The "Daiyu Data Circle" Salon: Pioneering and Exploration in CCTV's Data Journalism Post-print

Authors: Guo Junyi, Dai Yu, Xue Kui, Gao He

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

Abstract

Editor's Note: "Dai Yu Data Circle" is a data journalism community founded by veteran data journalist Dai Yu, bringing together China's top data news editors, team leaders from major media outlets, liaisons from leading think tanks and data source companies, and university professors who teach data journalism. What new forms of reporting are these Chinese data journalists pioneering? The community regularly hosts cutting-edge salons that reveal new trends in China's data journalism landscape.

Full Text

Dai Yu Data Circle Salon: The Creation and Exploration of CCTV's Big Data News

By Guo Junyi, Dai Yu, Xue Kui, Gao He

Editor's Note: "Dai Yu Data Circle" is a data journalism community founded by veteran data journalist Dai Yu, bringing together China's top data news editors, team leaders from major media outlets, liaisons from leading think tanks and data source companies, and university professors who teach data journalism. What new forms of reporting are these Chinese data journalists pioneering? The community regularly hosts cutting-edge salons that reveal new trends in China's data journalism landscape.

Host: Dai Yu, former editor at *Southern Window* and former head of the data journalism project at People's Daily's "Central Kitchen." Her team won second prize at China's first Data Journalism Competition and was longlisted for the 2016 Information is Beautiful Awards.

Guests: Guo Junyi, Deputy Producer at CCTV News Broadcast Editorial Department and pioneer of China's big data news.

Guo Junyi: Hello, friends from the data circle. Today, I'd like to introduce several groundbreaking and classic cases of CCTV's big data reporting, and then share some thoughts on the future of data journalism.

1. The "Supposedly Spring Festival Migration" Series

Many of you are probably familiar with our early collaboration with Baidu on the "Supposedly Spring Festival Migration" series. Many people remember seeing Baidu's Spring Festival migration map on CCTV, but few know how we actually produced these stories. Today, I can reveal the process.

In January 2014, just before the Spring Festival, Baidu had developed a "Baidu Spring Festival Migration Map" product. Back then, people might have found the migration map visually impressive, but it was essentially a data visualization product—not yet a news story. I must emphasize that a visualization product lacks journalistic value and therefore doesn't attract audiences.

The real challenge was how to mine actual news from big data. Baidu didn't provide us with news angles, so we had to dig them out ourselves. I spent considerable time examining the migration data for many cities—population inflows, outflows, popular routes—and eventually discovered something significant. Traditionally, China's Spring Festival migration pattern involved movement from mega-cities to smaller ones, and from coastal regions to central and western cities. This conventional pattern alone wouldn't constitute news. But I noticed that migration volume from Beijing to Chengdu was enormous, while outflow from Chengdu to Beijing also ranked among the top three for several consecutive days. From this data, we identified a new migration trend: reverse migration—from central and western cities back to mega-cities. Of course, this analysis was based solely on the mobile location big data provided by Baidu.

Since this was our first attempt at big data journalism, we felt uncertain about deriving a news angle from commercial company data. As China's national broadcaster, would CCTV be seen as authoritative if we published such findings without verification? Therefore, we supplemented the big data with traditional investigative methods, asking our Sichuan bureau reporters to interview Chengdu Railway Bureau staff. Their data also supported our finding: actual ticket sales confirmed that Chengdu's outbound volume was indeed exceptionally large.

Even with data from Chengdu Railway Bureau, we still didn't feel completely confident. Television requires visual evidence to support the story. So we asked Chengdu Railway Station staff to film live footage of the station's waiting hall, which perfectly corroborated our angle: the footage clearly showed many elderly people preparing to travel. With only six days until Chinese New Year's Eve, why would elderly people travel? They weren't going out to work or for tourism—they were likely making reverse migrations to spend the holiday with

their children in the cities where they worked.

Our first program thus used cross-verification to confirm the news angle and support our content. The next challenge was transforming this into an engaging visualization for broadcast. We considered various methods—filming computer screens, connecting laptops directly to studio screens—but none worked well. Eventually, we settled on a simple solution: screen recording. We used screen recording software to convert Baidu's Flash-based dynamic migration map into video format, which could then be broadcast directly on television.

The first program's success was enormous. Many viewers and colleagues in academia and industry had never seen this format before, so the response was significant and encouraging. Subsequently, we produced many more series under the "Supposedly" brand, covering Spring Festival, Chinese New Year, Two Sessions, World Cup, APEC, and more.

2. Big Data User Profiling

Let me share another program that I consider particularly well-executed—a "Two Sessions" report in collaboration with 360 Company. At the time, online fraud was a major concern, so we discussed with 360's Chief Security Officer, Dr. Pei, how to use 360's security big data for relevant reporting.

Dr. Pei mentioned they had abundant security big data, but these highly technical datasets couldn't directly become news stories. We media professionals needed to find the right angle. I suggested using 360's security big data to create a user profile—describing what kind of people were most vulnerable to online scams. 360 spent two or three days creating this profile.

The results were fascinating: boys aged 16 to 25 (post-90s generation) living in Guangdong, engaging in online part-time jobs, shopping, and gaming between 10 AM and 12 AM, were most likely to be defrauded. While interesting, this was just a data-derived profile. We wondered whether this matched reality.

Coincidentally, the day before we completed the profile, Sun Yat-sen University's official Weibo account posted that several male students had recently been defrauded while searching for jobs or part-time work online. This real-world case perfectly matched our big data-derived user profile, demonstrating the remarkable power of big data.

3. The Shanghai Bund Stampede Incident

I'd also like to share the case of the Shanghai Bund stampede. As you know, on December 31, 2014, a tragic stampede occurred at Shanghai's Bund, killing 36 people. Upon hearing the news, I thought about big data's early warning capabilities and whether we could use it to prevent crowd crushes at large events.

Crowd density mapping, which we now call heat maps, is based on LBS (Location-Based Services) big data from mobile positioning. I approached

Tencent Maps with this idea, and they cooperated to create heat maps showing crowd changes at Shanghai Bund on the day of the accident, as well as the day before and after, plus a crowd change trend chart.

Both visualizations clearly showed how crowds changed over time. If the relevant Huangpu District authorities had used big data for early warning, this tragedy might have been prevented.

I found this approach to preventing stampedes through crowd analysis so meaningful that I later discussed the idea with China Mobile's leadership. China Mobile subsequently asked Beijing Mobile to develop anti-stampede software specifically for Beijing's temple fairs, which was successfully deployed during the 2015 Spring Festival at the Ditan Temple Fair.

4. Data Micro-Commentary

Next, I'd like to discuss our latest approach. At the end of 2016, *News Broadcast* produced an annual series called "Point-by-Point Reform Annual Account," consisting of eight episodes. Here we experimented with a new format worth sharing.

The series title contains two "points": the first means "review" (using data to review reform measures and achievements), and the second means "commentary" (each data point was accompanied by a one-sentence comment). We later summarized this as "data micro-commentary," a relatively rare format in news reporting that represents a small breakthrough.

Host Dai Yu: Thank you, Teacher Guo, for these vivid cases. CCTV combines big data analysis, on-site interviews, data visualization, and visual language to create comprehensive products with broad influence. My question is: what do you think are the advantages and disadvantages of television media doing data journalism compared to other media forms?

Guo Junyi: Television's advantage lies in visual impact. TV can employ rich dynamic visualization effects to present content, including data, making the final product visually appealing and intuitive. However, the disadvantage is that in-depth investigative reasoning is difficult to unfold in television programs, and the production cost for each data news piece is relatively high.

Dai Yu: You mentioned costs. Some colleagues in our circle are curious: does CCTV pay data companies when collaborating on data journalism?

Guo Junyi: Our costs are mainly in post-production. Collaborations with big data companies don't involve financial transactions. For news media, the primary consideration in data journalism is news value—we want news value, while data companies want dissemination value. Both parties have cooperative needs. Of course, CCTV's large platform is also an important factor attracting companies.

Some data companies have complained to me that although media collaboration can uncover unexpected insights, when data conclusions derived from costly data processing and manpower aren't used by media, it feels wasteful. They consider the cost-effectiveness between data analysis and media impact.

That's understandable, but for media, news value is the first priority, and media space is limited. Big data journalism requires significant investment—Baidu assigned more than 20 people to cooperate with us at most—but the social impact of selected topics is also considerable.

Data conclusions not used in media reports can be published by data companies on their own WeChat official accounts. Of course, this requires better communication between media and data companies. If journalists can articulate their topic requirements more precisely, it can avoid wasting data conclusions.

Indeed, good data journalists need to harness big data for news purposes. Sometimes misinterpretation can lead to "false reporting." Some scholars from the Chinese Academy of Social Sciences have raised concerns that some academics don't recognize the scientific significance of data, still believing that only traditional social field surveys are more authentic. What's your view on the scientific validity of data journalism?

I've discussed using big data for legal studies with a law professor—they hadn't heard of it at the time, though relevant papers have since emerged. I'm not a sociologist, but I know traditional social surveys use sampling, while big data represents full samples. I believe full samples are more valuable. I'm not a statistician either, so I wouldn't deny the value of sampling, but since we have big data and analytical capabilities, why not try analyzing massive datasets? I think data source selection should prioritize authority and leadership in respective fields, and media can use traditional surveys and cross-verification to mitigate data source risks.

Perhaps the combination of data scientists' mining capabilities and journalists' field investigations represents the future collaborative trend for both the data and journalism communities.

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