The Development of Data Journalism in China: Influences, Motivations and Practice

Professor Scott Wright, School of Media, Film and Journalism, Monash University, Australia. Scott.wright@monash.edu

Associate Professor David Nolan, News and Media Research Centre, University of Canberra, David.nolan@canberra.edu

Abstract
Using semi-structured interviews with Chinese data journalists across party and commercial media, this article assesses the structure and practice of data journalism in China. In doing this, it responds to calls for further studies of data journalism in non-western contexts. It finds that Chinese data journalists face some of the same pressures and challenges that have been documented in other countries, including limited access to data and the constraints imposed by the screen-size of smartphones. However, these were often exacerbated through a combination of social and systemic factors - to the point that their impact is qualitatively and quantitatively different. Simultaneously, however, we find that in some cases Chinese data journalists, at least amongst party media, were protected from pressures such as audience demand, and encouraged to focus on state-of-the-art work. We conclude that what has emerged is a form of 'data journalism with Chinese characteristics', and that these characteristics emerge from the interactions between systemic, newsroom and social factors.

Keywords: data journalism, Chinese journalism, computer-assisted reporting, digital journalism, data-driven journalism

Introduction
Once described as the "new punk" (Rogers 2012) and "the future" of journalism (Berners-Lee 2010), data journalism has spread around the world (Rogers et al. 2017) as part of a quantitative turn in the field (Coddington 2015). Despite the emergence of a burgeoning field of studies on data journalism, Fink and Anderson (2015, 479) argue that because the vast majority focus on Western and democratic contexts we need to "broaden our analytical lens to include other countries, particularly those outside of North America and Europe". Appelgren et al (2019) note that this problem still persists, with a preponderance of work on Anglo-American and Northern European 'democratic corporatist') media systems. This article addresses this gap by considering the motivations for, influences on, and development of, data journalism in China. The article begins by discussing extant research focused on data journalism internationally. It then considers nascent research on data journalism in China, framed in the context of wider debates about how contextual factors influence data journalism. Building on this work, it draws on data from 18 semi-structured interviews to investigate influences on, and drivers of, the adoption of data journalism practices in China, how Chinese data journalists perceive its development, and how data journalism might be understood as both affected by and contributing to change in Chinese journalism. Our findings suggest that while some trends in Chinese data journalism mirror Western developments, contextual factors have contributed to a specific trajectory and practices that are particular to China.

Data Journalism

A wide range of studies have assessed the practice of data journalism. Fink and Anderson (2015) divided this work into three strands. The first strand addressed practical concerns, analysing possibilities and challenges, and its potentials for contributing to journalistic performance (Flew, Daniel and Spurgeon 2010). Research in non-democratic contexts and the global south often focuses specifically on barriers (Lewis and Nashmi 2009: 1207; Palomo et al. 2019: 1272) though the use of Western analytical prisms is often problematic (Mutsvairo 2019; Wright et al. 2019: 1296). A second strand considered how the emergence of data journalism was both enabled by, and had implications for, the development of new actors in the relations that constitute news production. These have included internal and external data specialists (Borges-Rey 2016; Felle 2016) and the role of NGOs and "big tech" particularly in the 'developing world' (Cheruiyot 2019; Palomo et al. 2019; Mutsvairo 2019: 1292). A third strand has taken an historical approach analysing how data journalism has changed over time (Parasie and Dagiral 2012; Coddington 2015) including the relations that influence its production (Wright and Doyle 2019).

In addition to these three strands, studies of the form and content of data journalism have found that the predominant topics are politics, society and business (Rogers et al. 2017). While data journalism awards often go to interactive stories (Young, Hermida and Fulda 2018), research suggests that most work is "largely superficial, institutionally sourced and non-remarkable" (Knight 2015, 70) and adheres to "traditional journalistic values, norms, and practices" (Tandoc and Oh 2017: 1008). This links to work on the production of data journalism, drawing a distinction between a small number of well-resourced newsrooms committed to team-based, often interactive data journalism projects (Wright and Doyle 2019; Appelgren and Nygren 2014), and the wider practice of individual data journalists producing stories (Stalph 2017) often under considerable time pressure (Fink and Anderson 2015, Appelgren and Nygren 2014).

In summary, a burgeoning literature has assessed data journalism. The focus of this research on the practice of data journalism, particularly 1) the skills and training of journalists 2) the position of data journalism in the newsroom, and 3) the techniques, forms and contents of data journalism stories - and the factors that have influenced this - informs this study. However, extending our analytical lens beyond North America and Europe to China requires a consideration of how different influences impact Chinese data journalism.

Contextual Factors Shaping Data Journalism: the case of China

The extent to which the practice of data journalism in China has similarities to North America and Europe, or has taken pathways and processes shaped by the dynamics particular to China's media environment, remains underexplored. The relatively small number of studies that have looked at data journalism in non-Western contexts find distinct cultural "orientations" at the region or country level (Mutsvairo 2019: 1289; Mutsvairo et al. 2019) whether it be in the Middle East (Lewis and Al Nashmi 2019), Latin America (Palomo et al. 2019), or Africa (Cheruiyot 2019). Appelgren et al (2019) link these differences to three factors: distinctive journalistic cultures, media market structures and the political contexts in which they emerge. This framework broadly mirrors the key variables identified by Hallin and Mancini (2004) as shaping media systems: the role of the state, the links between media

and political actors, the nature of media markets and the nature and influence of journalistic professionalism. Fink and Schudson (2015, 479) argue that such an approach can provide a "larger thematic framework to compare different understandings and practices of data journalism" (Fink and Schudson 2015, 479). However, Hallin and Mancini's approach has been subject to criticism, with key critiques centring on the limited attention paid to the role of technological change, in particular how digital technologies contribute to the transformation of media systems (Benson et al. 2012, Zhao 2012). In addition, focusing on China in particular, Zhao (2012) questions a 'one-size-fits-all' model, centred on an 'interventionist' model of the state regulating the media at 'arm's length'. Such a model, she argues, is inadequate for an authoritarian state where, despite a distinction between party and commercial media outlets, all media are state licensed, markets are shaped and promoted by both state policies and the marketization of state entities, and ownership structures are relatively opaque. Journalistic professionalism is likewise impacted by the state, which serves as a defining influence in processes of journalistic education and recognition (Pan and Lu 2003).

As we have elsewhere argued, however, it does not follow that all variables shaping Chinese media practices should be viewed as entirely dependent on the state (Simons et al. 2017). While there may be no clear distinction between market and state, China's promotion of a market economy has generated its own dynamics, including an increased need to be responsive to consumers and the emergence of powerful actors who exercise an influence that, though neither autonomous nor necessarily antithetical to state power, are nevertheless guided by distinct concerns and priorities. Likewise, journalistic professionalism should not be viewed as merely an extension of state power (Simons et al. 2017). On one hand, journalists are influenced by different, historically accrued and occasionally competing models of professionalism that derive from Confucian, liberal, party and market traditions (Lee 2005). On the other, Chinese reporters are increasingly subject to the influence of other traditions encountered through international media consumption, overseas training and education, and international civil society initiatives that offer data journalism training. The rise of market-based and consumer-oriented models has promoted professionalization in Chinese journalism, while also creating pressures that have produced sensationalism and some unethical journalism practices (Simons et al. 2017, Tong 2011). In addition, the meteoric rise of Chinese social media contributed to a flowering of investigative journalism at the turn of the century, fuelling optimism about an increased press freedom. Tempering such optimism, however, Repnikova (2017, p.145) argues that 'critical reporting...has been encouraged in China only insofar as it facilitates effective propaganda', while Tong (2019) suggests that even limited moves toward media liberalization have, in recent years, been eroded by a mixture of political, economic and technological forces.

A research agenda concerned to understand the development of 'data journalism with Chinese characteristics', then, must take account of the changing dynamics that contribute to broader processes of journalistic transformation in China. These dynamics involve changing formations of state power and political relations, transforming media markets and actors, and inherited and emergent models of professionalism. In addition, however, these dynamics also reciprocally affect and are affected by changing social relations in which the availability and affordances of digital technologies play a constitutive role.

Existing research into Chinese data journalism is limited, however. Zhang and Feng (2018) find that Chinese data journalists frequently incorporate data visualisation, but journalists rarely provide audiences with access to original datasets. The findings drawn from their

qualitative interviews present a mixed picture. Most respondents were enthusiastic about the possibilities of data journalism, viewing it as enabling new and attractive modes of storytelling, efficiencies in production and attracting career esteem. However, practitioners were concerned about the heavy time-investment involved in data journalism projects and their commercial viability. Some respondents saw data journalism as an imported tradition and were sceptical about its compatibility with China's media system in the longer term. The majority of respondents expressed concerns about the availability and quality of data. Several also noted that work pressures limited the time available to produce complex stories, and the positioning of data journalism as a supplementary add-on to existing organisational structures and practices (Zhang and Feng 2018, 12-14).

Du's (2019) case study of the Caixin VizLab found that the quantity of reports has declined year on year since 2015 from 375 down to 54, with lifestyle and politics (either international or non-sensitive) the most prominent topics. The most frequent sources for data were non-governmental institutions and then government, with most data drawn from pre-processed datasets rather than collected or scraped by the journalists (Du 2019, 115).

While both of these studies present valuable empirical data, neither study really details specifically Chinese characteristics to data journalism, and the broader dynamics that are shaping this. This study seeks to address this challenge. The approach adopted here is to focus in detail on the factors and forces that have shaped the development of data journalism in China, including the possible motivations and drivers for the adoption, maintenance and trajectory of data journalism practices, and how Chinese journalists perceive that development.

Methodology

To address this challenge, the study draws on 18 interviews with Chinese data journalists, of which 16 were trained in the US or UK. 10 of these were working for party media at the time of interview; two were working in commercial media but previously worked for party media; two were former party data journalists working in education; two worked for commercial media and two were independent data journalists writing on WeChat and occasionally for international commercial media. Interviewees were primarily from Shanghai and Beijing but two were from larger provincial cities. Interviewees were identified using a mixture of snowball sampling and searching online for people who identified as data journalists (we found limited evidence of data journalists working outside of the larger Chinese cities). A snowball technique helped us to overcome potential trust barriers. We were initially assisted by Professor Deng Jianguo of Fudan University, who also helped us establish trust with interviewees. Five interviews were conducted face-to-face in China, eight were via online audio calls and the remainder were by written response to questions. The face-to-face and audio-call answers were significantly longer (average of 9663 vs 1880 words) than written responses, likely because there were follow on questions and writing responses is more onerous and time-consuming.[i] This limitation was offset because they allowed us to broaden the range of interviews. The main questions were standardised, but other questions were tailored to the specific organisation, and the 'live' interviews were semi-structured with follow on questions asked in response to answers. Questions focused on background and training; the strengths, weaknesses and challenges that were faced; and the structure and practice of data journalism.

Following the interviews, responses were transcribed, and a thematic analysis was conducted to identify patterns in responses that emerged from the research. Thematic analysis constitutes a method through which "analysts make sense of recurring observations found

within data in order to interpret what is occurring" (Hawkins 2017, 1757). In this case, the mechanism of interpretation was both deductive and inductive, as the researchers both analysed the nature of responses to pre-given questions, and sought to identify themes that emerged from responses and from questions that responded to points raised by interviewees. To protect respondents, interviewees are kept anonymous and identified by numbers only.

The Development of Data Journalism in China

Data journalism has developed rapidly in China. It has faced a steep learning curve because there was no tradition of Computer Assisted Reporting and this has meant that data journalism is "less develop[ed]… at least 20 years later than the US" (interviewee 3). Similarly, interviewee 2 argued that the USA was "much better than us, I think [China is] far behind" but China is "doing much better than two years ago". Several interviewees cited leading western examples such as the *New York Times'* 'Upshot' and *The Guardian's* 'Datablog' as inspirations and that these were "carefully monitored" (interviewee 10). All interviews cited the influence of (particularly) US traditions of data journalism. This is perhaps unsurprising as many had graduated with a degree in data journalism from the USA or UK, and this international education and training has contributed to shaping the skills, practices and trajectory of Chinese data journalism.

There was strong evidence of a shared professional culture across outlets, with WeChat being a key networking and support channel. During an initial development phase from 2012-14 an international NGO, IREX, led weeklong training programs in data journalism (interviewee 14). The program was cut after a state crackdown on foreign NGOs. Data journalism training continued though, via Chinese J-Schools, a data journalism MOOC for the Asia-Pacific, and a (now defunct) Chinese data journalism support website.

Following this, respondents referred to a significant growth phase that ran from roughly 2014 to 2015, as data journalism began to be taken seriously in China. During the growth phase, interviewees considered data journalism to be a "buzzword" in the Chinese media. Interviewee 14 argued that the mentality was "we'll just invest a little bit of money and then we'll see how it goes" with data journalism treated more as "PR" or "cosmetic" that editors "can show to the world that they are ahead of the curve" rather than about "how data can empower people".

In the current phase, interviewees described a rationalization and, to some extent, constriction of data journalism in China, with a decline in the number of data journalists and teams. For example, the Xinhua data journalism team shrank in size from 30 to 10 as they transferred "the efforts and the strength" to short videos (interviewees 1, 18), while a data editor working for a commercial outlet noted that their team initially grew to four, before being cut back to one (interviewee 6). Others told us "some of the teams actually died" (interviewee 14), that "data journalism was not as popular" anymore (interviewee 3), and that even Chinese state media were not immune to economic and audience pressures experienced internationally (interviewee 12).

In spite of the numerous concerns and challenges outlined below, there remained some optimism. The "good sign", one interviewee stated, is that "the publications that still have data journalism teams […] focus more on data driven stories [and] not just [trying to] make a regular story look better" (interviewee 2), while another stated that teams that survived "are really serious, dedicated" (interviewee 14). Indeed, Chinese data journalists have begun to win international awards[ii], and our interviewees spoke at length about how they 'dug' or

'excavated' their favourite data-driven stories. Within the parameters of the system, much high-quality data journalism is produced. We now turn to consider the development of data journalism in China, starting with the organisational structures before turning to different challenges and opportunities.

The Structure of Data Journalism in China

In a similar vein to other countries, there were two main approaches to embedding data journalism within Chinese newsrooms. First, there are bigger, specialised data teams. The most prominent are arguably *The Paper* (8 people specifically focused on data journalism and 12 developers working across the website in July 2018), *Caixin* (4 data journalists alongside a team of developers in July 2018) and Yicai Global. These teams were broadly modelled on international examples. Often data journalists had specific technical skills and focus but covered a wide range of topics, but occasionally data journalists also had a topic beat such as health. Internal collaboration was the norm, though there was sometimes competition between different outlets from the same media group, consistent with long-standing observations of journalists acting as 'competitor-colleagues' elsewhere (Tunstall 1971). For example, data journalists working for one 'traditional' outlet were not invited to a training event run by a newer outlet because they are competitors (interviewee 4). Nevertheless, informal collaboration was routine: interviewee 2 commented that "we help each other frequently because we have really good personal relationships" while interviewee 5 told how they "went downstairs [to a sister outlet] for support".

Second, there are individual data journalists or small teams (2-3 people) from commercial media or print-focused party media undertaking all aspects of data journalism (interviewee 5). Another trend was the rise of independent data journalists on WeChat (interviewee 18). Interviewees 8 and 11 worked independently, while interviewee 14 discussed the emergence of a community of "self-media" data journalists "not formally employed as journalists" but "out of personal interest, publish a lot of things on their WeChat", and suggested this was "more dynamic" because there were fewer direct restrictions on such activity. Here, one can discern how processes of professional networking, distinctive occupational and professional concerns and opportunities afforded by new platforms all contribute to the development of Chinese data journalism, alongside institutional and economic influences.

Another important trend was that interns provided significant labour. While the teams at party outlets often had many interns (interviewee 1, 3), interns seemed particularly important for individual data journalists (interviewee 5, 8) and commercial media. One editor noted that: "we now have more than 30 online interns and freelancers" to help "deal with the short[age] of workforce" (interviewee 6). Another party outlet noted that a lack of staff was an issue, particularly compared to their peers (interviewee 9). Interns, notably, bring cheap labour and a relatively young, digitally native cohort into the space of data journalism production.

Data Journalism 'Talent' and Training

One common theme to emerge was discussion of data journalism 'talent', meaning the still relatively small number of journalists with adequate technical skills and understanding to lead the production of data journalism. Most data journalists interviewed had less than five years' experience. When asked about their personal strengths and weaknesses, most emphasized their technical strengths, but had less experience with reporting (Interviewee 1, 15). Some had no formal training in journalism: "I only learned by doing" (interviewee 8). One editor was conscious that "everyone needs to have a general understanding of data journalism", so all new journalists were given training and tasks to ensure they were adept at handling data

(interviewee 3). However, a common concern was that too much focus was being placed on technical excellence at the expense of journalism excellence, suggesting the adoption of data journalism, driven by economic and reputational concerns, had been viewed through a technological determinist lens. Interviewee 6 said there was "too much attention to programming skills and interactive design, and they forget what a good story should be"; interviewee 3 stated they "focused on presenting stories more in the beginning, and the data driven part [such as "digging stories"] is still a work in progress" and they needed to "grow news sense"; while interviewee 15 noted that "many local data journalists understand how to collect data and write about it, but visualising it as a coherent story rather than blocks of text interspersed with a bunch of charts is a challenge."

Most interviewees had received a degree in data journalism from the USA. They studied abroad because it was considered to provide the highest quality training, but also because there were no Chinese data journalism degrees until recently. There was a strong sense that Chinese data journalism practices were influenced by US traditions and training. However, interviewee 14 felt that "US training is not enough" because it fails to account for the peculiarities of the Chinese media system; that "the people [Chinese J-Schools] train up should be [a] better fit for the Chinese audience"; and that the prevalence of mobile-based consumption in China has also necessitated different training (see below). Several interviewees expressed hope that, with many Chinese universities now offering data journalism subjects, and some offering data journalism degrees, a distinct style of data journalism indigenous to China could develop (8, 11, 12, 14).

A key motivation for studying data journalism was to be competitive in the job market: people were "rushing [into data journalism] because this is a new area, [and] that means there's more chance" (interviewee 8). Before studying data journalism in the US, interviewee 2 worked as an unpaid journalism intern for 5 months as there were no paid jobs; studying data journalism was a way to avoid the "decline in the traditional industry". For some, there was also a perception that data journalism skills allowed them to "tell stories that other kinds of journalism cannot. It has its own unique specialities" that can "enrich journalism" (interviewee 3). Nevertheless, some lamented that data journalism graduates increasingly choose to work at "Alibaba or Tencent" and other tech companies that "are sucking up all of the advertising dollars...it's quite demoralising" (interviewee 11). Similar to journalists internationally, several interviewees expressed concern about the boundaries of the journalism industry in the face of competition from platforms, and anxieties about money, job stability and job satisfaction (interviewee 3, 4, 11, 13, 14, 18). Even senior journalists had shifted roles: "I couldn't get more resources for my data journalism as [my editor] didn't have data journalism as core business and data journalism was unable to make a profit" (interviewee 12). This situation has created a "bottleneck" with "more talents and technology and resources coming into the sector" but no "viable business model" (12), leading to fewer job opportunities and the increasing use of unpaid interns.

Internal training was generally limited for practicing data journalists as they were already the most skilled. Some drew on professional courses from universities or commercial companies. All of the data journalists considered "self-learning" as "super important" (interviewee 3). Interviewee 4, for example, noted that they continued to learn "after work or during work […] because it is a very practical area", while interviewee 1 used Google, YouTube, MOOCs, books, and also maintained a university email address that enabled them to access training. For interviewee 7 "most journalism students graduate with a decent knowledge of data" but "the best way to learn is on the job".

While the early push was to upskill beat reporters to normalize data journalism into everyday practice, there were mixed views on how successful this had been. Interviewee 1 felt that the skills of "normal reporters […] are not advancing" and interviewee 14 argued that rather than being normalised, data journalism was becoming "more and more specialized" and this was reflected by the prominence of team-based production. Indeed, there was often a sharp distinction made: data journalists were the only ones who did data journalism (interviewee 6, 2).

This speaks to a wider debate about the role and position of data specialists in newsrooms (Wright and Doyle 2019). The role of one individual data journalist, for example, was described as to introduce data and 'trends' to support other journalists' reporting. Although they sometimes pitched stories, their primary role was support (interviewee 5). Likewise, interview 4 said they were sometimes not credited for their work or listed as designers, and interviewee 2 noted some traditionalists "think we are not a journalist, just a designer." There were also complaints that "general journalists tend to want to use data to support their conclusion rather than allowing the data to speak for itself" (interviewee 7). Another, more structural factor, according to interviewee 6,was that "Portal Web […are not allowed to have journalists" focusing on "policies or social issues" and so "I wouldn't call myself a data journalist".

The Motivation for Data Journalism in China

The previous section has outlined some of the personal motivations for becoming data journalists. In this section, motivations at the newsroom and journalism level are considered. The question of why Chinese news organisations - particularly those backed by the Party - invest in and conduct data journalism is a multifaceted jigsaw puzzle that in some ways is reflective of journalism in China more broadly. As Zhao (2008) has documented in detail, since the 1990's the market has become an integral aspect of Chinese media, facilitating not only an emergent commercial sector, but also a significant recasting of party outlets that were increasingly expected to support themselves through the generation of revenue. This market orientation was influential in the uptake of data journalism. The original hope was that data journalism might attract or keep new audiences (interviewee 3, 6, 7, 12), and particularly paid subscribers (interviewee 1). However, this was not just simply to attract eyes - which for some Party media was, relatively speaking, less important - but that the 'factual' form of data journalism might have more of an influence on the audience and that they might appeal to party elites, and attract investment, through reportage on issues such as urban city development data (interviewee 2, 4).[iii]

One institutional motivation was to be "state of the art" and "to find stories, not just make a regular story look better", enabling new forms of interactive, exploratory journalism (interviewees 1, 2, 3). Interviewees said it was important that Chinese newsrooms keep abreast of the latest developments and trends and thus experimenting with data journalism was, at least for a few years, common (2, 4, 5, 8, 12, 13, 14, 15, 16). Interviewee 14 argued that while the original motivation for data journalism amongst senior editors was "PR […] to "show to the world that they are ahead of the curve, they are doing something that's really attractive", this had subsequently changed. Interviewee 8 argued that party media have to "explore any new formats that *might help* their communication campaigns". Whether the "trend" is data journalism or short videos, "they cannot afford to lose this battlefield" and, in the big picture, for these outlets the "investment is actually not very big".

A related motivation for some was to garner awards (interviewee 1, 2, 3, 4, 7). Interview 3 talked at some length about the importance of awards because while applying for awards was harder in other areas, "for data journalism, we can really talk with the international community" which can "help data journalism teams to grow" and can "give back to leadership". As one commercially-based editor explained: "Industry recognition by our peers helps immensely" because data journalism is "intensive and time consuming", leading to lower story output. While page views were "good", and continued to grow over time ("evergreen") and time spent on page was above average, "given the hours and manpower that goes into them, they are something of a luxury. Awards help justify their value" (interviewee 7). The importance of awards is that they deliver bankable kudos, a finding also repeated internationally (Wright and Doyle 2019). However, another commercially-based editor stated that "Awards are fine, but they are not important to us" (interviewee 6) and this position was also reflected by some party media (interviewee 9).

Challenges for Data Journalism in China

There are a series of (often interrelated) challenges that were thought to impact both the practice and popularity of data journalism in China. Some of these were perceived to be specific to the Chinese media system, while others were exacerbated by it.

The Audience for Data Journalism in China

One initial motivation for data journalism was to "attract users on mobile", "drive traffic", "renovate journalism" and, particularly for online news, to differentiate their reporting from print newspapers (interviewee 3). While one party outlet reported that their data journalism was "unexpected[ly] popular [sic] and ranked second in traffic rank in 2018" (interviewee 9), the general view was that this had not happened and data journalism struggles to attract an audience - or at least that this was "cooling" because, as interviewee 15 put it, they "gradually found that there is no essential difference between data journalism and traditional news stories." (interviewee 15).

The relative lack of audience interest has made it difficult to justify the resource costs involved. The following exchange between a journalist and an editor during a joint interview at a party outlet captures this concern:

> Journalist: Data journalism is more attractive within this industry, not to our audience. Because I've talked with friends who worked in the US, who work in this area, and why they are doing data journalism is because it attracts traffic and people like it. But for us, no, I…
>
> Editor: They don't like it! (laughs)
>
> Journalist: It's quite tricky that people are not interested in data journalism, yes, our audience is not interested in it, but I think the people who work in this area, they are interested in it.

They stated that while they "care about how many people read our articles" and recounted some successes, they generally "don't have very good numbers on page views compared to other beats". Nevertheless, they continued, "it doesn't bother us because we don't have KPIs. For us, it's experimental, so we don't have that pressure to attract traffic". Importantly, they also noted that because they were:

> the official paper … the first class or the first part of our audience is the government, the governors in [our region] or in China […] how they think about our articles is very, very important to us. So we need to produce some product to make the official happy – to focus on the achievements of the government. So that is a very big part of the audience. *Then* we [care] about the hot topic, what the public cares about.

The imperative for even party-based media to generate revenue is well documented, contributing to the adoption of market-based metrics. Here, the positioning of consumers as a secondary market behind the 'first part' of their audience, namely officials, is indicative of how China's media system has both insulated and shaped data journalism. All but one of the data journalists working at party media reported *not* having KPIs and that they were given freedom and time to do their work, and the main goal was to do "great work" that enhances the "reputation" of their newspaper (interviewee 1, 18), while interviewee 3 noted that they were given "a lot of freedom and space to learn and progress. They didn't just ask for traffic, or money, or impact "because they knew that not only on the technology side, but also on the data analysis side, we need to have time to grow" and so even though "less than 20% of our traffic comes from desktop" it "doesn't really guide us that much" and they continue to focus on larger projects that worked better on desktops (interviewee 3). While there might not be KPIs, the lack of audience interest "in long-term and in-depth journalism … makes me very depressed", though interestingly interviewee 18 argued that COVID-19 had rejuvenated data journalism in China.

One interviewee who has experience of both party and commercial media sounded a note of caution about the apparent insulation from market pressures amongst party media: data journalism is still "burning money [and] the big challenge for them is to figure out how they can sustain themselves" because "these kind of [Party] subsidies is [sic] not endless" and they might "end or limit the input of that money" if they do not see returns (interviewee 12). There has already been a shift in focus to short videos among party media - including using such videos to communicate data journalism with a view to boosting clicks (interviewee 1, 2, 3, 6, 14, 18). Yet, with their key audience being party officials rather than the public, interviewees 2 and 4 were focusing on urban development, a tactic also successfully used by Initium Media's Rising Data Lab: "If we focused more about the urban development in [region] we could continue to survive in the newspaper and maybe we could get more resource, money or other things from the government", while interviewee 18 noted that: "Of course, sometimes we have the reporting assignments from party officials."

The situation is arguably starker for data journalists working in commercial media, whose KPIs including "page views / subscribers / hot articles / commercialization", were "very important, And it affects [...] how [the outlet] will be in the future." (interviewee 6). Failure to meet these had already led to staff cuts. This journalist reiterated the views of the editor and journalist at the party outlet: "To be honest, data journalism is not popular in China, [the] general audience have no clue what that is" (interviewee 6).

Mobile Phones and Social Media

One of the key influences on data journalism practice around the world has been the rise of mobile phones and social media for news consumption and sharing, and it is routine for data journalists to produce a desktop and a mobile-friendly version of a story (Wright and Doyle 2019). In China, the journalists believed that this was even more extreme because the vast majority of news is consumed on mobile, and often people only use mobile. This follows the

remarkably rapid diffusion of mobile phones in China, a phenomenon that has been supported and extended by the popularity of mobile apps - in particular, the virtual ubiquity of WeChat, which has rapidly come to play a central role in social life, commercial transactions and everyday governance (Harwit 2017, Huang and Zhang 2017). The impact of mobile was mentioned by every data journalist. Interviewee 1 explained that: "After I came back to China [from the US] the first thing is mobile first. It is very important […] not saying I need to change the content, but I need to change the design of the story" and "before we publish a project we have to test it as many times as we can on WeChat". "Almost all" consumers, they added, "just read it in the smartphones", so the mobile version will "shape how much people will pay attention to it". In particular, readers struggled with scrolling, and reading "1000 words … feels like they are reading 5000". This had clear implications for the nature of data journalism stories, as even for "long form projects […] it doesn't need to be long."

Interviewee 14 noted that while in the West it is now sometimes "mobile first...[i]n China, in many cases this is mobile only". Likewise, "if elsewhere in the world you can still do two different versions [one for mobile, one for desktop], in China you can only think [about] what works for mobile. And if it doesn't work, you kill the story". Interviewee 3 noted that WeChat exacerbated a shift to mobile first and mobile only data reporting, with news aggregation app, Toutiao (interviewee 6) and Weibo also cited (interview 18).

It was noted that designers "really hate mobile" because "the width is so limited" and it "impacts the design of data journalism projects very much" so "we have to come up with ways to use that platform" (interviewee 3). Interviewee 14 felt China was leading the way globally on designing data journalism for mobile, but compromises in functionality were still "quite a problem [and] I think this is why in China data journalism is not very popular." More specifically, interviewee 9 stated that "Screen size determines our layout" and interviewee 14 noted that "in order to optimise for mobile we need to sacrifice some of the possibilities". Similarly, interviewee 18 had to "re-edit and redesign my long-term journalist stories, cutting them into pieces", with page loading time also a concern. In response, the trend has been to "Make it simple" (interviewee 9), do less interactives, and "Even for a very big project" they sometimes just "present it as articles […] 'cause it's easier for the users to read" and they are now "very cautious about starting an interactive project" (interviewee 3), while commercial outlets had stopped doing interactives completely because "nobody sees it" (interviewee 6).

Another factor here was the rise of 'self-media', namely freelance content published by individuals through apps such as WeChat and Toutiao. These have "lowered the threshold for content production, so […] we need to choose topics that are more interesting and more concerned" (interviewee 15). Similarly, interviewee 9 stated that for "WeChat, we have to adjust our content to make it more down to earth", while interviewee 6 stated that "sadly, the media environment is getting worse due to censorship and the rise of self-media".

Data Literacy

Several data journalists were concerned that Chinese audiences struggled to comprehend data visualisations. Interviewees 4 and 6 noted that while education in China had improved fast, data literacy was relatively limited and they "always got feedback about not understanding the [complex] chart" (6), that "the data story is too difficult for most of the audience" (4), or that data visualisations were too "tough, cold and complex to understand" (18). For interviewees 9 and 10 the goal was simplification; interviewee 5 tried "to limit [it] to one message in each graph" for clarity; interviewee 1 was keen to expand their audience and so had to make [visualisations] at the level for the mass audiences" but "data literacy had

improved"; and interviewee 7 argued that it is "less about readers' literacy than about our ability to tell stories with data. If a general reader cannot understand one of our graphics then we have failed at our job." Nevertheless, interviewee 3, who did not have explicit audience metric targets, argued that this could amount to a form of dumbing down that could undermine the value of data journalism: "I think a lot of the times we cannot just present it as the most... the simplest version, 'cause otherwise it doesn't mean anything. It changes the meaning of the story." This reflects a tension between the technical possibilities of data journalism, and the realities of audience consumption preferences - particularly when driven by social media and aggregators and consumed on mobile phones.

Data Access and Quality

Access to data was widely cited as a challenge by Chinese data journalists, though it had improved significantly (10, 16, 17, 18), leading interviewee 9 to state that it "is easy now in China. 99% can be found in 'data.stats.gov.cn'". Nevertheless, most suggested there are still many issues, including "regional disparities" (18) and that "[c]ollecting the data is always the hardest part" (interviewee 17).

Firstly, "a lot of [government] data is not fully disclosed" and sometimes "data is missing" (5, 14, 17), "is not well organised or structured" or not available for download, so "80% of my time is spent in data mining and data cleaning" (2) and the "limited data we can work with" is compounded by the "limited topics that we can write about" (6). No journalist directly questioned the accuracy of government data sources. However, interviewee 11 noted that much government data was "aggregate percentages rather than individual level or any kind of micro level data. So it is not useful from a data journalism perspective" and that this is reflected in some Chinese data journalism, which is limited to "bar charts, percentages" or "graphic illustrations of percentages [...] and this isn't what I would consider data journalism." Interviewee 17, likewise, noted that the focus for some outlets was image and "form [rather] than the data itself. In many assignments, we are still asked to 'put figures into visually appealing graphics' even when the data does not make much sense".

Secondly, there were concerns about using commercial data. Interviewee 2 stated that "we don't trust their data, they won't give us their full data", and interviewee 3 noted that

> we are really cautious about working with companies to get their data because [...] they usually communicate [...] through their PR department, and they have very strong agenda setting. They are really cautious about their data, and they don't give us the raw data. It's actually no better than negotiating with government agencies.

Another concern was that businesses were basically using their data to gain favourable coverage, and data was being used as a form of power. Interviewee 12 noted that commercial companies came to them about stories and they were "confused about the purpose or motives of these companies. Does [the company have] some real demands on data journalism, or just want to use [data journalism] as a tool to ingratiate themselves with [redacted] as a media company for some favourable media coverage?" This was considered to have "muddied the water". They ultimately decided to stop working for the media and set up a kind of native advertising or integrated marketing company that creates data journalism content for commercial companies. A related concern here is the potential for bribery: "as a business reporter [...] half of their salary is the red packets" – cash given in envelopes by actors that seek to influence journalists (interviewee 11).

While there were challenges, data journalists used several strategies to get around restrictions and gain access to data. Interviewee 5 stated that: "a lot of times I need to go to [the local] library to read through the physical books, the year book, they have a lot of good data in there but they are not electronic [...] So I need to type in" the data. Others suggested that data journalism was an area where investigative work could be conducted that would otherwise be difficult. Interviewee 4, for example, stated that the government: "put their data on their website - they don't know you can get the data by scraping it. They don't know it, so we don't tell them!"

The Question of Censorship

The nature and impact of censorship has become a key debate within Chinese media studies. The impact of censorship on data journalism received some fascinating, divergent responses. While all of the journalists reported that censorship impacted what they could write about, there were different perspectives. Interviewee four had a fairly relaxed approach: "we would like to publish this story and see what happens. I think the worst maybe is they would put it down." Interviewee 2 cited an example where they linked data on a kind of public order violation to some officials, but were told to drop the story. Most journalists argued that data journalism did not afford them more freedom (e.g. 6, 9, 12, 15, 18). As interviewee 15 put it, "The freedom of reporting is not determined by the way the story is presented and excavated, but by the right of the news organization", while interviewee 6 responded to a question about whether there was more freedom with a blunt "no". The realities of "limited choice in topics" and "red line[s]" initially made it hard for people who had studied or interned abroad (interview 18).

Despite this, there was some evidence that, on occasion, data journalism did expand journalists' freedom. Interviewee 1, for example, argued that sometimes data journalism

> can find some things, and you can see some trends - maybe this is a little bit sensitive. Sometimes, you put it there, and leave it alone in your reports. And maybe this is a very obvious point, you should write down as texts. But if you write down and it's very sensitive for the report, it can be 404 [censored]...But if you find the data trend and you just put it in the graphics, the point is very obvious, and you just leave it there. Maybe readers can read it and understand, can get your point, and these reports can be safe.

In such responses, it is possible to discern the influence of what we have elsewhere discussed as a particularly Chinese mode of professionalism (Simons et al. 2017) that, while not informed by an overtly adversarial relation to government, nonetheless remains committed to a journalistic mission to ensure the public is informed. There was a strong sense of frustration - and some resistance - around the challenge of censorship from some data journalists, but for others it was more akin to resignation about the nature of the Chinese media system. A minority felt that data journalism afforded more freedom in that messages could be left in a chart, and there were active attempts to work around restrictions. On a practical level, however, it was considered possible to work within the constraints of the system to conduct data-driven investigative reports - where resources allowed - and they often spoke passionately about their own investigative reporting (1, 2, 3, 4, 18).

Conclusion

This article has found that data journalists in China experience pressures and challenges that, in some respects, are similar to other countries, but in other ways are either exacerbated in the

Chinese context or wholly different. This research broadly supports Appelgren et al.'s (2019) analysis that journalistic culture, market structures and political context are crucial contextual factors shaping the take up and practice of data journalism beyond liberal-democratic media systems (see also Mutsvairo 2019). As we have emphasised, however, it is the *interaction between* these factors that matters, and how the dynamics of such interaction, which also influences and is influenced by technological developments and affordances, plays out over time.

Interviewees generally argued that their Chinese audiences had relatively limited interest in data journalism – at least compared to Western contexts. This was explained by a variety of factors including data literacy, the near ubiquitous use of mobile phones, and the impact of social platforms such as WeChat and Toutiao. These were considered to be more important in China than the west and led to a push to simplify stories, minimise scrolling, limit interactivity and, for some commercial media, a focus on popular topics. However, some party media pushed back against these trends, investing significant resources into large, cutting edge interactive and narrative projects no matter the audience. The goal was to push the parameters of data journalism practice and conduct state-of-the-art projects that could – and occasionally did - win international awards – to please party officials.

Another set of challenges related to the restricted political environment. While not a uniquely Chinese issue (Tong and Zuo 2019: 7; Mutsvairo et al 2019) most journalists faced challenges with access to data, using creative methods to get around these challenges so that they could 'dig' and conduct investigative reports – "to dance with chains" as Du (2019: 120) puts it. Views on the impact of censorship were more mixed. Some journalists argued that they could avoid censors by placing messages in charts without explaining in the text (though the audience might miss them too). Others argued they could conduct data-driven investigative reporting within the constraints of the Chinese media system - though sometimes this required them to push boundaries. Journalists talked at length about their best data-driven investigative stories. As indicated by the awards, in spite of the constraints, Chinese data journalists have conducted many cutting-edge investigative reports.

Overall, however, the trend has been one of rationalisation for data journalism: both the number of teams and the size of them is reducing and there is a growing reliance on interns. This may help to explain Du's (2019) finding that the number of data journalism articles produced by Caixin had declined significantly. We also saw limited evidence of data normalisation in China - if anything it had become a more specialist activity. Whether data journalism can remain viable - at least in its current form - depends either on it finding a more effective business model, or on the largesse of the Party. While there is some limited evidence of data journalism offering new opportunities for publication, including investigative journalism, such possibilities remain circumscribed by the economic and political constraints faced by Chinese data journalists.

References

Appelgren, E., and G. Nygren. 2014. "Data Journalism in Sweden." *Digital Journalism* 2 (3): 394–405.

Appelgren, E., C. Lindén & A. van Dalen. 2019. "Data Journalism Research: Studying a Maturing Field across Journalistic Cultures, Media Markets and Political Environments". *Digital Journalism*, 7 (9): 1191-1199

Benson, R., Blach-Ørsten, M., Powers, M., Willig, I., & Zambrano, S. V. (2012). "Media systems online and off: Comparing the form of news in the United States, Denmark, and France." *Journal of Communication* 62 (1): 21-38.

Berners-Lee, T. 2010. "Analysing data is the future for journalists, says Tim Berners-Lee" *The Guardian.* 22 November 2010. https://www.theguardian.com/media/2010/nov/22/data-analysis-tim-berners-lee

Borges-Rey, E. 2016. "Unravelling Data Journalism." *Journalism Practice* 10 (7): 833-843.

Cheruiyot, D., S. Baack & R. Ferrer-Conill. 2019. "Data Journalism Beyond Legacy Media: The case of African and European Civic Technology Organizations". *Digital Journalism* 7 (9): 1215-1229.

Coddington, M. 2015. "Clarifying Journalism's Quantitative Turn." *Digital Journalism* 3 (3): 331–348.

Du, R. 2019. "Dancing with Chains: A case study of data journalism in China." In *Data Journalism in the Global South*, edited by B. Mutsvairo, B., S. Bebawi, and E. Borges-Rey, 109-123. Basingstoke.

Felle, T. 2016. "Digital Watchdogs? Data Reporting and the News Media's Traditional "Fourth Estate" Function." *Journalism* 17 (1): 85–96.

Fink, K., and C. W. Anderson. 2015. "Data Journalism in the United States: Beyond the 'Usual Suspects'." *Journalism Studies* 16 (4): 467–481.

Fink, K., and M. Schudson. 2014. "The Rise of Contextual Journalism, 1950s–2000s." *Journalism* 15 (1): 3–20.

Flew, T., C. Spurgeon, A. Daniel, and A. Swift. 2012. "The Promise of Computational Journalism." *Journalism Practice* 6 (2): 157–171.

Hallin, D. C., and P. Mancini. 2004. *Comparing Media Systems: Three Models of Media and Politics*. Cambridge: Cambridge University Press.

Harwit, E. (2017) "WeChat: social and political development of China's dominant messaging app." *Chinese Journal of Communication* 10 (3): 312-332.

Hawkins, J. M. 2017. Thematic Analysis. In *The Sage Encyclopedia of Communication Research Methods*, edited by M. Allen, 1757-1760. Thousand Oaks, CA: SAGE Publications.

Huang, H. & X. Zhang (2017) "The adoption and use of WeChat among middle-aged residents in urban China." *Chinese Journal of Communication* 10 (2): 134-156.

Knight, M. 2015. “Data journalism in the UK: a preliminary analysis of form and content.” *Journal of Media Practice* 16 (1): 55–72.

Lee, C.C. (2005) The conception of Chinese journalists: ideological convergence and contestation. In: *Making Journalists*, edited by H. De Burgh, 107-129. London: Routledge.

Mutsvairo, B., S. Bebawi and Borges-Rey 2020. *Data Journalism in the Global South.* Basingstoke: Palgrave.

Norman P. Lewis & E. Al Nashmi 2019. “Data Journalism in the Arab Region: Role Conflict Exposed.” *Digital Journalism* 7:9: 1200-1214.

Pan, Z., and Y. Lu 2003. Localizing professionalism: discursive practices in China’s media reforms. In: *Chinese Media, Global Contexts*, edited by C.C. Lee, 215-236. London: Routledge.

Parasie, S., and E. Dagiral. 2012. “Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago.” *New Media & Society* 15 (6): 853–871.

Mutsvairo, B. 2019. “Challenges Facing Development of Data Journalism in Non-Western Societies”. *Digital Journalism* 7 (9): 1289-1294,

Repnikova, M. 2017. “Media openings and political transitions: Glasnost versus Yulun Jiandu.” *Problems of Post-Communism* 64 (3-4): 141-151.

Rogers, S. 2012. “Anyone can do it. Data journalism is the new punk.” *The Guardian*. 24 May 2012: http://www.theguardian.com/news/datablog/2012/may/24/data-journalism-punk.

Rogers, S., J. Schwabish and D. Bowers. 2017. *Data Journalism in 2017: The current state and challenges facing the field today.* Accessed 23 June 2018. https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf.

Simons, M., D. Nolan and S. Wright. 2017. “’We are not North Korea’: Propaganda and professionalism in the People’s Republic of China”. *Media, Culture & Society* 39 (2): 219-237.

Stalph, F. 2018. “Classifying Data Journalism.” *Journalism Practice* 12 (10): 1332-1350.

Tandoc, E.C. and S-K. Oh. 2017. “Small Departures, Big Continuities?.”*Journalism Studies* 18 (8): 997–1015.

Tong, J. 2011. *Investigative Journalism in China*. New York: Continuum.

Tong, J. 2019. “The Taming of Critical Journalism in China: A combination of political, economic and technological forces.” *Journalism Studies* 20 (1): 79-96.

Tong, J. and Zuo, L. online first. “The Inapplicability of Objectivity: Understanding the Work of Data Journalism.” *Journalism Practice*.

Tunstall, J. 1971. *Journalists at Work: Specialist correspondents: Their news organizations, news sources, and competitor-colleagues*. London: Constable.

Wright, K., R. Zamith and S. Bebawi. 2019. “Data Journalism beyond Majority World Countries: Challenges and Opportunities”. *Digital Journalism* 7 (9): 1295-1302.

Wright, S. and K. Doyle. 2019. “The Evolution of Data Journalism: A case study of Australia.” Journalism Studies 20 (13): 1811-1827.

Young, M.L., A. Hermida and J. Fulda. 2018. “What Makes for Great Data Journalism?” *Journalism Practice* 12 (1): 1–21.

Zhang, S. & J. Feng 2019. “A Step Forward? Exploring the diffusion of data journalism as journalistic innovations in China”. *Journalism Studies* 20 (9): 1281-1300.

Zhao, Y. 2008. *Communication in China: Political economy, power, and conflict*. Rowman & Littlefield Publishers.

Zhao, Y. 2012. “Understanding China’s media system:. In: *Comparing Media Systems beyond the Western World*, edited by D. Hallin and P. Mancini, 143-174. Cambridge: Cambridge University Press.

[i] We followed up with three written interviewees to clarify points. These are not included in the word counts.
[ii] See, for example: http://caixinglobal.com/2018-06-08/caixin-wins-gen-2018-data-journalism-award-for-best-large-data-journalism-team-a-first-for-a-chinese-media-outlet-101267017.html#:~:text=Caixin%20VisLab%20was%20awarded%20the,field%20of%20data-driven%20journalism.&text=Caixin%20VisLab%20was%20the%20only%20Chinese%20team%20nominated
[iii] Responses suggest that urban data news is popular with editors as stories focus on largely positive news about the speed of China’s development.