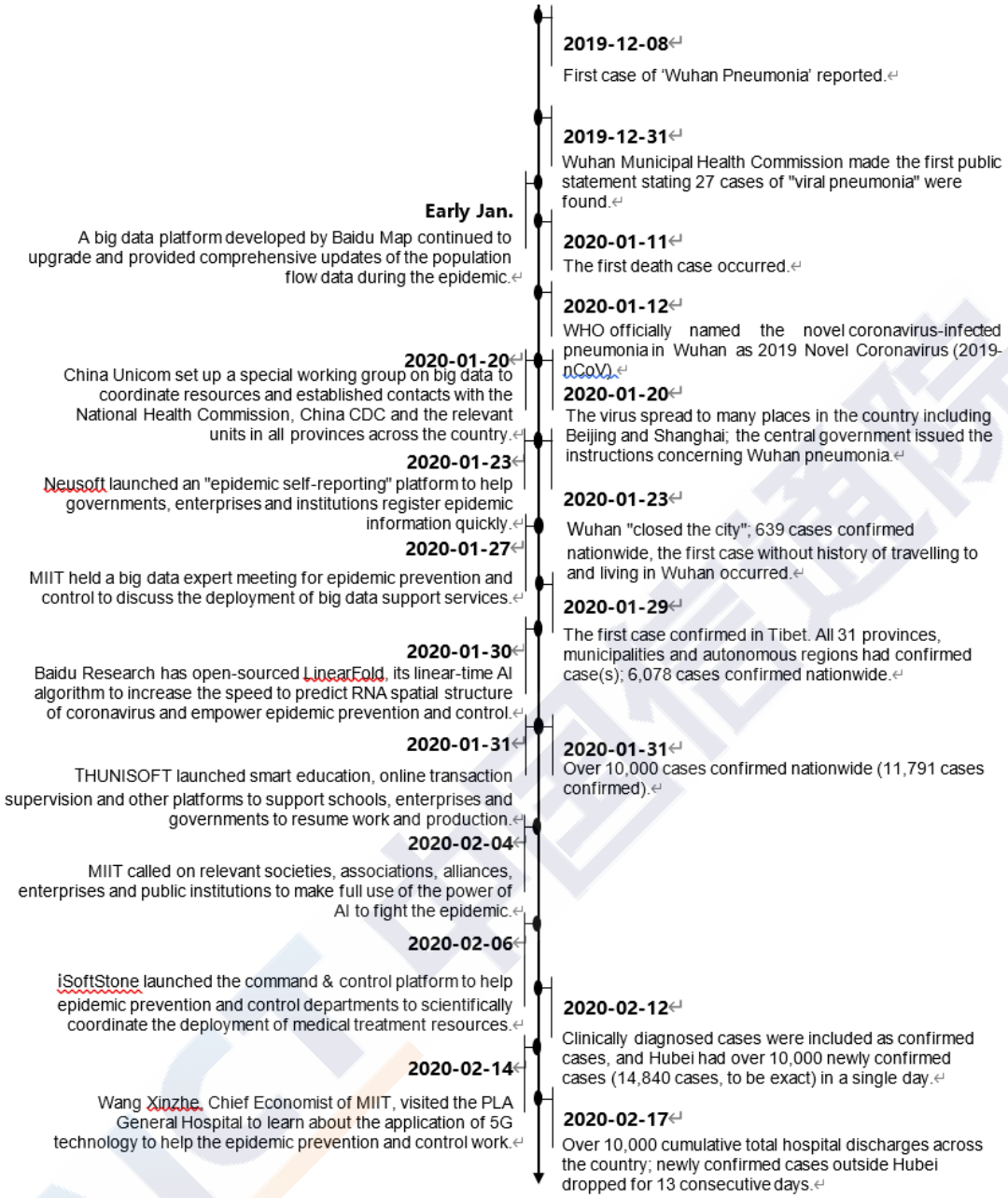
Table of Content

I. Summary

1. Timeline for Using Technologies to Fight the Epidemic

The Covid-19 epidemic broke out suddenly at the beginning of 2020. Due to the highly contagious, widespread and risky nature of the disease, the epidemic prevention and control (EPC) work is faced with arduous tasks, limited time, and a grim situation. In this battle against this epidemic, the fast-evolving next-generation information and communication technologies (ICTs) such as big data, cloud computing, and AI are being deeply integrated with transportation, healthcare, education, and other fields in an accelerating manner. They have become powerful weapons to fight the "epidemic", making the organization and implementation of EPC more efficient and effective.

This report summarizes the time axis, which includes the timeline for the development of epidemic and the timeline of the cases of using data and intelligent technologies for EPC. It can be seen that with the development of the epidemic situation, the data-driven EPC is rapidly unfolding, new application scenarios continued to emerge and the application scope continued to expand.



Source: CAICT

Figure 1 - Time Axis of Using Technologies to Fight the Epidemic

2. "Data" Axis of Using Technologies to Fight the Epidemic

By analyzing and summarizing these application scenarios, we found that "data" can empower EPC in the following ways.

(1) Provide Strong Support for the Dissemination of EPC Knowledge

With the help of mobile Internet and smart phones, people can access various data anytime and anywhere, such as epidemic updates and scientific knowledge about epidemic prevention. Local governments have regularly released the epidemic updates through e-government platforms, Weibo, WeChat official accounts and other channels. Various news clients, social networking platforms, search engines, and short video platforms have also actively supported the release and dissemination of epidemic information. Moreover, many "Internet + health" platforms have launched online consultation services to facilitate netizens to consult doctors about the prevention and treatment of Covid-19, which have effectively alleviated the difficulties faced by the people to seek for medical advice caused by the shortage of medical resources during the epidemic, and avoided cross-infections caused by the gathering due to the influx of people with common illnesses into hospitals.

(2) Quickly Track the Movement Trajectory of "Affected" People

By integrating the information of telecom operators, Internet companies, transportation departments and other organizations, big data technologies can analyze the movement trajectory of the people. Specifically, on the basis of data analytics and mining technologies, on the one hand, the movement trajectory of a patient can be drawn based on data that includes geographic location and time stamp information, such as mobile signaling data; on the other hand, according to the movement trajectory of the patient in a period before the diagnosis date and the people staying with the patient for a relatively long period, close contacts can be inferred based on big data analytics. By conducting comprehensive analysis of the movement trajectories of diagnosed patients, suspected patients, and related contacts, we can present an accurate picture of the movement of different types of people who roam in and out of a region. This not only provides powerful guidance for precise treatment, but also an accurate basis for predicting high-risk areas and potential high-risk areas.

(3) Predict the Epidemic Trend and Conduct Attribution Analysis

Based on the relevant data of high-risk population, combined with the number of newly diagnosed, suspected, death, and cured cases, and with the help of big data analytics models and practical techniques such as transmission dynamics models, dynamic infection models, and regression models, it is possible to analyze and display the heat distribution of the infected and the heat distribution of the risks of close contacts and make important analysis and judgement on the trend of the epidemic, such as the peak and turning points of the epidemic. We can use emerging AI technologies such as deep learning, combined with a large amount of data such as movement trajectory information, social networking information, consumption data, and exposure history, to establish scientific models, find spatiotemporal collision points based on the patient's diagnosis time, close contacts and other information to calculate the transmission path of the disease, and provide a theoretical basis for the attribution analysis of infectious disease.

(4) Help Local Governments Implement Scientific and Targeted Policies

Based on big data analytics and combined with algorithmic models, we can predict the spread speed and trend of the epidemic and thus provide local governments with an effective basis for dynamic monitoring and management, overall planning of medical materials and reserves, guarantee of the supplies for people’s livelihood, and formulation of traffic control policies. For example, based on the travel data of people in high-risk areas during the Spring Festival, epidemic analysis and early warning can be realized and accurate prevention and control can be fulfilled by tracking the trajectories of diagnosed patients, suspected patients and close contacts. At the same time, big data analysis can be used to assess and predict the short-term and long-term impact of the epidemic on socio-economic operations, establish a rapid and efficient economic emergency response mechanism, and help the government introduce various measures such as tax cut, fee reduction, and special subsidies in a timely manner so as to alleviate the risks of the broken capital chain caused by the epidemic situation and possible difficulties in continuous

operations faced by SMEs, and strive to maintain a smooth and orderly production and life[1].

(5) Promote Case Diagnosis and Epidemic Research

Using big data, AI and other related technologies, we can effectively accelerate the data analysis and computing work for host prediction and drug screening and greatly improve the efficiency of virus research and treatment.

[1] https://mp.weixin.qq.com/s/vuH0jg7L15rDyNNyKjrgvQ, *ICTs Help Win the War of Epidemic Prevention and Control*

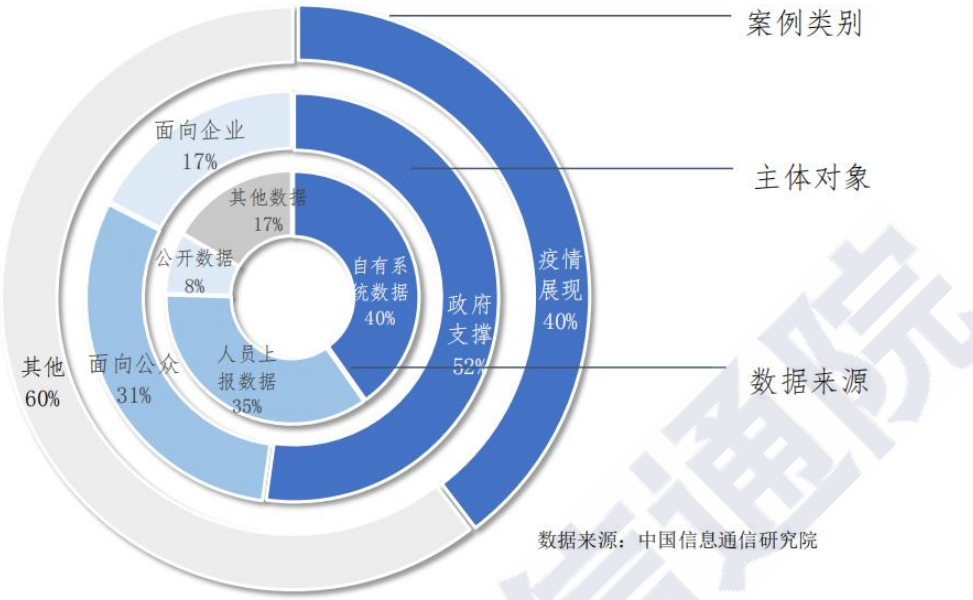
II. Application Analysis

This chapter summarizes the data application cases in five major aspects, including epidemic analysis and display, the epidemic prevention and control (EPC), improving efficiency of healthcare and medical treatment, increasing convenience and benefits for the people, and resuming work and production. The cases are obtained from Internet channels and corporate reporting channels. More than 200 cases are collected and analyzed. The selected cases have achieved good results in practical applications and played an important role in the EPC process.

1. Epidemic Analysis and Display

The epidemic data we see every day when we open the major news apps on mobile phones are typical applications of epidemic analysis and display. It can be said that the epidemic-related data are the basis for carrying out various aspects of epidemic situation analysis, management and control. Therefore, many companies have first performed multi-entity, multi-channel, and multi-dimensional display of the epidemic based on the data from various sources.

According to the survey results of this report, 39.6% of the enterprise cases are related to visualized display or visualization of collected data, among which more than 80% have performed more in-depth data analysis on the basis of visualization; in terms of data sources, there are public data (8.1%), personnel reporting data (36.1%), self-owned system data (41.3%), and other system and third-party data (17.1%). There is an overlapping of data source channels among the various types of cases; by service objects, 52.2% of the enterprise cases provide data collection and analysis services for the governments, 30.4% provide information display services for the public, and 17.4% provide data analysis and display services for the enterprises.



案例类别	Case Categories
主体对象	Service Objects
数据来源	Data Sources
疫情展现	Epidemic Display
其他	Others
政府支撑	Government Support
面向公众	For the Public
面向企业	For Enterprises
自有系统数据	Self-owned System Data
人员上报数据	Personnel Reporting
公开数据	Public Data:
其它数据	Other Data
数据来源：中国信息通信研究院	Source: CAICT

Figure 2 - Epidemic Analysis and Display

This section discusses the value of epidemic analysis and display applications from the perspective of three types of service objects.

(1) Government Support

In terms of government support, the epidemic analysis and display applications mainly include such core functions as the display of epidemic-related information, personnel flow, vehicle flow, epidemic-related resources, and logistics information within the scope of government control.

Through real-time and comprehensive display of these important information, technologies can support the governments in the prevention, management and control of the epidemic.

Since the launch of the " Population Big Data Platform for EPC" on January 26, China Mobile Beijing Branch (also known as Beijing Mobile) has provided CPC Beijing Municipal Committee and Beijing Municipal People's Government, 13 districts and counties, and relevant commissions and bureaus in Beijing with epidemic-specific analysis services, such as users from epidemic areas, users returning to epidemic areas, users from other provinces, users returning to other provinces, monitoring and distribution of non-resident users, and profiles of various groups of people and distribution heat maps. It has provided data support throughout the entire process from accurate epidemic defense, screening, monitoring, and advocacy for high-risk groups, potential high-risk groups, and potential risk groups. Hangzhou Dt Dream launched the "City Brain" project in Quzhou, Zhejiang Province. Through the big data analysis cockpit, it can display the grid-level distribution of local personnel, and monitor and screen key non-local vehicles in the city, with false negative rate of less than 1%. In addition, the epidemic analysis and display platforms of Neusoft, Sefonsoft, DataOJO, idatasing, LongShine and Meishu Tech also provided strong support for local governments in their EPC work.

(2) Public Services

From the perspective of public services, the epidemic data analysis and display applications mainly include such core functions of display of epidemic information, population migration, epidemic-specific services, public opinion monitoring and evaluation, and public information reporting and display. These applications can disseminate the dynamic epidemic information to the public in a timely manner and provide effective methods for EPC.

Baidu Map’s Migration Big Data platform has expanded the open access services from 100 to over 300 cities. The platform contains rich data indicators, including source, destination, migration scale index, and migration scale trend map. It even supports querying the data about the ranking and percentage of the migration origins and destinations of a city since the Spring Festival and

offers comprehensive and three-dimensional migration big data services and support. At the same time, "Citizen Epidemic Risks Reporting System", which was introduced by JD Cloud, enables citizens to take pictures and report potential risks, and the geographic location will be automatically identified. The generated data are safe and isolated. The system has a dedicated data backend management feature, enabling citizens to promptly and efficiently report the epidemic-related information. Various epidemic analysis and display services and systems were launched to provide strong support for public epidemic prevention work, including the epidemic columns introduced by WeChat and Alipay, "Community Electronic Pass for Epidemic Prevention" launched by Baidu, the " EPC Services on the National Integrated Platform" introduced by Taiji Group, the “Covid-19 Dynamic Display System” developed China Mobile Xiong'an ICT Co., Ltd, and the epidemic analysis and display systems released by Beijing Percent IT, istarshine, and YUN (www.sxdata.com.cn) have all provided strong support for public epidemic prevention work.

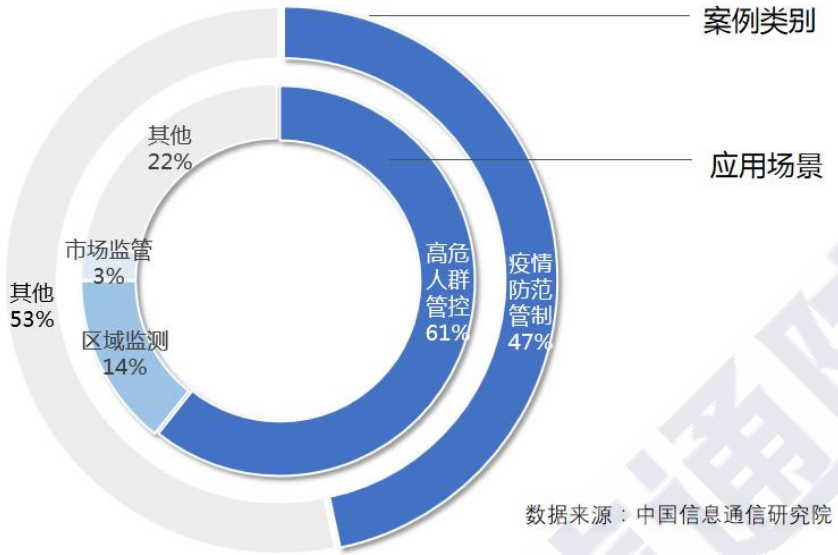
(3) Enterprise Services

In terms of enterprise services, most enterprises achieved effective prevention, control and management of the epidemic situation within the enterprises through self-built or purchased epidemic analysis and display products.

CASIC Network and Information Development Co., Ltd has independently established an enterprise epidemic management and control platform, focusing on real-time display of Covid-19 information, screening of high-risk groups, auxiliary diagnosis of diseases, and tracking of charitable donations.

2. Epidemic Prevention and Control (EPC)

Big data analysis and display has largely completed the data collection and organization work. Through deeper data modelling, analysis and mining, the value of data can be further tapped on the original basis. Epidemic prevention applications use data to identify high-risk groups, perform such function as area monitoring and market supervision, and provide support for the decision-making of government departments.



案例类别	Case Categories
应用场景	Application Scenarios
疫情防范管制	Epidemic Prevention and Management
高危人群管控	High-Risk Group Management
市场监管	Market Supervision
区域监测	Area Monitoring
数据来源：中国信息通信研究院	Source: CAICT

Figure 3 - Epidemic Prevention and Control (EPC)

According to the survey results of this report, the proportion of technical solutions and scenarios that can effectively support the epidemic management and control reached 46.7%. It can be seen that the use of various technical means to improve the means and efficiency of the epidemic management and control has become the most popular application scenarios during the epidemic. In general, among the various application scenarios of epidemic management and control, the most important is the monitoring and control of high-risk populations. The proportion of cases with such a feature reaches 60.7%; other management and control means, such as area monitoring (14.3%) and market supervision (3.1%) also have some level of applications. Another important result shows there are 32.1% cases, in which effective technical means and solutions are developed through technical R&D and application adaptation, thus increasing the support capacity for epidemic management and control. Intelligent outbound calls, image recognition, rapid

integration of microservices, high-dimensional machine learning, knowledge graph, spatiotemporal data analysis, and visual representation technologies have all become powerful means.

(1) Management and control of high-risk population is a top priority.

Effectively identifying the movement trajectories of high-risk groups and contacts through location data and various behavior data can fundamentally reduce the spread of the epidemic, which is also a very important task for government departments at all levels.

China Unicom Big Data (CUBD) has developed a series of data models such as transmission risk analysis and spatiotemporal correlation analysis. Through integrated analysis of multi-dimensional data, CUBD can identify and analyze the trajectory of population in specific areas and the contact range of confirmed cases, effectively supporting the government's regional EPC work. An epidemic search tool developed by China Telecom Corporation Limited Cloud Computing Branch Corporation can provide the path of high-risk population in the past 14 days and make risk judgements on close contacts through data analysis of the telecommunications GIS system, which effectively supports the EPC work of government departments. In addition, TravelSky push notification system, which can send an alert to air travelers if any of their fellow passengers is confirmed with Covid-19, the “Covid-19 Spread Monitoring System” introduced by Meyer Pico and the “Epidemic Screening Management and Reporting System” launched by Sugon Cloud Computing Group Co., Ltd all have the high-risk population screening and monitoring features, providing strong support for government departments at all levels in their EPC work targeted at high-risk population.

(2) Technologies Are Applied in Various Management and Control Scenarios

In addition to the monitoring and control of high-risk population, market supervision, and health tracking of people in an area are also important application scenarios for epidemic prevention products.

Thunisoft's "Online Transaction Supervision" system, based on the smart collection and analysis of online transaction information, has helped Yunnan Provincial Administration for Market Regulation effectively control the price fluctuations of online transactions during the epidemic and accurately conduct market supervision on the EPC products, thus ensuring the protection of people's livelihood. CUBD has developed a system to monitor the compliance of monitoring staff in EPC, which can provide real-time alarms on mask wearing, people gathering, and abnormal body temperature and enable governments to understand the epidemic situation in real time. JD Cloud has developed an epidemic information release product, which provides "multi-dimensional", "visualization" and "five in one" (people, address, events, objects, organizations) information release and EPC services for public management institutions, provides the grassroot organizations with a dynamic and grid-based assessment of the health dynamics of workers returning from the epidemic area, and achieves accurate area epidemic prevention and management.

(3) Technologies Provide Integrated Support

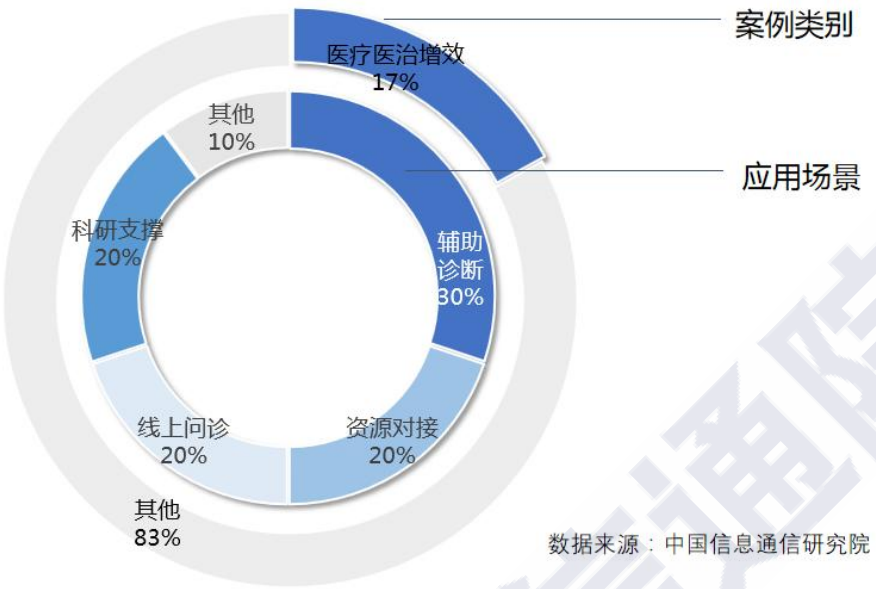
The technological capabilities of tech companies are the core driving forces for the breakthroughs in epidemic prevention work. The wide applications of AI image recognition, intelligent outbound calls, knowledge graph, secure multi-party computation (MPC), and microservices, have effectively improved the efficiency and safety of EPC work.

Based on AI image recognition technology and infrared thermal imaging technology, Baidu has developed an AI temperature measurement system, which can quickly screen the forehead temperature of multiple people at the same time and issue warnings, facilitate the rapid screening of crowds in crowded area, and effectively reduce the gatherings in public places. At the same time, Baidu quickly launched a "free intelligent outbound call platform for EPC", which can provide outbound calling services for three major scenarios: check of mobile personnel, local residents screening/return visit, and notification to specific groups. It can effectively support such application scenarios as community situation check, notification and return visits. It is hundreds of times more efficient than manual calls. This platform has been

used by more than a dozen regions, including Beijing and Shanghai. Relying on the advantages of Alibaba Cloud's platform, the "Epidemic Information Collection System" can effectively use the backend microservice module function through visualized drag and drop operations and quickly support the work of the health committees in 11 cities in Zhejiang province. In addition, Didi designed a special "orange video dashcam" (in-car video recorder) to collect the pictures and detect whether the driver is wearing a mask through an AI recognition algorithm. The "Intelligent Epidemic Combat Platform" developed by DataExa based on the knowledge graph, the "EPC Analysis System" built by Sefonsoft based on the three-dimensional city model, and the "Smart Analysis Platform for Disease Control" developed by Insight Smart Technology Co., Ltd. based on spatiotemporal big data and MSC technologies have effectively supported the epidemic management and control work of government departments, enterprises and public institutions at all levels.

3. Improving Efficiency of Healthcare and Medical Treatment

In the fight against the epidemic, big data and intelligent technologies have been fully applied to scenarios directly related to healthcare, such as disease diagnosis, medical research, and medical assistance. It is the most severe test of big data technologies. According to the survey results of this report, nearly 17% of the cases are the applications to improve the efficiency of medical treatment. These applications include resource matching, auxiliary diagnosis, online consultation, scientific research support, and others (including contactless temperature monitoring applications based on image analysis to identify high-risk population timely). Among them, auxiliary diagnosis refers to assisting or accelerating the diagnosis of confirmed cases through AI technologies; online consultation refers to reducing the pressure on medical institutions through intelligent consultation services; scientific research support refers to improving the efficiency in scientific research through open algorithms, models, or providing computing storage, and helping genetic testing, and vaccine R&D. Please refer to Figure 4 for the specific shares of different application scenarios.



案例类别	Case Categories
应用场景	Application Scenarios
医疗医治增效	Improving Efficiency of Healthcare and Medical Treatment
辅助诊断	auxiliary Diagnosis
资源对接	Resource Matching
线上问诊	Online Consultation
科研支撑	Scientific Research Support
其它	Others
数据来源：中国信息通信研究院	Source: CAICT

Figure 4 - Medical and Medical Treatment Applications

This section mainly uses specific cases to introduce the specific applications of algorithm, computing power, AI, intelligent consultation and other technologies in disease research, auxiliary diagnosis, and online consultation.

(1) Algorithms and Computing Power Assist Disease Research

From the perspective of scientific research, AI, big data and other technologies are coming into prominence in virus structure analysis and vaccine R&D. On January 30, Baidu Research has open-sourced LinearFold, its linear-time AI algorithm to epidemic prevention centers, gene testing institutions, global scientific research institutions, and it also opened the fastest RNA structure prediction website for free. It has reduced the time taken to predict the novel

coronavirus' RNA secondary structure from 55 minutes in traditional algorithm to just 27 seconds, improved the speed by 120X and saved the waiting time by 2 orders of magnitude. It has greatly improved the efficiency of scientific research and strongly supported epidemics prevention and control. In addition, UEC supported the virus sequencing by providing China CDC with much-needed big data computing and storage resources to ensure large-scale parallel sample analysis, data preservation and management. At the same time, it also provided technical support for the Institute of Microbiology of the Chinese Academy of Sciences (IMCAS) by establishing the process to analyze the evolution relationship of virus genome, and providing a visual display function of the evolution tree to monitor virus mutations in real time and trace the source of the virus host.

(2) AI Accelerates Disease Diagnosis

From the perspective of diagnostic support, many AI technologies are applied to the process of disease diagnosis to help determine the conditions and shorten the diagnosis time. On February 1, Zhejiang Provincial Center for Disease Control and Prevention (Zhejiang CDC) launched an automated genome-wide testing and analysis platform. Based on the AI algorithm developed by the Alibaba DAMO Academy, it can effectively shorten the genetic analysis of suspected cases and accurately detect virus mutations. In addition, HKUST Xunfei's Intelligent Medical Assistant can provide online analysis for medical records of the primary clinics it covers in Anhui province. The Assistant can analyze medical records from the dimensions of fever, cough, dyspnea, epidemiological history (relevant history in Wuhan), imaging, and blood routine, screen potential patients, and provide decision-making reference for Anhui Provincial Health Commission. The AI system introduced by inferVision for Covid-19 can help imaging doctors to screen suspected cases more efficiently, reduce the risk of cross infections among patients waiting in the hospitals, and assist in the diagnosis of early-stage patients when virus detection kits are short of supply. The system was first applied to Wuhan Tongji Hospital, and went live in various regions, including Shenzhen Third People's Hospital (National Clinical Research Center for Infectious Diseases) and other medical institutions in the frontline to combat the epidemic.

(3) Intelligent Consultation Connects Supply and Demand and Improves Logistics Support

From the perspective of medical assistance, big data technologies have been used to reduce pressure on hospitals, reduce crowd gathering, and integrate medical care resources. Both Alibaba and JD have introduced the epidemic service robots, which can provide users with much-needed online consultations and dissemination of epidemic knowledge, reduce the workload of medical staff and alleviate the pressure on hospital outpatients. Baidu's smart health division launched the "smart consulting assistant" to improve the efficiency of doctors' diagnosis and treatment by providing standardized pre-examination paths. It also provided platforms with free open APIs, such as online health consultation platforms, government EPC platforms, and Internet hospitals to fully support smart applications in medical and treatment domains. The "Covid-19 EPC Command & Control Platform" launched by iSSTech helps improve the prevention and control capabilities at the grassroot level by integrating medical rescue resources, and optimizing the supply of prevention and control materials.

4. Increasing Convenience and Benefits for the People

Life service applications are also a key breakthrough for data-driven EPC. Many Internet companies use the O2O service model to form a mapping of offline activities to online activities, use big data technologies to collect, classify and store the massive amount of life service data, and provide residents with contactless food delivery, real-time epidemic maps, Internet medical services and other services. While facilitating the normal life of residents, it also ensures the health and safety of various services.

(1) E-commerce Platform Guarantees Easy Access to Safe Food.

Affected by the Covid-19 epidemic, online shopping is more popular than offline shopping because it reduces the risk of cross infection. Fresh food e-commerce companies worked with partners to break barriers to the sales of agricultural products, increase commodity supply, and make residents' "vegetable baskets" more secure.

E-commerce platforms include Hema Fresh, Dingdong Fresh, Missfresh, Meituan grocery shopping, Tmall Supermarket, Yonghui grocery shopping, JD to Home, and many others. Most platforms choose supermarkets and stores based on the distance to the buyer, and provide all-day delivery service. According to statistics, in terms of product richness, more than half of the platforms vary in the degree of product richness in different stores; in terms of delivery time, more than 80% of platforms can fulfill delivery within 2 hours; in terms of contactless delivery, all platforms can provide contactless distribution services to reduce the risk of cross-infection caused by face-to-face contact; in terms of scheduled distribution, due to shortage of materials and limited transport capacity, all platforms provided scheduled distribution services in the early stage of the epidemic, but with the gradual alleviation in the shortage of materials and limitation in transport capacity, about 30% of the platforms gradually realized the "buy and deliver" service.

At the same time, in order to ensure the safety of delivered goods, multiple platforms have upgraded "contactless delivery" to "secure contactless delivery". Take Meituan Takeaway as an example. On the basis of “contactless delivery”, the health information of chefs, packers, riders, tableware disinfection, and other health information will be presented more transparently to users in the form of "electronic card + physical card" on the merchant and delivery sides. This ensures contactless delivery and realizes a secure delivery model in which the food safety related information is visualized and traceable during the whole process.

Compared with "secure contactless delivery", the "unmanned vehicle delivery" reduces the potential contact risk brought by riders. Baidu provides free support of the self-driving cloud service matrix on Baidu’s Apollo platform to the enterprises using self-driving vehicles in their anti-epidemic operations, helping them to better carry out the R&D of anti-epidemic vehicles. Based on this, Neolix, a partner in Baidu's Apollo ecosystem, launched an unmanned vehicle in Haidian Hospital, and began to provide contactless food delivery services from February 14th, minimizing the staff's field work and exposure. The "unmanned vehicle delivery" provided by the "White Rhino" in Zhongguancun was launched at the "Jiayuan Convenience Takeaway

Supermarket" in Wenquan Town, Haidian District, Beijing. After ordering online, the unmanned vehicle can deliver the goods to the gate of the community, and after the identity of the resident is confirmed, the resident can pick up goods by opening the corresponding door.

(2) Map Application Platforms Guarantee Travel Safety

Since the occurrence of the Covid-19 epidemic, many map application platforms with big data resources have quickly launched travel guides that are conducive to EPC and meet users' travel needs in special periods. Statistics show that more than 80% of the platforms have functions such as "epidemic map display", "fever diagnosis consultation", "co-passenger information inquiry", and some platforms provide functions such as "the latest entry policies of countries". Baidu Map was launched on February 6th with a special map of "epidemic-affected communities". As of February 12, it covered more than 200 cities and towns. In the " epidemic-affected communities" section, you enter the name of the community, it will provide epidemic reminders of the surrounding areas, including the announced epidemic sites, the distance to the nearest epidemic sites, and the surrounding places with crowds of people. With the help of the "heat map layer" of Baidu Map, you can also view the real-time population flow density to help residents decide the travel route and avoid crowded places such as business districts and transportation hubs, as shown in Figure 5. At the same time, "Xiaodu Contactless Elevator Service" uses voice recognition algorithm and automatic control technology, allows you to call the big buttons in the elevator, which greatly reduces the risk of virus transmission due to elevator contact and brings more convenience to the people.

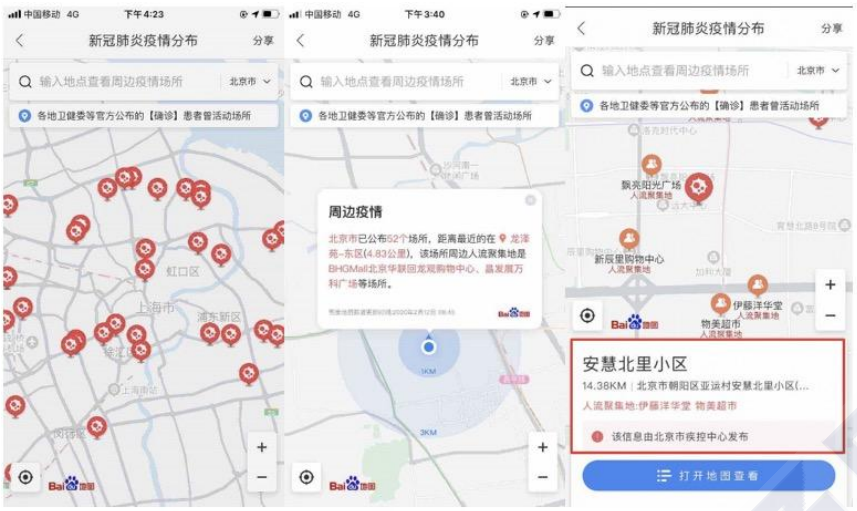


Figure 5 - Schematic Diagram of "Epidemic-Affected Communities" on Baidu Map

With the increase of returning workers, the subway, which is one of the main modes of commuting, will increase the risk of cross-infection if there are too many passengers. Baidu Map cooperated with the Beijing Municipal Transportation Commission and launched a subway passenger flow inquiry service, which is convenient for users to check the congestion level of subway cars at any time through Baidu Map and provide an effective reference for safe travel. AutoNavi launched a feature on February 12 in some cities to enable users to check the full load status of traffic flow in subways, which effectively distinguishes different degrees of congestion, allowing residents to grasp the real-time load of the subway stations and lines and choose the appropriate mode of transportation.

(3) Internet Healthcare helps residents to seek medical treatment

In the field of healthcare, many domestic Internet medical service platforms have made full use of 5G and other ICTs to link doctors and experts across the country to conduct online consultations, online inquiry, health science popularization, and psychological assistance counseling. It enables people to seek medical advice and treatment at home and allows IT to offer "smart help" in EPC. Data from MIIT show that at present, more than 190 public medical institutions and nearly 100 corporate Internet hospitals across the country provide online free consultations, which have alleviated the pressure of offline hospitals. Haodf (Good Doctor Online), DXY, Chunyu Doctor, Ping An Good

Doctor and other enterprises have pooled resources of over 10,000 medical experts in the fields of respiratory, infectious diseases, and internal medicine, providing free consultation for patients. They also opened hotlines for frontline doctors and nurses and provided them with psychological counseling. JD Digits has launched an epidemic inquiry robot. With the help of AI, the robot can learn the professional knowledge related to the epidemic, perform basic analysis and judgment on user questions, and offers services such as epidemic prevention science popularization. At present, the inquiry robots have been used in the WeChat official account or Apps of nearly 20 institutions, and have landed in 23 smart communities in cities like Hangzhou and Suzhou. At the same time, Baidu's smart health, Ali Health and other Internet platforms have provided smart Q&A for the epidemic. Beijing, Tianjin, Zhejiang and other regions are also actively setting up online doctor consultation platforms to tackle the epidemic.

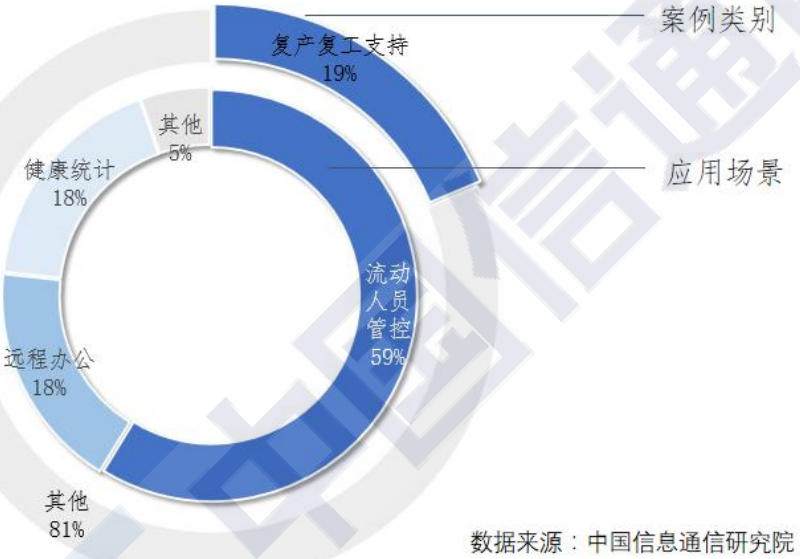
In order to help people buy drugs, multiple platforms have offered services such as fever clinic inquiry and drugs delivery services. Baidu Map launched "Fever Clinic Map", which allows users to view a list of nearby medical institutions that open fever clinics and their business hours. It currently covers more than 300 cities. At the same time, AutoNavi has launched a mini program for drugs delivery in more than 200 cities across the country, and cooperates with partners such as Fengniao Paotui, Dingdang Kuaiyao, and UU Paotui to provide users with fast and safe home delivery services. Alibaba Health launched a "buy medicine at home" service on the Taobao App on February 6 to meet the needs of chronically ill patients seeking medical consultations and prescriptions via the Internet and having the drugs delivered to home.

5. Managing Resumption of Work and Production

Affected by the sudden outbreak, the operation of the catering, accommodation, culture & entertainment, transportation and other industries slowed down or even stopped. The length of time for employees to await job assignments has been extended. Related upstream and downstream industries have also been affected. As positive progress has been made in the prevention and control of the epidemic, local governments and enterprises are gradually

focusing their efforts on resuming work and production so as to resume normal economic and social operations as soon as possible.

According to the survey results of this report, the enterprise cases in this field mainly focuses on the resumption of production, work, and study for enterprises, governments and schools, and assist the government in the EPC work after resumption, as shown in Figure 6:



案例类别	Case Categories
应用场景	Application Scenarios
复工复产支持	Supporting Resumption of Work and Production
流动人员管控	Management of Mobile Personnel
远程办公	Remote Work
健康统计	Health Statistics
其它	Others
数据来源：中国信息通信研究院	Source: CAICT

Figure 6 - Applications for Resumption of Work and Production

(1) Communication Big Data Itinerary Card Provides Geographical Location Query

As the peak of the return journey approaches, preventing cross-infection caused by the movement of people is the focus of epidemic prevention work

in all regions and departments at this stage. In order to simply, quickly, authoritatively and accurately prove the itinerary information in the past 14 days, CAICT and three basic telecom operators worked together to launch a "communications big data itinerary card" service based on telecom big data, providing the 1.6 billion mobile phone users nationwide with the proof of their journey in the past 14 days. The user scans the QR code and enters the mobile phone number and can realize one-stop data query across operators. It is easy and fast to operate and there is no need to install software. It can help all regions and departments identify the risks posed by migrant and returning workers in a timely manner, facilitate governments to take targeted policies, and help the regions make scientific and efficient arrangements for resuming work and production. The communication big data itinerary card was launched on the afternoon of February 29, and as of the evening of March 1, 53,577 queries have been completed.

(2) Health Codes Boost the Orderly Resumption of Work and Production in Cities

At present, many regions have adopted digital management measures to analyze and judge the situation of resuming work, production and study during the epidemic. Multi-level and sector-specific measures are taken for key areas and places to help governments and enterprises to resume work and production in an orderly manner. For example, Hangzhou and Shenzhen have successively launched "Health Code" to perform community and travel management during the epidemic. Citizens or returning workers can apply for health codes on Alipay and WeChat platforms as an electronic pass to enter and exit the community. It requires application only once and can be used across the city. Public places can adopt multi-level and categorized management based on health codes. Recently, under the guidance of the e-Government Office of the General Office of the State Council, Tencent and Alipay will soon launch a nationwide integrated government service platform for health information codes for EPC, which is expected to resolve inconsistent data standards and unavailability of cross-regional data caused by multiple collection channels. It will enable inter-provincial and inter-regional connectivity of anti-epidemic service and better support for national EPC work.

At the same time, the three major operators have launched applications for querying the personal itinerary during the epidemic. CUBD launched the "Health U Code", allowing users to query their 14-day itinerary, report their health status and generate health codes at different levels.

(3) Remote Work Supports the Coordinated Operation of Government, Enterprise and Schools.

In the area of resumption of work and production, remote work and personnel information reporting are the two main directions. Baidu announced that its intelligent remote office platform - Baidu Hi will be open to the public on February 11 and will provide free services such as high-definition audio and video conferencing, enterprise cloud disks, enterprise IM and application centers and platforms for enterprises in affected areas such as Hubei. It aims to meet the growing demand for remote work during the epidemic, support enterprises to quickly restore production capacity, and reduce the impact of the epidemic on enterprises and social and economic development. In response to the delayed resumption of schools, work and production caused by the epidemic, Alibaba's DingTalk has introduced a series of features based on the actual scenarios of different enterprises, such as customized daily health check-in, 100-person high-definition video conference, release of emergency notice and other functions. Based on the demands for resumption of schools, it also provides group live broadcast and other tailored services. Thunisoft provides integrated solutions for legal business scenarios, including cloud video services and legal-related video services for courts, procuratorates, political and legal affairs commissions, judicial and administrative agencies. It includes online court trials, video-based judicial and administrative mediation, remote video interviews by the procuratorate and other features.

(4) Big Data Guarantees an Accurate Increase in Production and Capacity of Enterprises

State Grid can analyze and predict the resumption of work and production in cities through electric power big data. Taking State Grid Zhejiang Electric Power as an example, it dynamically monitors and accurately analyzes the resumption of work and production in various regions and industries based on

the historical and current power consumption data of the enterprises in the power consumption information collection system. The electric power big data empowers the increase in production and capacity. It can quickly determine power supply schemes, and provide valuable data support and decision-making references for government departments to guide EPC and resumption of work and production. Baidu's intelligent quality inspection system helps a factory in Changzhou resume work safely. The 24X7 unattended intelligent inspection equipment is nearly 10 times more efficient than manual inspection. This system has effectively solved the problems that most workers cannot return to the factories during the epidemic period, and reduced the risk of epidemic infection in the workshops.

III. Deep Insights

1. Data Capacity Is the Basis for Epidemic Prevention and Control.

Good and abundant data are the basis for developing applications for EPC. This section analyzes the data capabilities reflected in the relevant cases collected in this report from three aspects: data collection, data interoperability, and open data.

(1) Data Collection

Traditional data collection approaches have exposed some weakness in the face of large amounts of data. The traditional collection of health data usually starts from the community health center at the grassroots level. The information is manually filled in by the community health workers, goes through district health committees and then municipal health committees, and finally comes to the provincial health committee and the national health committee. On the one hand, it increases the burden on the data collection staff at the grassroots level and reduces the efficiency of data collection. On the other hand, it is difficult to quickly verify the correctness of the data from the source, which increases the cost of data quality management at the subsequent stages.

With the help of informatization, many places have adopted online information reporting systems and intelligent outbound call platforms instead of manual reporting to reduce the labor cost of data collection and shorten the data flow journey. Beijing launched an online reporting service on February 13. Residents can register personal information on Alipay. So long as the residents stay in the same community, they do not need to repeatedly report personal information, but they need to make a daily health check-in[2]. At the same time, Baidu, Ali, Beijing Laiye Network Technology and some other enterprises have launched intelligent outbound call platforms for EPC. After automatically asking residents about the situation of leaving and returning to Beijing and their physical conditions, the system can automatically file the

[2] https://new.qq.com/omn/20200213/20200213A0K1Z200.html

information based on keywords and quickly realize the collection of epidemic data of residents in their respective jurisdictions. In addition, enterprises and individuals also use crawler technologies to collect dynamic data and quickly provide services such as epidemic data visualization, epidemic situation analysis, and epidemic extrapolation. The main users, the advantages and disadvantages of different data collection technologies are shown in the table below.

Table 1 - Comparison of Data Collection Technologies

Data collection method	Main Users	Advantages	Disadvantages
Paper form	Governments	Comprehensive and timely collection of data at the beginning of the epidemic and when there is only a small amount of data	● Inefficient data collection ● Cumbersome data collection process ● Poor storage availability ● Data are difficult to verify
Online Reporting System	Governments, enterprises	● Fast acquisition of large amount of data ● Easy to store and process data ● High reliability of collected data	Data update is not timely
Provided by Existing Information Systems	Government, enterprises	● Fast data acquisition ● Real-time data ● High authenticity of data	Complex interfacing process
Smart Outbound Calls	Governments	Fast acquisition of large amount of data	● Difficult to collect data comprehensively ● Inaccurate data ● High development costs
Crawler	Enterprises, individuals	● Fast acquisition of large amount of data in real time ● Capture more relevant data	● High technical threshold ● Risks in data reliability ● Non-public data are not available

In order to ensure the comprehensiveness and accuracy of data collection, automated collection tools and data quality verification means should be used. For multi-source data, the main source should be identified. The data should be captured from the corresponding system or platform to reduce self-reported data and repeated collection. At the same time, big data technologies should be used to achieve fast and real-time collection of various types of epidemic information. By providing multiple collection methods such as ETL, API, message queue, and data flow, as well as mobile reporting data collection functions and ensure real-time and accurate access to information including population movement information, real-time traffic information,

location information of key prevention and control personnel, logistics information, online public opinion and other types of data so as to meet the needs for data resources during this special period.

(2) Open Data

In response to the public's urgent need for epidemic data since the outbreak of Covid-19, many local government departments have made the epidemic-related data publicly available in a timely manner. This has not only helped eliminate public panic, increased self-protection awareness, supported government prevention and control measures, and enhanced government credibility, but also provided a steady supply of "fuel" for enterprises, professional organizations and individuals for the analysis.

At present, various provinces and cities have made a variety of epidemic data available to the public, including statistical epidemic data and data of medical institutions. This has provided data users with an effective way to capture the epidemic situation and search medical information. The statistical epidemic information of each province and city basically covers data entries such as date, location, time period, newly confirmed cases, cumulative reported confirmed cases, suspected cases, and newly discharged cases. At the same time, some provinces and cities have categorized statistical epidemic information to facilitate users to find data. For example, Shenzhen made available 9 data sets and divided the epidemic information into statistics of sources of daily confirmed cases, statistics of daily confirmed case, and daily diagnosis and treatment.

Currently, many provinces and cities are making structured data publicly available and establishing data standards and metadata to improve the availability and ease of use of epidemic data. For example, Shandong Province has established a data standard for the data set called "Information of Covid-19 Confirmed Cases in Shandong", which has specified the English name, data format, and Chinese description of each data entry; Shenzhen has established metadata for each data set, including key word, release time, update date, resource format, data provider, and data maintainer, to achieve fast data retrieval.

At the same time, various provinces and cities are basically able to disclose the latest epidemic situation every morning, and the epidemic information is updated within 24 hours. Multiple data download methods are provided to ensure the timeliness and availability of data. The health commissions of the municipalities directly under the central government have released epidemic data on the official website from January 20 or January 21, 2020. The epidemic information is released twice a day on average in Beijing and Tianjin. Different provinces and cities have adjusted the time and frequency of information release based on the progress of the epidemic. Beijing, Shanghai and other regions have provided API calls and supported the download of mainstream formats such as xls, xml, json, and csv.

However, the insufficient open data in confirmed cases and the lack of open geospatial data in various provinces and cities have greatly limited the ability of data users to further analyze and explore epidemic data. At the same time, the epidemic data released by some governments are fragmented and disclosed at different times, located on different pages, and appearing under different names[3], increasing the difficulty for data users to extract and organize data.

Medical research institutions have released the latest Covid-19 research results in a timely manner. On February 28, the research paper written by the team led by Dr. Zhong Nanshan, Academician of the Chinese Academy of Engineering, was published online in a top international medical journal, the New England Journal of Medicine (NEJM). The researchers have analyzed the distribution characteristics of patients at different age groups, symptoms of infected persons, contact history, imaging manifestation, treatment methods, clinical outcomes (including case fatality rate) and other aspects to help the public understand the diagnosis and treatment of Covid-19 and the details of the development of the epidemic[4]. As tech companies have a large amount of high-precision user behavior data, they use big data technologies to continuously release analysis reports related to resumption of work and production. On February 24, AutoNavi released the Analysis Report on Work Resumption Index based on Driving Activities. The number of people who

[3] "Which regions have made truly 'usable' and 'easy to use' data available? What can we do with these data? "
[4] https://m.huxiu.com/moment/86665.html?f=app_24h_ios_friends

actually used the navigation service was used as a data source to reflect the heat of work resumption[5] in 20 major cities from the perspective of navigation. The "Covid-19 Search Big Data Report – Resumption of Work" released by Baidu provides the public with a reliable reference[6] in terms of understanding the current status of various industries, the true needs of employees and enterprises, and industry development opportunities based on the popularity of keywords searched on "Baidu".

(3) Data Interoperability

With data interoperability, key epidemic information can be consolidated to effectively support the government's epidemic research and judgment and resumption of work and production. Hangzhou was the first to launch health codes nationwide, using red, yellow, and green QR codes as digital health certificates. Citizens or returning workers can apply the health codes online. The information for submission includes the current location, travel, and contact with patients, and will be reviewed by the government's backend system. Health code can be used as a certificate to travel around. The health code is an outcome of the interworking of data. Each region can scientifically evaluate the proportion of resumption of work and production based on the dynamic change of health codes and dynamically adjust the strategy of resumption of labor and production[7].

In the EPC work, data interoperability has been promoted between some governments and between government departments. This has greatly simplified the data collection process and improved the efficiency of EPC. The “close contact meter”, which was jointly developed by CETC, the e-government office of the General Office of the State Council, and the National Health Commission, went live on February 8 and obtained the data resources from the National Health Commission, the Ministry of Transport, China Railway and CAAC. The authority of the data is guaranteed from the source. The general public can check whether they are close contacts of patients with Covid-19 by

[5] https://www.xuexi.cn/lgpage/detail/index.html?id=12469248233056609514&_
[6] http://cn.chinadaily.com.cn/a/202002/14/WS5e466b18a3107bb6b579fa63.html
[7] https://baijiahao.baidu.com/s?id=1659038766757094374&wfr=spider&for=pc

entering personal information[8]. MIIT has coordinated the three basic telecom operators to build up a telecom big data platform in support of EPC and break down the barriers to the interoperability of their signaling data. At the same time, data sharing between telecom big data and health and disease control management departments was implemented. Based on the data of health and epidemic prevention departments, further research and judgment were made. Ji'nan Municipal Health Committee has developed a system to analyze the people visiting the fever clinics, bringing together the data of 86 major hospitals and more than 3,000 primary health institutions in the city. Combined with health data and data, the comprehensive tracking and continuous analysis of priority groups are achieved[9].

Some service-oriented enterprises have broken down data silos of enterprise and personal data within their platforms, and data exchange between enterprises is gradually realized. Umetrip has gathered the flight data of various airlines. On January 30, it launched a new push notification system that sends an alert to air travelers if any of their fellow passengers is confirmed with Covid-19. The records of flights with confirmed cases will be updated and an alert will be will sent to all the passengers on board the same flight[10]. China Railway, which owns the big data of its ticket sales platform – 12306, worked with local governments and health departments to identify close contacts of confirmed cases on trains and arrange the transport capacity for peak-shifting returns[11]. China Tower, in collaboration with 360 Total Security, introduced "Intelligent EPC Platform". Combined with train and flight data and billing data, the platform can sort out travel records of the cases with time stamps and close contacts and replace traditional epidemics investigation method[12], which is shown in Figure 7.

[8] http://www.sastind.gov.cn/n112/n117/c6808856/content.html
[9] https://baijiahao.baidu.com/s?id=1659111197539074743&wfr=spider&for=pc
[10] https://baijiahao.baidu.com/s?id=1657145992333744288&wfr=spider&for=pc
[11] https://www.xuexi.cn/lgpage/detail/index.html?id=8145762554441060628&item_id=8145762554441060628
[12] https://mp.weixin.qq.com/s/fMfEc3N69GhBQwFPVzafKw

Figure 7 - China Tower Intelligent EPC Platform

However, data interoperability still faces many obstacles. For example, an effective data interoperability mechanism has not been established, and the requirements and scope of data interoperability cannot be clarified in a short period of time. Due to the barriers in interfaces and rules of related technologies, it is also difficult to support the rapid connectivity of large volume of data.

Therefore, local governments should clarify the requirements for data interoperability, establish resource matching and distribution mechanism, determine the resources necessary to meet the needs, integrate the data capabilities of existing IT systems, and eliminate technical barriers to data interoperability in the future. For example, MIIT has immediately established a leadership group for EPC with a focus on the telecom big data support service to coordinate related work. MIIT organized provincial and municipal communications administrations, CAICT and basic telecom operators to establish a joint EPC working mechanism of the communication industry to strengthen joint coordination of the communication system between province and ministry. At the same time, under the State Council's epidemic joint prevention and control mechanism, a joint working mechanism was

established with the National Health Commission for the sharing of epidemic-rated telecom big data. The information is shared with relevant departments regularly or on demand, which has effectively promoted information sharing between government departments.

In addition, data cooperation between governments and enterprises should be strengthened so that the original fragmented data in the fields of public travel, social media, telecom services, and other areas, which are related to EPC, can be deeply integrated and effectively used. As the EPC situation gets tough after the resumption of work and production, the data should be used to reduce second-generation infections and block third-generation infections. At the same time, a large amount of epidemic-related data owned by social media and civil organizations, should be fully utilized. Big data and AI, based on these data and combined with other sentinel surveillance data, and traditional public health data resources, can help us analyze and identify the causes.

2. Data Analysis Gets Deeper.

The depth of data analysis applications is generally assessed in terms of its operational difficulty and output value, and is generally divided into four types in the industry:

Descriptive analysis: This type of analysis only describes what happened. The information is usually conveyed and presented in a visualized way, such as the epidemic distribution information displayed on big screens;

Diagnostic analysis: This type of analysis traces the cause of an event. Generally, the cause of the problem is further analyzed based on descriptive information, such as the attribution of the virus infection;

Predictive analysis: This type of analysis predicts possible events. It needs to model variable data and predict the possibility of events through prediction models, such as predicting the development trend of the epidemic situation;

Prescriptive analysis: This type of analysis can guide the next-step decision. It is more complex and valuable. It usually relies on a combination of descriptive, diagnostic and predictive, and other analysis behaviors to determine the best solutions. For example, the local government can take the most appropriate actions based on the comprehensive understanding of local conditions, the surrounding epidemic situation and other factors.

During the anti-epidemic period, the data applications mainly showed the following characteristics:

The early stage of epidemic control is a critical period for laying a solid data foundation. Most of the applications emerging at this stage are mostly information collection and platform construction projects, such as the construction of various self-reporting systems and big data platforms. Due to time and data restrictions, more than 60% of the data applications before February 2020 were preliminary descriptive applications, such as epidemic distribution maps, and population mobility displays. There were a few diagnostic applications, mostly related to attribution analysis of virus infections, and judgment of high-risk groups based on the spatiotemporal attributes of mobility of people. For example, at the beginning of the epidemic, after a passenger was confirmed with the disease, the aviation and railway departments would send out an alert to the passengers on the same flight and train, who required special attention and quarantine.

With the construction of the platforms and the gradual accumulation of data, diagnostic applications have become more mature, and predictive applications have gradually increased. By the beginning of February 2020, the proportion of descriptive and diagnostic applications reached 75%, and the data and analysis were more granular. District-level and even community-level applications began to appear. For example, the “one district, one code” system launched by Beijing Percent IT. Residents complete the filling of personal information and the system automatically analyzes and assists manual management when the residents enter and leave the community. Predictive applications, such as the prediction of potential source of infection and the path of infection, and the trend of the infected, have appeared and are gradually used as reference for government decision-making. Based on

high-dimensional machine learning technologies, 4Paradigm established a city-district and county-level digital twin system, which can simulate the impact of key decisions by considering various factors that impact the epidemic such as traffic control, time for resumption of work, and drug delivery, thus providing the basis for policy development.

Predictive and prescriptive applications will play an important role in the epidemic control. Predictive and prescriptive applications are highly complex. Although such applications are fewer, they have higher value. During the EPC process, big data and AI technologies are used in the detection and analysis of viral genes. Based on the characteristics of viral genes, predicting viral protein structure can assist drug screening, which can greatly improve the efficiency of diagnosis and epidemic research. Baidu Research has open-sourced LinearFold to scientific research institutions and opened the fastest RNA structure prediction website for free. It can greatly improve the efficiency of scientific research and help epidemics prevention and control.

The maturity of various descriptive, diagnostic, and predictive applications has laid the foundation for the government to develop targeted policies and take targeted actions. The road, railway, and flight traffic analysis provided by the transportation department and the trajectory analysis of close contacts provided by the communication operators can effectively support the governments to take the targeted prevention and control measures. Materials support platforms, "health code" data, electricity consumption data of enterprises, and other analytics applications can also support the government to take special response measures in various links of production to guarantee the resumption of work and production. With the further opening of the data in each link and the continuous accumulation of experience, more predictive and prescriptive analysis applications will play a role in EPC.

3. Tech Companies Show the Power of Technologies in EPC

(1) The Value of Big Data Becomes Self-evident.

The "human-to-human" transmission of the disease has placed the monitoring and management of individuals at the center. How to access, describe and

analyze an individual's trajectory, and the overlapping of positions of multiple people has become the most basic and important data for analytics. Therefore, both the mobile signaling data and the location data obtained by the Internet Apps have become important data resources. Many telecom and Internet companies have also developed corresponding products based on such data. Specifically, due to the mobile phone real-name registration reform, the comprehensiveness, authenticity, and real-time nature of telecom big data has fully demonstrated its value.

Telecom big data come from the basic communication data in the course of telecom operations. Using telecom big data for the analysis and statistics of population flow is of great significance for supporting the analysis and judgment on the epidemic situation, making EPC arrangements, monitoring mobile personnel and developing targeted policies. Telecom data has the following characteristics: First, comprehensiveness. China has 1.6 billion mobile phone users. It is characterized by huge user base, wide coverage, and large amounts of data. For example, the existing public communication network generates about hundreds of billions of telecom data records every day. During the holidays such as the Spring Festival, people are highly mobile and the amount of data will be larger. Second, authenticity. After years of real-name authentication, telecom big data systems have basically realized "one number corresponding to one person". The reliability and authenticity of locating the person and confirming the identity of the person have been greatly improved through the base station signaling data, laying a foundation of authentic data for the analysis of epidemic. Third, real-time. Telecom big data systems can collect, summarize and process telecom-related data in real time, provide various data analysis results in a timely manner, and provide fine-granularity data support for EPC. For the dynamic movement of people, it can analyze and predict the dynamic movement of key populations such as the diagnosed, suspected and close contacts to support the arrangements for EPC.

As the largest mobile operator in Beijing, Beijing Mobile has a large amount of authentic user data. On the premise of complying with national laws and industry norms, Beijing Mobile Big Data Center launched an integrated solution for EPC. Its epidemic prevention platform supports the collection and

real-time processing of signaling data in a centralized manner. Through the fusion of local user data and roaming data of other provinces and regions, it can help related departments better understand the flow of affected personnel across the country and make reductive analysis on the complete trajectory chain. Due to strong correlation between the real-name registration and natural person, the timeliness of epidemic prevention data has been greatly improved. Through the continuous acquisition of mobile phone data, a deep connection between natural person and geospatial and behavioral characteristics is established, daily insights into the flow of people in key areas are obtained, and detailed comparison and analysis of data are conducted on a daily basis.

(2) Fast and Efficient Response

The Covid-19 epidemic came aggressively. In the face of such an emergency, the rapid response capabilities of tech companies are a reflection of their strength. From system bus to service-oriented thinking, from modularization and loose coupling to Zhong Tai and micro-services, technological innovations always develop towards an efficient and agile direction. Such innovations are also the most important means to solve the needs of rapid response.

Whether Baidu Map has upgraded its migration big data platform from 100 cities to 300 cities with almost no delay, or the Yida platform of Alibaba Cloud has turned invisible microservice modules in the backend into the drag and drop operations in the front-end, or telecom operators have rapidly formed epidemic big data solutions based on their long-term data and service accumulation, it is a reflection of their rapid response capabilities.

(3) Capabilities of Intelligent Technologies Continue to Improve.

Big data and AI technologies complement each other and can jointly improve the level of intelligent technologies. Intelligent methods and equipment are constantly developing, and assisting humans to better complete more tasks has become an important direction for future development. During the epidemic, technologies such as AI intelligent temperature measurement equipment, intelligent robots, intelligent unmanned vehicles, intelligent

algorithm models, smart applications, and intelligent displays. have helped us accomplish more things that humans cannot.

The increasing demand for intelligence and the continuous innovation of intelligent technologies have also led to the development of smart economy. After experiencing the three stages of the invention and popularization of the PC, the PC Internet, and the mobile Internet, the digital economy is evolving into a new stage of smart economy with AI as its core driving force. The smart economy will bring new vitality to the global economy and is the core engine that drives the global economy forward[13].

Looking at the epidemic up to date, many tech companies have fully unlocked the potentials of AI technologies to fight the epidemic in terms of epidemic analysis and prevention, medical treatment, and resumption of work and production. For instance, AI temperature measurement system, Baidu Map's migration big data platform and other applications are impressive. The applications of related AI technologies may find suitable commercial landing scenarios after the end of epidemic. This shows the role of intelligent development in promoting social and economic development.

4. "Open Source and Crowdsourcing" Becomes a Unique Landscape

In the early days when the Covid-19 began to spread, information about the epidemic exploded, and there was a diversity of release channels. The information of medical supplies shortage was not effectively collected and released, which has brought great difficulties to material distribution and donation. At this time, a group of volunteers formed a R&D team. Based on their own professional advantages, they used crowdsourcing and collaboration to initiate open source IT application projects related to epidemic prevention and aimed to build reliable and efficient release channels. it has created a unique landscape in the tech communities for EPC.

(1) Rich Project Types

[13] "Smart Economy Brings New Revolution, Industry Intelligence Needs to Go Through Three Stages", Robin Li, Wuzhen World Internet Conference

According to the 46 open source projects counted by the open source platform OpenSourceWuhan, the epidemic information category accounted for 37%, the news recording category reached 30%, followed by the help information category 9%. Specifically, more than half of the epidemic information category are real-time epidemic maps of provinces and cities. Among the 14 news-recording projects, there is <2020 Covid-19 Memory>, which records the media, non-fiction and individual narratives, and also <2020nCov_individual_archives>, which records individual's life.

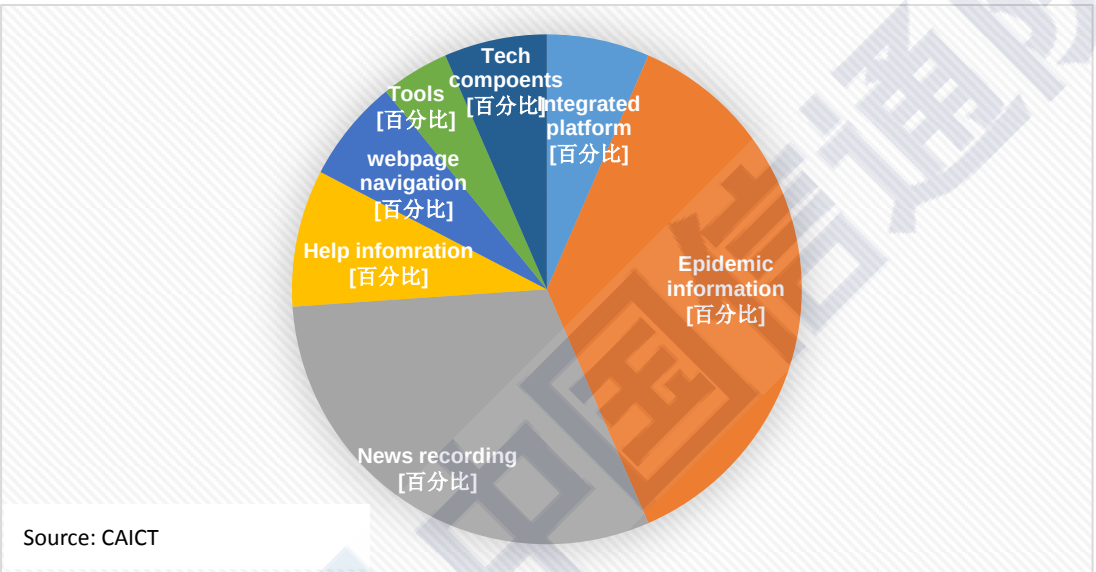


Figure 8 - Statistics of Open Source Projects

(2) Fast and Efficient Operation

A major public health event is of a more complex type among public social events. It involves many types of resources to be mobilized, it is difficult to coordinate and arrange different resources, and it takes long time for deployment and implementation. The open source and crowdsourcing models allow supply and demand to find each other, thus improving efficiency significantly. Most open source projects for EPC can get their systems launched in two to three days, and the number of people involved in the projects range from a few to thousands. Take the wuhan2020 project as an example. The project was initiated on January 25. It attracted the participation of hundreds of people in two or three days, and got the system launched on Jan. 28, then the system was iteratively upgraded. According to statistics, there are

currently more than 1,000 volunteers participating, and 39 contributors to the project on Github. The other star project, "2020 Covid-19 Memory: Reports, Non-Fiction and Individual Narratives", uses Github to store various reports. It stored the reports from more than 70 media and attracted more than 100 participants. 2020 Wuhan Individual Stories included more than 300 logs in a short time. In contrast, the normal outsourced informatization project usually needs about one week to launch.

(3) Two Channels to Access Data

The data of open source EPC projects generally come from two channels. One is publicly available data, which is crawled manually or by machine, and then processed. For example, the wuhan2020 project uses a large number of volunteers to manually crawl various data on the Internet. The data will be recorded on Github after being reviewed by the reviewers; and the project DXY-COVID-19-Crawler is a tool for crawling data. It crawls data on the web in real time and exposes API interfaces to other projects; most of other epidemic map projects use crawlers to crawl the epidemic data published by various provinces and cities. Another is from the user-generated information and data, such as 2020 Wuhan Individual Stories, which calls on ordinary people in Wuhan to contribute their diaries as data sources.

(4) Diversified Modes of Operation

Projects involving multi-person collaboration require a set of processes for division of labor, including collaboration specifications, division of labor, and tools for communication and collaboration. It mainly includes code hosting and collaboration, communication between developers, data and information collection, synchronization, and review, as well as daily communication and formal meetings between volunteers. Most open source projects for EPC use Github, Gitee and other code hosting platforms for development and collaboration. Github is a frequently used platform. Slack, WeChat, DingTalk and other tools are used by volunteers as daily communication tools. Many open source projects use collaborative tools such as graphite documentation to collect data and text, as well as some image processing tools for design. At the same time, open sourcing has also become an important mode of

operation for tech companies. Baidu PaddlePaddle is the first open source mask-wearer face detection and classification model in the industry. This model can effectively detect all faces with and without masks in crowded areas and determine if the person is wearing a mask. Taking advantage of China ’s largest AI open ecosystem, Baidu has launched an AI developer anti-epidemic guardian plan to mobilize the strength of developers and partners to develop various types of EPC products, and realize the full output of intelligent capabilities.

(5) Rising Influence

Open source projects have achieved huge social effects in the EPC process. This report makes statistics of the top 7 open source EPC projects in terms of the number of stars on Github and Gitee. As shown in Table 2, these projects include comprehensive information display, news recording, information navigation, epidemic maps, information crawlers and other aspects. As of February 21, two projects had more than 6,000 stars on Github, and more than 800 forks, both of which topped Github Trending.

Table 2 - Functions and Influence of Open Source Projects

Item Name	Category	Functions	Number of Stars	Number of Forks
2019ncovmemory/nCovMemory	News Recording	2020 Covid-19 Memory: Reports, Non-Fiction and Individual Narratives	6600	880
wuhan2020 project	Integrated Information Platform	To collect information from various hospitals, hotels, factories, logistics, materials donation, financial donation, epidemic prevention, treatment, and updates, and release it in a unified manner so that the parties can communicate with each other and effectively allocate social resources.	6000	870
DXY-COVID-19-Crawler	Information Crawler	Covid-2019 epidemic data real-time crawler and API	1000	207
wuhan2020-timeline	News recording	Record the timeline of the progress of Covid-19 in Wuhan since December 2019.	890	150
2019-ncov	Epidemic Map	Visualization	390	81
2020nCov_individual_archives	News recording	Back up ordinary people's records during the epidemic.	340	45

xinguan2020	Integrated Information Platform	Visitor registration system, disinfection inspection & registration system, student health report management system, disciplinary violation reporting system, material management system, hospital patient management system	206	59

In summary, open source projects have shown unique advantages and characteristics in the EPC. The operation and management of open source projects makes large-scale social collaboration possible, allowing ordinary people and people with professional skills to participate in the EPC work. Open source projects can quickly organize effective forces in a short period of time to build various information platforms, and their response speed and iteration efficiency are higher than conventional outsourced projects. The smooth operation of open source projects requires a set of collaborative processes and incentive mechanisms, which usually involve division of labor, collaboration specifications, and use of tools. Their democratic coordination methods and decision-making mechanisms can stimulate the enthusiasm of participants to take part. From the perspective of collaboration tools, they are mainly divided into code hosting, data and information collection and aggregation, and daily communication tools. These collaboration tools have effectively improved the efficiency of distributed work and provided a strong guarantee for remote online collaboration. From the perspective of data and information sources, the main sources of open source projects depend on publicly available information on the Internet and voluntary contributions of volunteers. Generally, data are collected manually or by machines. In the face of manually collected and contributed information, an audit mechanism is needed to verify the validity and authenticity of the information. Based on the observation of the actual project operation, these open source projects had huge influence and played significant roles. A large number of applications have emerged in the display and visualization of epidemic information, the release of hospital needs for materials, recording media reports, and recording personal life. The information and code of open source projects are open to the society in accordance with certain protocols, which greatly improves the utilization rate.

IV. Problems and Challenges

Although big data and intelligent technologies can play a significant role in the EPC process. However, from the perspective of the response, especially based on the role played by Internet companies in the data-driven EPC process, there is still big room for further improvement.

1. Limited Data Sources and Backward Collection Methods

From the cases of Internet companies, it is not difficult to see that the data used by most enterprises in the EPC basically derive from data disclosed by governments. However, at the current stage, the dimensions of government public data are not rich enough to meet the needs of the public.

In the era of big data, the needs of the public for information have changed. In the face of the epidemic, the public focuses their attention on not only on information about prevention and control work, self-protection knowledge, and related policy documents, but also on specific and quantified epidemic-related information and data[14].

For example, in addition to the statistics of the overall situation of a place (how many suspected, diagnosed, critical, discharged, or death cases have been accumulated in various provinces, cities, districts, and counties today? How many new or reduced cases today? How many people are excluded from suspected cases? How many close contacts were released from observation?), the public also want to know the data of the individual patients (where did these patients come from? Where did they go? When did they get sick? Which hospitals did they go to? What are the symptoms? What is the current situation?). However, only based on the information disclosed by the government and data provided by enterprise cases, most places are still unable to meet these requirements.

In addition, some government departments and grass-roots organizations still use manual methods to collect epidemic-related data. When collecting data,

[14] "Which municipality has disclosed the most complete, detailed and standardized epidemic data? We made a comparative analysis." Zheng Lei and Chen Yao.

they usually use paper forms and telephone inquiries, which not only increases the burden on the grassroots staff, but also cannot guarantee the authenticity of the data. For example, when collecting personal information of patients, if the mobile phone number left is not distinguished as the patient's number or the number of the patient's family, it may lead to a large number of inconsistent results between the ID number and the mobile phone number when verified on the operator's system. This will add difficulties to the accuracy and authenticity of subsequent data analysis.

2. Lack of Data Governance and Low Data Quality

Judging from the data used for the epidemic analysis, the current data obtained through public channels are not standardized, the statistical coverage, data statistics time and other dimensions are not unified, which brings great difficulties to the analysis. Data sources are mixed and duplicated, data quality is not high, and data credibility needs to be improved.

In the process of EPC analysis, huge amounts of data are being produced every moment. An inadvertently misregistered ID number, an undocumented home address, hiding information about one trip, and an unknown close contact may trigger a new epidemic. However, the lack of attention to the preprocessing stage of big data and the unstandardized data processing are likely to increase the difficulty to access authentic data. Judging from the existing data reporting forms and the quality of reported data, there are still many unstandardized and unreasonable situations in the data reporting process.

The above reasons have resulted in poor data usability, poor data quality, and inaccurate data. The significance of using big data for prevention and control analysis is not only to collect large amounts of data and information, but also to perform good preprocessing of collected data so that data analysis and data mining personnel can extract valuable information from the usable big data, which should the ultimate goal of big data analysis for EPC.

3. Insufficient Privacy Protection and Compliance Risks

With the escalation of the Covid-19 EPC situation, various public awareness-raising and prevention methods have been adopted in various places. Among them, the screening and reporting of returnees from key affected areas and confirmed patients is a very important measure, which can help health supervision institutions understand the situation in a timely manner and quickly cut off the transmission linkage. Although the relevant departments have hidden the personal privacy information of the patient when producing the co-passenger query tool, the data were still leaked to varying degrees in the process of reporting the primary data. Not long ago, many documents containing personal information were suddenly reposted on social networking platforms such as WeChat and Weibo, including the names, photos, employer units, schools, home addresses, mobile phone numbers, and ID numbers of relevant personnel. This brought great difficulties to the lives of returnees and confirmed patients. Many people received harassing calls and abusive text messages. It also exposes the weak protection of personal data.

Articles 110 and 111 of the General Principles of the Civil Law of the People's Republic of China stipulate that citizens enjoy the right to privacy and personal information is protected by law. No organization or individual shall illegally collect, use, process, or transmit the personal information of others, or illegally trade, provide or disclose the personal information of others. Article 12 of the Law of the People's Republic of China on Prevention and Treatment of Infectious Diseases also stipulates that the disease prevention and control institutions and medical institutions shall not disclose relevant information and materials related to personal privacy.

In the special period of EPC, the public's right to know, public safety and citizens' right to privacy need to be considered in a balanced manner. Based on the actual needs of protecting public interest in EPC, specific information of special groups can be disclosed, but the principle of compliance should be followed for the disclosure, the disclosure is really based on necessary ground and meet the purpose of EPC, and the disclosure means should be reasonable and appropriate. Negative effects should be minimized.

In order to protect the personal information in the joint prevention and control of the Covid-19 epidemic, we should actively use big data including personal information to support the joint prevention and control work. The Cybersecurity Administration of China (CAC) issued the "Notice on Doing a Good Job in Protecting Personal Information and Using Big Data to Support Joint Prevention and Control"[15] on February 4, 2020, which clearly states that personal information collected for EPC and disease prevention and treatment purposes shall not be used for other purposes. No unit or individual may disclose personal information such as name, age, ID number, etc. without the consent of the person whose data are collected.

The notice requires that all regions and departments must attach great importance to the protection of personal information. Except for agencies authorized by the health department of the State Council in accordance with the Cybersecurity Law of the People's Republic of China, the Law of the People's Republic of China on the Prevention and Treatment of Infectious Diseases, and the Regulations on Response to Public Health Emergencies, no other unit or individual may, on the grounds of EPC and disease prevention and treatment, collect and use personal information without the consent of the person whose data are collected. Where laws and administrative regulations provide otherwise, such provisions shall prevail.

The notice also requires that the collection of necessary personal information should refer to China's national standard in this field – the Personal Information Security Specification and adhere to the principle of data minimization. The collection should be limited to key groups such as diagnosed patients, suspects, and close contacts in principle, and is generally not targeted at all population in specific regions. The de facto discrimination against specific geographical groups should be prevented.

The notice clearly states that personal information collected for EPC and disease prevention and treatment purposes shall not be used for other purposes. No unit or individual may disclose personal information such as name, age, ID number, phone number, and home address without the consent

[15] http://www.cac.gov.cn/2020-02/09/c_1582791585580220.htm

of the person whose data are collected, except for the disclosure of masked data out of the needs for joint prevention and control.

4. Insufficient Data Circulation and Serious Data Silos

The core of big data is connectivity. Only by integrating multiple sources of data, can the value of big data be unleashed. In EPC, we have also seen the problems of fragmented and scattered data, consolidated but unconnected data and connected but unused data. This is largely due to the mismatch between the data source and the data development capabilities, that is, the party that holds the data lacks sufficient data development and utilization capabilities, while the party with the data development capabilities does not have enough data. As a result, the role of data cannot be maximized due to insufficient circulation.

Moreover, the data of different enterprises and government departments are isolated in the form of "silos", and the data needs to be efficiently integrated. Data fragmentation is one of the challenges faced by local government departments at all levels and major Internet companies when using big data to prevent epidemics. Judging from the cases of multiple Internet companies, the data are often scattered in different departments, and these data are stored in different data warehouses. The data technologies of different departments may also be different, which leads to the isolation of internal data in different silos within a region or with an enterprise. If these barriers are not removed, it would be difficult to unleash the value of big data. Big data requires the association and integration of different data to better play its advantages. How to break through the barriers of data silos, and realize the sharing of technologies and tools is the key to unleash the value of big data in epidemic prevention.

V. Summary and Suggestions

General Secretary Xi Jinping pointed out that the outbreak is a big test of China's governance system and capabilities. EPC has presented not only a major challenge to governance, but also an important opportunity to optimize the governance system and enhance governance capabilities. In order to strengthen the role of next-generation ITs such as big data in EPC, the following suggestions are made in this report.

1. Enhance Data Accumulation and Maximize the Value of Data Applications

Why did some regions and departments encounter such challenges as insufficient data sources and difficulties in collecting data during the EPC? It can be said that this is closely related to the failure to collect and accumulate basic data in the work of relevant departments. For example, if a medical institution has corresponding real-time data statistics on the patient's conditions, it can issue early warning and assist in judgment of the epidemic when data abnormalities such as the surge in patients with similar conditions occur. The departments at the grassroot level and community organizations should also strengthen the check of community residents, build a comprehensive big data support mechanism, have the basic data capabilities to respond to emergencies, and avoid multiple and repeated data collection.

2. Improve Data Quality and Strengthen Data Governance Capabilities

During the epidemic, the format of the epidemic-related data disclosed by various regions were not unified, and even conflicting conversions and re-corrections occurred, creating great difficulties to the analysis of epidemic-related data. Therefore, harmonizing data standards is critical. It is necessary to clearly define the source of the data, the format of the data, and the requirements for data management. At the same time, it is necessary to standardize not only the data source, but also the data processing process. On the premise of ensuring data authenticity, we should make full use of the existing big data platforms and big data development results and consolidate and integrate data in timely and effective manner.

3. Protect Privacy and Ensure Data Compliance

During the EPC process, while protecting the public's right to know, personal information must be protected. First, to clearly define the implementing agency that collects and uses relevant personal information, as well as the relevant work processes and specifications, and ensure that the collected information is used only for EPC and disease prevention and treatment purposes. Second, to clarify the sharing scope and use rights of each department and related units in the process of data sharing, and measures such as anonymization, de-identification, and access control should be adopted to restrict data sharing behaviors. Third, to disclose the relevant masked information in a proper manner on the premise of protecting the personal information of the parties involved when disclosing epidemic information to the society. Fourth, to strengthen the confidentiality awareness of the information collection staff at the grassroot level, control data compliance from the source, and prevent the leakage of personal information.

4. Promote Technologies and Explore Innovative Application Scenarios

Epidemic prevention and control (EPC) can provide more scenarios for the application of various digital technologies. For example, robots can be used to achieve intelligent sterilization, primary screening of fever, isolation ward rounds, drug distribution and other functions, and intelligent sensing detection technologies can be used to achieve accurate and efficient measurement of body temperature and other functions. At present, there are not enough enterprise cases in this regard. Therefore, it is necessary to encourage all parties to use the innovative ICT applications more efficiently to support joint EPC. First, local governments should actively use 5G, cloud computing, big data and other technologies to further improve EPC. They should establish big data monitoring and analysis platforms to analyze the trajectory of confirmed patients, track their contact history, identify the virus transmission route, and predict the development trend of the epidemic situation. Second, to encourage medical institutions and medical scientific research institutes to apply AI technologies, optimize algorithms and computing power, and help R&D of online diagnosis, viral genome sequencing, and vaccines/drugs. Third, to encourage cloud computing, big data, AI and

other related enterprises to strengthen joint research, develop innovative applications in epidemic detection, analysis, early warning, prevention and control, and improve the effectiveness of epidemic prevention and treatment.

5. Seek Opportunities and Promote the Development of Digital Economy

The epidemic presents both challenges and opportunities. In this epidemic, whether with the rise of "stay-at-home economy" triggered by the surge of traffic from fresh food delivery, e-commerce, online entertainment, and education platforms, or with the resumption of work and production, Internet and IT enterprises and industrial Internet platforms have begun to show their power in helping enterprises to restore their "horsepower" as quickly as possible, all highlighting the importance of the digital economy. The rapid development of the digital economy will help ensure the stable and orderly progress of current living and production, and reduce the negative impact of the epidemic on China's real economy.

New models and new business forms have seen the most active investment in the digital economy. During the epidemic, home quarantine led to the development of new models and new business forms such as online entertainment, online education, online office, and telemedicine, and changed the mode of social operation and people's behaviors. As the epidemic eases, economic and social production returns to normal, and huge online market demands such as education, medical care, and office work will attract a large influx of social capital, which is expected to spawn super apps like Baidu App, WeChat, Didi, etc.

The epidemic situation has delayed the resumption of work and production for a large number of enterprises. Facing recruitment challenge and expensive labor, many manufacturing enterprises have to take measures to solve problems by introducing production equipment, reforming employment mode, improving management efficiency, and improving workers' skills. For example, Supor responded to the difficulty in recruitment by upgrading automation levels, building up human-machine mixed-line work environment and reducing labor costs. The epidemic may become a booster for traditional enterprises to embark on digital transformation and upgrading. Due to labor shortage and

restrictions, enterprises are forced to use digital means to maintain production, thereby breaking the impasse of "being afraid of transformation", " not knowing how to transform", "not having means to transform" and thereby promoting the leapfrog development of investment in digital transformation.

Concluding Remarks

On February 23, General Secretary Xi Jinping hosted at a meeting to advance the work on coordinating the prevention and control of the COVID-19 and economic and social development. At the meeting, he pointed out that the current epidemic situation is still grim and complex, and the prevention and control are at the most critical stage. Whether it is to analyze and judge the development and trend of the epidemic situation, accurately implement the EPC strategy, or provide convenient applications & services for the people and promote the orderly resumption of work and production, we can see that data and smart applications are playing a key role in this battle without gun smoke. Here, we call on the tech communities to unite together and give full play to the important role of data and intelligence in the prevention and control of the epidemic, be determined to fight and win the battle against the epidemic, and secure finishing the building of a moderately prosperous society in all respects and completing the 13th Five-Year Plan to achieve the victories in both EPC and economic and social development!

Acknowledgements

Cloud Computing and Big Data Research Institute, China Academy of Information and Communications Technology (CAICT) is the lead organization in writing this report. Other entities participating in the writing includes: Baidu, Alibaba Cloud, China Unicom Big Data (CUBD), China Telecom Corporation Limited Cloud Computing Branch Corporation, China Mobile Beijing Branch (Beijing Mobile), JD Cloud, Didi, Hangzhou Dt Dream, Chengdu Sefonsoft, Neusoft, Taiji, Inspur, Huawei, Thunisoft, CASIC , Tencent, HKUST Xunfei, TravelSky, 4Paradigm, Sugon Cloud Computing Group Limited, Beijing DataOJO, Xiamen DataExa, Beijing Xiangshu Technology, China Mobile Xiong'an ICT Co., Ltd, Insight Smart Technology Co., Ltd, idatasing, LongShine and Meishu Tech. We have also received the strong support of many other entities and experts in writing this report, and I would like to thank them for their support!