# Chinese Open Source Data Collection, Big Data, And Private Enterprise Work For State Intelligence and Security: The Case of Shenzhen Zhenhua

**By:**

**Christopher Balding**

**&**

**Robert Potter**

**&**

**Et. Al.**

**Both authors contributed to the writing and research of this paper and names are only listed alphabetically. Et. Al. refers to others who contributed and we wish to acknowledge but whose names need to remain anonymous.**

## Summary

China is collecting vast amounts of open source data to support influence and intelligence operations through private enterprises it then sells to state institutions. Here we present one database collected on 2.4 million individuals around the world from sectors China deems as targets for a variety of purposes ranging from political influence to intellectual property targeting. The data appears used to support Chinese intelligence, military, security, and state operations in information warfare and influence targeting. The data covers a broad array of public and non-public data with classifications and rankings on individuals and institutions designed to assist Chinese analysts. The company also provides big data analytics as well as other functionality to support Chinese military and intelligence analysts. Individuals and institutions in open liberal democracies need greater understanding and privacy rights based upon the asymmetric information warfare being undertaken by the Chinese Communist Party and state security intelligence.

## Introduction

China big data and surveillance efforts domestically are widely acknowledged. It has always been assumed that China gathers large amounts of information domestically, but its efforts on foreign national and institution monitoring have received little attention. China through arms such as the United Front Work Department (UFWD) and others under a variety of institutions such as the foreign ministry engage in broad monitoring of foreign individuals and firms. To date however, little evidence has been produced on specifically how China is monitoring foreign individuals and institutions.

Here we present leaked data from a Chinese company serving the Chinese government, security, and military clients monitoring foreign individuals and institutions. This paper will focus less on analyzing the data as it is too vast to be easily analyzed for any narrow sense. Instead this paper will focus on describing the contours of the data, how it appears to have been collected, and how it is likely used by clients of the company. For security reasons, we will not be revealing names of individuals and other key information that may present various types of security risks.

## Overseas Key Information Data Base

The Overseas Key Information Database (OKIDB) was created by a company called China Revival based in Shenzhen, China. The data was provided to Christopher Balding by a China based source connected to the company China Revival. At no point were any methods like phishing, hacking, malware, or unauthorized passwords used to obtain the data.

The data was analyzed on multiple levels to try and better understand its importance and reliability. Numerous methods were used to authenticate the data, that it was compiled by China Revival, and the various methods used by them to compile the database.

China Revival claims on their website to have records of more than 2.4 million individuals, 650 thousand organizations, 2.3 billion news articles, and 2.1 billion social media posts. The data we have matches in scope and scale what China Revival claims to have.

While we do not have access to their source code that would show us directly how they are compiling OKIDB, we were able to impute by deduction various means they would have or could have utilized to construct this database. Most of these means are widely known within the data collection industry and utilize relatively straight forward data harvesting techniques.

The database is constructed around numerous existing databases or platforms and harvested into one large database with multiple points of overlap. Significant amounts of data comes from the Dow Jones owned database Factiva. We do not believe China Revival is using or obtained Factiva data in an authorized manner but we cannot say this with certainty.

China Revival appeared to crawl information platforms to build out personal and professional profiles of key individuals globally. This crawling capability appeared to be built by China Revival and adapted for their needs. Personal data was built out around key individuals to provide familial and business contact links to provide analysts greater understanding of potential targets.

The data was crawled from such well known platforms as Facebook, Twitter, and LinkedIn as well as others. In addition to personal information, OKIDB logged information on posts, likes, and reTweets. This allowed for a wide variety of relationship and key person targeting.

OKIDB logged photographs of individuals where possible in third party open links for analyst reference where possible. The photos were taken from public sites the targets maintained accounts on such as Twitter, Facebook, or LinkedIn. These photos from public sources of the targets appeared crawled and subsequently stored on publicly available Chinese servers with links in the database for each individual.

The database also maintained records on think tanks and university specifically for key personnel and work output. It maintained links to papers, blog posts, videos, and other content by institutions and personnel. One part of the database even marked individuals as "Important". Most individuals and institutions were scored though we were unable to reverse engineer the scoring algorithm though we were able to determine some patterns in the scoring. For reasons of privacy, we will not be publicly disclosing the names of those individuals, institutions, or the scoring patterns.

Large amounts of the data was open source data, there was approximately by our estimates 10-20% of the data was not publicly or easily available from public sources. Some of the non-public data on individuals we remain unsure where the data comes from or how it was obtained. We have reason to believe some of the data comes from unauthorized data access such as hacking but we cannot be certain. Non-open source data had a tendency to tie to higher security individuals but not always.

The database included large amounts of public sector employee records. This includes everyone from known politicians to political aides to low level military personnel. The breadth of data capture was quite extensive.

The database extends well beyond pure storage in functionality. Designed to assist the Chinese government, security, and intelligence services, OKIDB adds in multi-layered functionality to help target and link individuals. Though not extensive, we found analyst notes about certain targets. Certain indexes had classifiers for individuals or institutions such as importance.

It also assisted in a variety of relationship mapping. For instance, it recorded broad family relationships and work history. It also had other more complex big data capabilities that allowed relationship and network mapping from business networks to personnel linked to a carrier ship to social media influencers.

The database appeared to have links to related databases. We are not sure if this was client specific, a higher security database, or maintained by a separate company with which Shenzhen Zhenghua had a partnership. However, there were links in the database found linking to other databases.
It was also notable information conspicuously absent from the data. Many institutions and individuals known to have deep interest or links to China were generally absent from the data. Based upon the consistency of this finding we believe that these institutions and individuals were kept in separate databases for numerous reasons.

It was also interesting the number of names found in the database that comes from organized crime. There is no clear reason for the inclusion of such inclusion of a significant number of organized crime figures at both the widely known level as well as lesser known figures in various criminal organizations.

While the database draws from existing database and crawls significant amounts of information, it is also clear the database has extensive human interaction on many levels. We have reason to believe the individuals or specific classes of individuals are specifically targeted. We cannot say who exactly determines the targets and what mix there is between the human and algorithmic interaction but human analysis here plays a not insignificant role.

The database appears to focus on individuals and institutions China deems influential or important. The general classes of individuals and institutions that appear in the database are known to be of significant interest to China across policy domains. From politicians and their families to professors and think tanks to scientists and tech leaders to organized crime figures, all are individuals and institutions Chinese security, intelligence, and influence operations are known to be interested in targeting.

### How is OKIDB Data Used?

Due to the nature of the data we have, analysis of how the data is used is more inductive however, we can still draw clear conclusions about the Chinese state is using OKIDB data.
Shenzhen Zhenhua claims on their website and related material that their clients come from military, security, intelligence, and Chinese state institutions. Based upon the background research we have completed about Shenzhen Zhenhua, its employees, and linked individuals and institutions we believe this claim to be accurate.

To verify this claim, we conducted numerous types of background research on a variety of entities. This included, but is not necessarily limited to, open source intelligence checks on key individuals and linked entities. Available evidence supports Shenzhen Zhenhua claims it is closely tied to these Chinese state as well as military and security agencies such as the People's Liberation Army and the Ministry of State Security.

The database functionally appears to be used for a few purposes, again drawing clues from Shenzhen Zhenhua statements. A fundamental purpose appears to be information warfare something Shenzhen Zhenhua talks about openly. The database includes a big data analytics layer that allows analysts to track key influencers and how news and opinion moves through social media platforms. China via multiple channels has moved actively into public information platforms attempting to influence the debate and narrative about China. The data collected about individuals and institutions and the overlaid analytic tools from social media platforms provide China enormous benefit in opinion formation, targeting, and messaging.

From the assembled data, it is also possible for China even in individualized meetings be able to craft messaging or target the individuals they deem necessary to target. Open source data to intelligence agencies has become enormously valuable. This is why intelligence agencies globally have warned against LinkedIn recruitment via Chinese institutions in one specific instance. What is notable about the database of social media, platform, and personal data Zhenhua assembled is the breadth of individuals in specific fields China is seeking to influence.

We were unable to see direct evidence of Chinese agencies using this data to craft information warfare campaigns, messaging, anonymous account usage, or individual influence targeting. This is because OKIDB is input data but provides no direct evidence of the activities that use this data. However, the OKIDB is designed for that purpose and institutions accessing the database are known to engage in influence and intelligence activities. It is likely OKIDB is informing their decision making processes about operations, targets, individuals, and institutions.

## Conclusion

China is known to be building a techno-surveillance authoritarian state domestically. Here we provide the first direct evidence of data collected by China on its monitoring and data collection on foreign individuals and institutions for purposes of intelligence and influence operations.
The unique blend of civil-military fusion pushed by China that works with private firms to engage in state policy activities such as intelligence gathering should be concerning. Foreign individuals and institutions working in sensitive or influential sectors need to be aware of how China is targeting them for influence operations. China is using a variety of firms and channels to gather data to inform its influence and intelligence operations.

Open liberal societies fail to grasp the threats embodied in Chinese authoritarian communism by ignoring non-traditional warfare and influence operations. The information warfare being touted by Zhenhua targets key institutions in democracies such as the children of politicians, universities, and key industrial sectors. These flow into information transmission and policy formation. Open liberal democracies would be wise to improve data privacy and security and understand the threats.