

## Postprint: Classic Cases and Development Trends in Data Journalism, 2016

**Authors:** Dai Yu, Li Jiaxuan, Cao Ziyi

**Date:** 2023-10-08T00:00:00+00:00

### Abstract

of Fallen Officials Nationwide Since the 18th Party Congress.” This database covered publicly disclosed information on over 19,000 officials at or above the county/division level reported by disciplinary and judicial authorities between November 8, 2012 and December 31, 2015. In addition to query functions, the H5 provided statistical displays, such as crime-type distributions and provincial comparisons of fallen officials.

### Full Text

#### 2016: Classic Cases and Development Trends in Chinese Data Journalism

By Dai Yu, Li Jiaxuan, Cao Ziyi

2016 marked a year of vigorous development for data journalism in China. Whether in traditional print publications or web portals, market-oriented media or state-run outlets, industry practitioners or academia—all began exploring data journalism, producing impressive works.

A clear “advanced” trend emerged in 2016: database journalism became the highlight, with media organizations building exclusive “news databases” equipped with web-based interactive query and matching functions. Previously, most data journalism presented data conclusions and statistical results as the subject of reporting, leaving the public unable to access primary data sources for their own queries and analysis.

This shift occurred for three main reasons. First, traditional print media offered poor interactivity and could not present large-scale data within limited space. Second, audiences had limited capacity to interpret data, and media outlets saw little necessity in making databases public. Third, media organizations lacked

both the capability to build and report on databases and the attention toward database development.

However, this situation changed markedly. A representative example was the H5 interactive piece released by Phoenix News during the escalating Wei Zexi incident, which allowed users to query “how many Putian-system hospitals are in your city.” As public criticism focused on the Putian hospital network, this database journalism piece timely addressed everyone’s primary concern. The query function enabled each person to extract “news from their own surroundings,” triggering widespread dissemination.

Similarly, Tencent News’s “Fact School” column launched “Hello Discipline Commission, Well Done—A Summary of Fallen Officials Nationwide Since the 18th Party Congress.” This database covered publicly disclosed information on over 19,000 officials at or above the county/division level reported by disciplinary and judicial authorities between November 8, 2012 and December 31, 2015. In addition to query functions, the H5 provided statistical displays, such as crime-type distributions and provincial comparisons of fallen officials.

Database journalism is characterized by large data volumes and comprehensive cases, allowing almost every reader to query results relevant to themselves or their locality without predetermined conclusions—achieving a “thousand faces for thousand people” reporting style. However, from another perspective, such data journalism is time-consuming, leaving insufficient time for interview interpretation and analytical depth.

In addition to query-based databases, 2016 also saw the emergence of matching-style gamified database journalism H5 pieces, such as the joint production by China.com.cn and Tuzheng Data on National People’s Congress representatives. This reporting package included two databases: one of online public opinion survey results about young people’s views on NPC representatives, and another NPC representative database containing manually collected information on representatives’ current positions and industries gathered from public news reports. The H5 “Who Can Represent Me?” was based on this representative database, designing a matching algorithm according to audience understanding of representation. Readers input their gender, age, ethnicity, industry, party affiliation, and province to automatically calculate matching scores and display information on the most compatible representatives.

In fact, the long-term development value of database journalism should far exceed that of a single breaking news story. Databases accumulated through lengthy collection processes can be permanently published and regularly updated as web pages. Unfortunately, current practices regarding database preservation, subsequent mining, and structured public release remain inadequate. Over time, this type of database journalism may struggle to secure sustained investment and support, potentially limiting improvements in mining depth and data quality.

Alongside the database journalism boom, Chinese data journalism made clear

strides in data visualization in 2016. The People's Daily Central Kitchen's H5 "Complete Earthquake Records 1976-2016" used screen recording to achieve three-dimensional display of millions of global earthquake data points on mobile devices, allowing audiences to interactively explore earthquake distribution patterns in China over nearly four decades.

However, the true breakthrough in visualization design and interactivity belonged to "Do You Really Know Your City?"—a quiz-style H5 from CBN-weekly's New First-tier City Research Institute. This piece included lifestyle data for over 300 cities, such as per capita video viewing time on Youku, subway operating hours, average spending per meal or movie, and counts of restaurants, civil airports, pet shops, super high-rise buildings, and coffee shops, inviting audiences to guess specific values. The H5 broke through conventional chart formats like bar or pie charts in its interactive guessing mechanism, enriching the presentation beyond simple narrow-block selection or linear dragging. For instance, audiences could rotate clock faces, dashboard gauges, or circles along arcs, or vertically adjust four columns of "abacus-bead" style numbers for guessing. These interaction forms were rarely seen in previous data journalism.

Having reversed the traditional notion of "heavy interpretation, light data," China's data journalism field in 2016 moved toward "data is king" in database journalism. While continuing efforts to break through in data analysis, interpretation, and interviewing, the field still saw few in-depth investigative reports integrating multi-party verification, and to some extent retained weaknesses in data interpretation and relative absence of data-driven interviewing. Data journalism, after all, involves numerous divisions and complex processes, encompassing data collection, analysis, interview interpretation, reporting or interactive copywriting, visualization, and technical support. In early development stages, it is likely to emphasize certain aspects while neglecting others.

In terms of overall quality, Caixin's "From Regulation to Stimulus: A Decade of Housing Market Cycles" remains exemplary. From a journalistic perspective, this work featured mature, sober writing covering case interviews, data analysis, and policy research, with relatively complete content sections. Even in data visualization, it achieved high standards, experimenting with 3D interactive displays using 联动 chessboard and curve charts to present price fluctuations across 70 cities from July 2005 to July 2016. The interactive design addressed audience needs—simply entering a city name and household monthly income automatically calculated how many years of saving without spending would be required to purchase an 80-square-meter home there. Even the labeling of data sources and calculation methods was detailed and standardized.

Beyond moving toward "more complete data" and "better design," Chinese data journalism in 2016 also strove for "faster and newer" goals. Phoenix News's "Live US Election Vote Count" achieved a "screen-flooding" effect on election day, allowing audiences to check real-time vote results and comparisons between the two candidates via mobile phones. The H5 featured accurate data, stable systems, and clean design, representing something pioneering in China's data

live-streaming. It also demonstrated that Phoenix News's data journalism works generally had strong news value, seemingly not deliberately emphasizing the data aspect but rather applying data journalism's timeliness more prominently.

Overall, Chinese data journalism in 2016 indeed advanced in the three main directions of data, journalism, and visualization, with particularly strong momentum on mobile platforms. However, this progress could only be described as within expectations and reasonable—building upon past standards with certain efforts and improvements, yet not reaching explosive growth or clear breakthroughs.

As exemplified by the South China Morning Post's 2016 Information is Beautiful Design Competition Community Award (determined by global data visualization community online voting) gold-winning work "China Provincial Economic Growth 1996-2015," the data might not seem novel to Chinese audiences but gave foreigners intuitive understanding of China's growth speed in international voting, truly achieving journalistic value.

Looking ahead, data journalism may see more visualization progress in 2017, though comprehensive in-depth reporting may not necessarily achieve greater mining depth. It is believed that more Chinese data journalism works will appear on the international news stage. However, regardless of how data journalism develops, it must ultimately move toward comprehensive investigative reporting that integrates databases, news points (new data discoveries), interviews, and visualization. DT Finance's report "100,000 Registration Data Points Show: Even the Internet Can't Solve Difficulty in Seeing Doctors" analyzed which provinces' hospitals better support online appointment booking, which hospitals have "overloaded" appointment volumes, and which hospitals received low ratings—all from registration data.

**(Author Bio: Dai Yu, former editor at Southern Window, former head of the People's Daily Central Kitchen data journalism project, led team to win second prize in China's first Data Journalism Competition and was longlisted for the 2016 British Information is Beautiful Design Competition)**

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv — Machine translation. Verify with original.*