



Reuters Institute Fellowship Paper

University of Oxford

Showing the way:

How big Pakistani media groups are leading the surge in data journalism

By Maleeha Hamid Siddiqui

April 2018

Michaelmas and Hilary terms

Sponsor: Thomson Reuters Foundation

TABLE OF CONTENTS

1. INTRODUCTION.....	4
2. METHODOLOGY	5
3. MEDIA INFLUENCERS.....	6
4. DATA JOURNALISM: A BACKGROUND.....	8
5. DATA JOURNALISM IN PAKISTAN.....	10
5.1 FINDINGS.....	11
5.2 CHALLENGES.....	21
5.3 RECOMMENDATIONS.....	25
5.4 FUTURE.....	27
6. APPENDIX.....	29

ACKNOWLEDGEMENTS

THIS research would not have been possible without the generous sponsorship of the Thomson Reuters Foundation, the Reuters Institute for the Study of Journalism and the University of Oxford. I would like to thank my employer, the Dawn Media Group for letting me experience life at Oxford University.

My utmost thanks to my supervisor Dr Chris Westcott for giving me enough time and space to formulate my ideas and offering his advice as and when needed. I will always cherish our discussions and sharing personal stories about journalism practices in our countries.

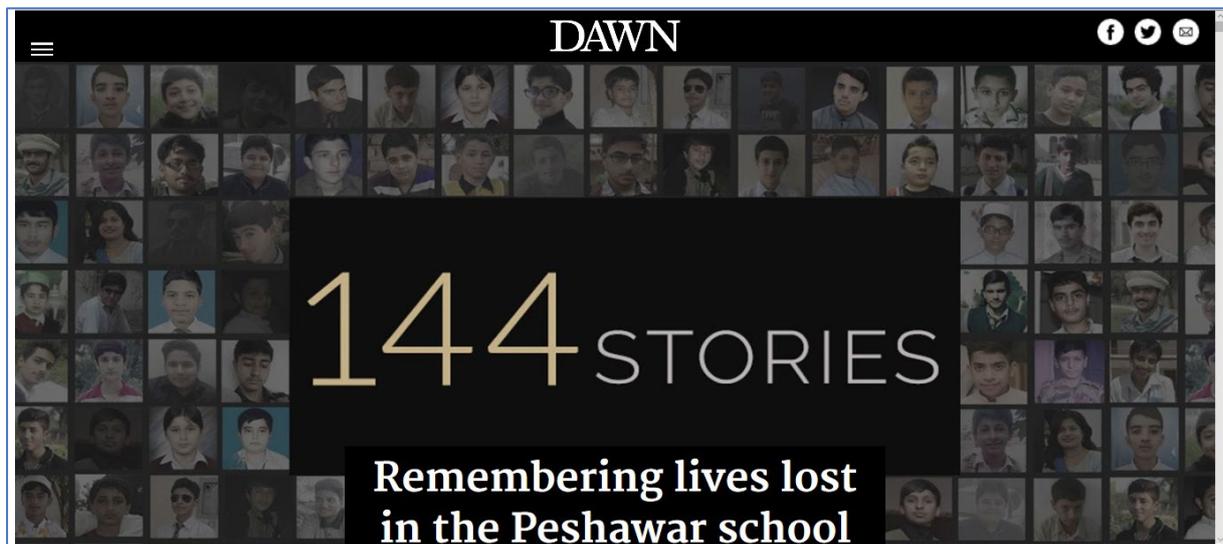
Many thanks to Meera Selva, director of the fellowship programme, David Levy, director of the Reuters Institute, and to the administrative staff of the institute, especially Louise Allcock for her kindness and humour.

Gratitude to my co-fellows, post-doctoral research fellows and past fellows for their camaraderie and humour. I would also like to acknowledge my friends in the UK who warmed my heart immensely especially during the long, cold winter. I am grateful for the support of my friends back home, they are too many to acknowledge, but my immense gratitude to Imran Ayub.

And finally, I will always be indebted to my family, especially my mother. I am extremely fortunate to have been born in a household where education is tremendously valued.

1. INTRODUCTION

The year 2015 is arguably a milestone year when data journalism came into its own in Pakistan. A year earlier the outlawed Tehrik-e-Taliban of Pakistan attacked an army-run school in Peshawar, killing 141 people, 132 of them children¹. To commemorate the anniversary of the brutal attack, the digital editions of the English language dailies created groundbreaking data-driven stories, as part of their larger narrative, employing up-to-date visualisations and data scraping tools. Off the top of my head I can recall “144 Stories”² and “A Year On – For Peshawar”³.



I have always been associated with traditional print media but was quite intrigued by this new form of storytelling. What I found even more incredible was the immediate response and feedback these stories received from readers.

This paper, therefore, is an opportunity for me and others to look at this new form of news narrative in Pakistan. It will primarily examine three questions. First, how are traditional big media organisations in Pakistan utilising data journalism to tell stories? Second, how are data journalists pushing the envelope, and third, what are the prospects for data journalism in Pakistan?

¹ Accessed at <http://www.bbc.co.uk/news/world-asia-30491435>

² Accessed at <https://www.dawn.com/news/1223313/144stories-remembering-lives-lost-in-the-peshawar-school-attack>

³ Accessed at <http://labs1.tribune.com.pk/for-peshawar/a-year-on/>

2. METHODOLOGY

To examine how journalists in Pakistan are utilising data journalism, the internal and external challenges they are encountering while writing data stories, the themes they have worked on, and whether there is a future for data journalism in Pakistan, I decided it was best to interview as many of those involved as possible.

Also, there is hardly any literature (books, reports, research papers) available on data journalism in Pakistan. This is not surprising as data journalism in the country is relatively new. Hence, it was only by interviewing those at the coalface that one could “create new knowledge”⁴ about data journalism in Pakistan.

In total I got in touch with 19 media professionals, international and Pakistan-based, via email, Twitter and Facebook Messenger from October 2017 onwards while I was in Oxford. Grouping them into three categories (heads or senior position holders in digital newsrooms, trainers and data journalists), I sent them queries accordingly. All 19 responded and 11 were subsequently interviewed via email, phone and WhatsApp from October 2017 to March 2018.

To examine the scope of subjects covered by data news and stories, I carried out content analysis. Around 50 pieces of output published from leading media organisations in Pakistan, India, the UK and the US were examined, out of which 28 stories from Pakistan were closely studied. The analysis period was from November 2016 until March 2018.

⁴ This phrase was planted in my mind by my journalist friend Aurangzaib Khan, actively involved in journalist training.

3. MEDIA INFLUENCERS

In order to understand the current news environment of Pakistan it is critical to be aware of major influencers⁵ that impact the workings of media, highlighting some of the perils of doing journalism in the country.

a) Army

Media groups and organisations in Pakistan have no choice but to be on the right side of the army. Otherwise, incurring the wrath of the army can be detrimental to their survival. For instance, in 2014, Geo TV of the Jang Media Group learnt the hard way when it triggered the Pakistan military's fury by airing accusations that the head of the army intelligence agency, the Inter-Services Intelligence (ISI), was allegedly behind the near fatal shooting of its star TV talk show host Hamid Mir. Cable TV providers were pressurised to drop Geo TV from their line up and advertisers, too, deserted the channel⁶.

The attack on Mir and the stern treatment meted out to Geo TV were warnings to other television channels and media groups to tone down the criticism of the military with respect to the controversial military courts⁷ and Balochistan⁸, a province gripped by an insurgency since 2006 when its former governor and chief minister Nawab Akbar Bugti was killed in a military operation after a prolonged conflict with then President Pervez Musharraf⁹. Taking the cue from Geo TV, Pakistani media has since turned a blind eye over the human rights abuses in the province¹⁰.

b) Judiciary

The adverse judicial approach to the freedom of expression in the media is increasing as underlined by Human Rights Watch (HRW) in 2012. According to the HRW World Report on Pakistan, both the Lahore and Islamabad high courts barred the media from criticising the judiciary or giving airtime to critics, through threats of contempt of court proceedings. This situation allegedly came about in the aftermath of a corruption scandal involving the son of then Supreme Court Chief Justice Iftikhar Muhammad Chaudhry¹¹. Thus, newspapers viewed the judiciary as an institution of coercion and censorship¹². This thin-skinned approach continues to persist. Earlier this year the Lahore High Court issued a written warning to the electronic media watchdog, Pakistan's Electronic Media Regulatory

⁵ The *Cambridge Dictionary* defines an influencer as one who changes the way that other people behave

⁶ "Pakistan press freedom under pressure from army" by Jon Boone, *The Guardian*, accessed at <https://www.theguardian.com/world/2015/sep/14/pakistan-press-freedom-army-journalists-military>

⁷ "Military courts get Supreme Court nod" by Nasir Iqbal, *Dawn*, accessed at <https://www.dawn.com/news/1198533>

⁸ "Living like a fugitive" by Idrees Ali, Dana Priest, *The Washington Post*, accessed at http://www.washingtonpost.com/sf/investigative/2015/07/25/living-like-a-fugitive/?utm_term=.d074b5652889

⁹ "Akbar Bugti's death and the revival of the Baloch insurgency" by Shah Meer Baloch, the *Herald* magazine, accessed at <https://herald.dawn.com/news/1153862>

¹⁰ "The simmering conflict in Balochistan" by Hafeez Jamali, the *Herald* magazine, accessed at <https://herald.dawn.com/news/1153776/the-simmering-conflict-in-balochistan>

¹¹ Accessed at <https://www.hrw.org/world-report/2013/country-chapters/pakistan>

¹² "Media quandary: Can judges be criticised?", *Express Tribune*, accessed at <https://tribune.com.pk/story/472132/media-quandary-can-judges-be-criticized/>

Authority (Pemra) for not stopping TV channels from broadcasting anti-judiciary speeches by disqualified Prime Minister Nawaz Sharif and his daughter, stating it was equivalent to ridiculing and scandalising the judges¹³.

c) Militant and separatist groups

On October 26, 2017 militants from a banned separatist group lobbed a hand grenade at a newspaper office and attacked a vehicle carrying newspapers in Turbat city, Balochistan province. Next day, a grenade was hurled at a press club in Balochistan¹⁴. During that period no newspapers were delivered in the province following calls to their offices by the group threatening that they would prevent the distribution of papers to prevent the publication of news about their activities¹⁵.

These incidents highlight the extent of terror in practising journalism in a country where journalists routinely face threats from militant and separatist groups all wanting to manipulate the media to further their agendas.

d) Government

The Pakistani government is no angel when it comes to ramping up the pressure on Pakistani news media. For instance, in November 2017, the government launched a crackdown to quell a large protest by a religious group at a road junction located between the capital Islamabad and Rawalpindi city¹⁶. As part of the crackdown, orders were given by then Prime Minister Shahid Khaqan Abbasi and the Islamabad administration¹⁷ to shut down TV transmissions and suspend social media platforms and mobile phone signals.

¹³ "Diatribes against judiciary: notices to Nawaz, Maryam", *Dawn*, accessed at <https://www.dawn.com/news/1382226>

¹⁴ "Eight injured in attack on newspaper agency office" in Turbat by Behram Baloch, *Dawn*, accessed at <https://www.dawn.com/news/1366538>

¹⁵ "Newspapers not delivered in Balochistan amid threats", *Dawn*, accessed at <https://www.dawn.com/news/1366137>

¹⁶ The name of the group is Tehreek-i-Labbaik Ya Rasool Allah, little known until it participated in a by-poll against wife of disqualified prime minister Nawaz Sharif, Kulsoom Nawaz and came in third.

¹⁷ "Pemra blocks all TV channels" by Sehrish Wasif, *Express Tribune*, accessed at <https://tribune.com.pk/story/1568146/1-media-blackout-faizabad-sit-pemras-orders/>

4. DATA JOURNALISM, A BACKGROUND

a) The concept of data journalism

Alexander Benjamin Howard in his report *The Art and Science of Data-Driven Journalism* states that data journalism is the application of data science to journalism, where data science is defined as the study of the extraction of knowledge from data. Howard also provides a broader definition which is that data journalism is finding stories in numbers. It is treating data as a source to augment reporting along with accounts of human witnesses, quotes from officials and experts, and visualisations to present it¹⁸.

Another description of data journalism is given by Melissa Bell of Vox Media. “Data journalism is, simply, journalism based off of data. Just as some stories are based off of conversations from people and others are based off of conversations from documents, there’s a space for stories based off of raw data. The result can take many forms: text article or a data visualisation or a video or something else entirely¹⁹.”

b) It is not a modern invention

Data journalists agree that using numbers and statistics have always been an integral part of reporting. An early instance of using data in journalism was an investigative exposé by *New York Tribune* editor Horace Greeley published on December 22, 1848. The story had a table in two columns which showed excess mileage paid to each member of the US Congress. The Congressmen were doing nothing illegal. There was a law which provided for mileage reimbursement which was computed by calculating the distance by the usually travelled route. However, Greeley argued with the help of the table that the law needed to be reformed as the mileage reimbursement was calculated decades earlier and now with the increasing prevalence of trains and steamships, travellers could travel much faster. Hence, the excess mileage was a waste of the taxpayer’s money that needed to be curbed²⁰.

c) How it became a buzz phrase

The use of the phrase “data journalism” gained momentum when World Wide Web inventor Tim Berners Lee, said “data journalism is the future” at the 2010 launch of the first UK government datasets to the general public for all spending by departments of over £25,000²¹. In 2014, data journalism entered mainstream discourse driven by the highly

¹⁸ Pages 4, 5 of *The Art and Science of Data-Driven Journalism*, accessed at <http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf>

¹⁹ “What is Data Journalism” by Melissa Bell for Vox Media, accessed at <https://www.vox.com/2015/2/4/7975535/what-is-data-journalism>

²⁰ “How Big Data busted Abe Lincoln” by Scott Klein, ProPublica, accessed at <https://www.propublica.org/nerds/antebellum-data-journalism-busted-abe-lincoln>

²¹ “Analysing data is the future for journalists”, accessed at <https://www.theguardian.com/media/2010/nov/22/data-analysis-tim-berners-lee>

publicised relaunch of Nate Silver's FiveThirtyEight.com website to provide coverage of politics, economics, sports, science and lifestyle²².

d) Why it matters

With continued international moves towards greater transparency, governments throughout the world, including Pakistan, are increasingly publishing information online. If we believe that journalists are watchdogs for democracy then data journalism has an important role to play. Data reporters can be effective when they hold the Establishment to account²³.

Secondly, data creates possibilities to tell important stories, allowing for far more complex investigations by using software to find connections. In the hands of advanced data journalism practitioners, algorithms can analyse trends, and the relationships of influence, power or sources²⁴ mapped.

And finally, data journalism is a new set of skills at a time when basic skills from traditional journalism aren't enough²⁵.

²² "Status update: Building FiveThirtyEight" by Nate Silver, accessed at <https://fivethirtyeight.com/features/status-update-building-fivethirtyeight/>

²³ Page 137 of *Data Journalism: Mapping the Future*

²⁴ Page 8 of *The Art and Science of Data-Driven Journalism* by Howard

²⁵ "Why is Data Journalism important?", *Data Journalism Handbook*, accessed at http://datajournalismhandbook.net/1.0/en/introduction_2.html

5. DATA JOURNALISM IN PAKISTAN

(Findings, challenges, recommendations & future)

Prior to researching the status of data journalism in Pakistan, I was under the impression that the media groups had fallen into a comfort zone. But I was pleasantly surprised to find that a body of work in data journalism has been built up in recent years. Not only that, I found that despite working under severe constraints²⁶, data journalists had written stories which pushed the envelope on the situation within the country.

In respect of data journalism Pakistan shares a more or less similar media landscape with Russia where, despite its poor ratings on indices of freedom of speech, democracy and transparency, data journalists are able to work within the parameters of what is, in fact, not such a closed country. According to Reuters Institute fellow Anastasia Valeeva's paper "Open Data in a Closed Political System: Open Data Investigative Journalism in Russia", Russia was ranked 35 out of 94 in the Open Global Data Index 2016. Conversely, in the Open Budget Index 2015, Russia ranked 10 out of 102, scoring better than Italy and Germany. Moreover, in terms of financial transparency, Valeeva wrote that Russia was competing with front runners such as the UK and the USA²⁷. By publishing open data in the public interest, some Russian journalists were able to write investigative data stories.

Another instance is China which has low rankings in press freedom lists. Yet, media companies backed by Chinese business tycoons are publishing stories that are a mixture of propaganda and investigative reports, some of which have national influence. The public-private funding partnership in China's media sector has created a space for editorial staff to negotiate and collaborate with government authorities on pushing reporting boundaries for mutual benefit²⁸.

²⁶ See Chapter 3, Media Influencers.

²⁷ Accessed at https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Valeeva_Open%20data%20and%20investigative%20journalism%20in%20Russia.pdf

²⁸ "Is censorship stifling China's media" by Vincent Ni, senior producer, BBC World Service, accessed at <https://reutersinstitute.politics.ox.ac.uk/risj-review/seminar-report-censorship-stifling-chinas-media>

5.1 FINDINGS

a) Large media groups leading the surge

The first and foremost key finding from my research is that in Pakistan it is the big news organisations, such as the Dawn Media Group, the Jang Group, and Express Group, that are spearheading data journalism, headed by their digital newsrooms.

“Some large media groups you have identified [Dawn, Jang and Express] are, actually, leading the surge in the use of data journalism. It is the digital teams that are taking the lead in data-driven news projects. Some of these teams are more visionary than others. It’s exciting to see these news organisations deliver projects that push the boundaries of traditional Pakistani journalism.”

— Waqas Naeem²⁹, programme manager, Media Matters for Democracy, Islamabad-based media research and advocacy non-profit.

This happened because, perhaps, there was recognition within the media groups that fundamental changes were necessary to cope with new and more challenging market environments³⁰. In the case of legacy media such as newspapers those environments include an increasing reliance on digital media, mobile devices and social media platform with intense competition for attention when accessing and finding news³¹.

b) Taking on sacred cows

Reporting on sacred cows (military, judiciary, government, militant groups) has a long history in Pakistan. It is why the country has arguably some of the most courageous journalists in the world³². This tradition of journalism continues with data-driven investigative stories. The most prominent was the Panama Papers investigation project in which more than 370 journalists collaborated from around the world, including Pakistani journalist Umar Cheema, who wrote the story for *The News* of the Jang Group. It was based on a 2.6 terabyte trove of data held by the Panama-based company Mossack Fonseca leaked to German newspaper *Sueddeutsche Zeitung* which was then shared with the International Consortium of Investigative Journalists (ICIJ).

The leaked documents, published on April 4, 2016, showed that three children of then prime minister of Pakistan, Nawaz Sharif, were linked to several offshore companies that owned properties in prime locations in London, and that the companies used the properties as collateral to secure a loan worth millions of pounds³³. This led to a probe into his family’s wealth and culminated in his disqualification as prime minister by the country’s Supreme

²⁹ Email interview on March 4, 2018

³⁰ “Leading Change: Why Transformation Efforts Fail” by John P. Kotter, *Harvard Business Review*.

³¹ “Challenges and opportunities for news media and journalism in an increasingly digital, mobile and social media environment” by the Reuters Institute for the Study of Journalism, 2016, accessed at <http://reutersinstitute.politics.ox.ac.uk/our-research/challenges-and-opportunities-news-media-and-journalism-increasingly-digital-mobile-and>

³² ‘The courage of Pakistani journalists’ accessed at <https://www.theatlantic.com/international/archive/2011/09/the-courage-of-pakistani-journalists/245358/>

³³ “Panama Papers: Pakistani PM Nawaz Sharif’s son defends offshore companies”, BBC, accessed at <http://www.bbc.co.uk/news/world-asia-35959895>

Court, leading to his immediate resignation on July 28, 2017. Cheema’s story led to one of the most significant political consequences of the Panama Papers data leak.

“Banned outfits in Pakistan operate freely on Facebook”³⁴ by *Dawn* in 2017 is another example of taking on sacred cows. Using Facebook search, the names of 41 banned sectarian, terrorist and anti-state organisations were searched in Facebook pages, groups and profiles, and it was discovered they were accessible to every user on the social media platform³⁵.



This investigative story uncovered the extent to which 41 of the 64 banned outfits have a presence on the platform disseminating their ideologies through more than 700 pages, groups and user profiles³⁶. In terms of reader engagement, the story was shared 55,000 times on Facebook³⁷ and generated over 300³⁸ comments. While it is difficult to gauge the story’s explicit impact, the findings uncovered a disturbing fact³⁹. Official authorities were allowing these groups to operate as they stoked hatred against people of other communities encouraging the murder of people on allegations of blasphemy, exposing the Pakistani state’s muddled priorities⁴⁰. But these same officials were hounding individuals for expressing views against the country’s armed forces on social media⁴¹.

“A Year On — For Peshawar, NAP in numbers”, Tribune Labs⁴² online multimedia project on the first anniversary of the Army Public School terrorist attack in Peshawar that killed 144 children and staff, and the subsequent formulation of a counter-terrorism

³⁴ Accessed at <https://www.dawn.com/news/1335561/>

³⁵ “Methodology”, accessed at <https://www.dawn.com/news/1335561/>

³⁶ Editorial accessed at <https://www.dawn.com/news/1336405>

³⁷ Figure obtained from the online site of the story.

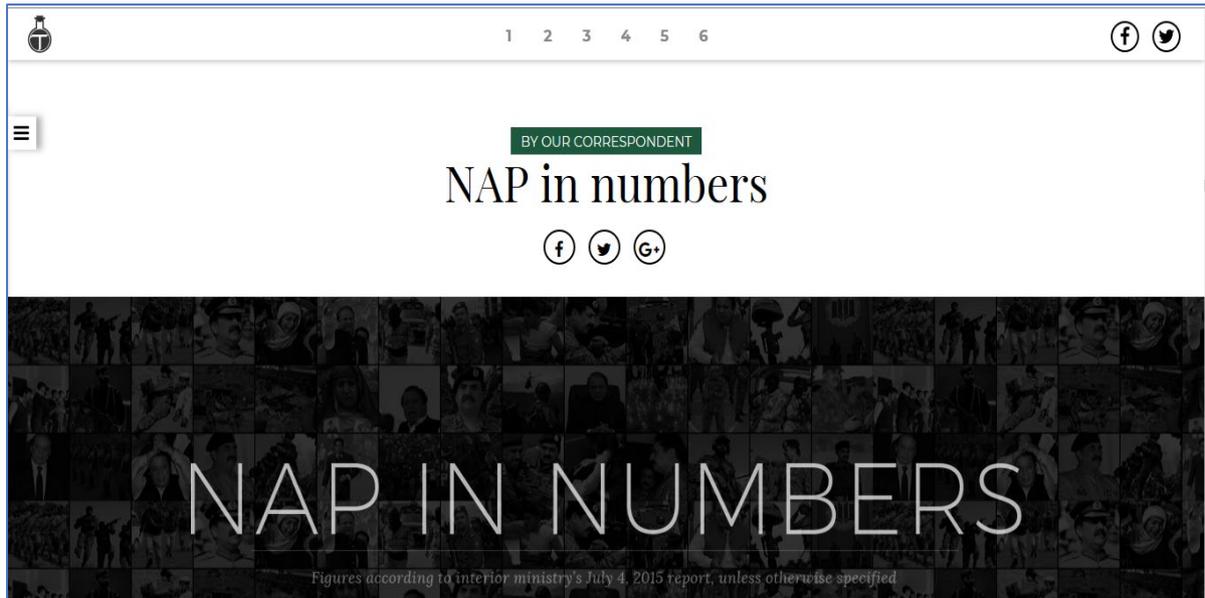
³⁸ Figure estimated from story posted on Dawn.com’s Facebook page.

³⁹ Some notice of the story was taken at the government level when according to a news report the Parliamentary Committee on National Security sought a briefing on the banned groups activities on Facebook based on Dawn’s story, accessed at <https://www.dawn.com/news/1336340>

⁴⁰ “Editorial: Banned groups on Facebook”, *Dawn*, accessed at <https://www.dawn.com/news/1336405>

⁴¹ “FIA launches crackdown on anti-army campaigners” by Amir Wasim and Zulqernain Tahir, *Dawn*, accessed at <https://www.dawn.com/news/1334626>

⁴² Digital media site of Express Tribune, Express Media Group.



policy called the National Action Plan (NAP), is a great example of utilising data as a part of a larger story and crucially, a powerful instance of accountability of military and government policies. Along with *Dawn's* similar data visualisation story, "144 Stories"⁴³, which was done three years later, it implicitly highlighted continued military and government counter-terrorism policy failures.

Former editor of Tribune Labs, Shayan Naveed⁴⁴, explained to me how they worked on the project:

"For Peshawar"⁴⁵, "was our biggest project, consisting of 17 stories, which can be quite overwhelming for the reader. We divided these stories into four sections. The fourth section is titled "A Year On" which is a review of the progress made on the government's 20-point National Action Plan. "NAP in Numbers" is the second story in the section. Reporter Azam Khan filed a 1,000-word story which was just data after data of the government's progress. We decided to make an infographic instead of a written story because a) it would be a break from all the other stories which were text based; b) More people were likely to read it if it's in the form of an infographic. In fact, it was one of the most commented on and shared pieces of the project. It includes information on how many prisoners were executed since the attack and how many of them were terrorists; how many suspects had been arrested since the attack; how many terrorists were killed in [Operation] Zarb-e-Azb since the attack; how many religious seminaries exist in Pakistan and how many are sealed and so on."

While the Geo TV interactive data story "Explainer: How 10 PML-F votes helped one MQM-P senator retain his seat" showed how rival candidates voted against party lines during the 2018 Senate elections, it was also an indirect critique of a political system.

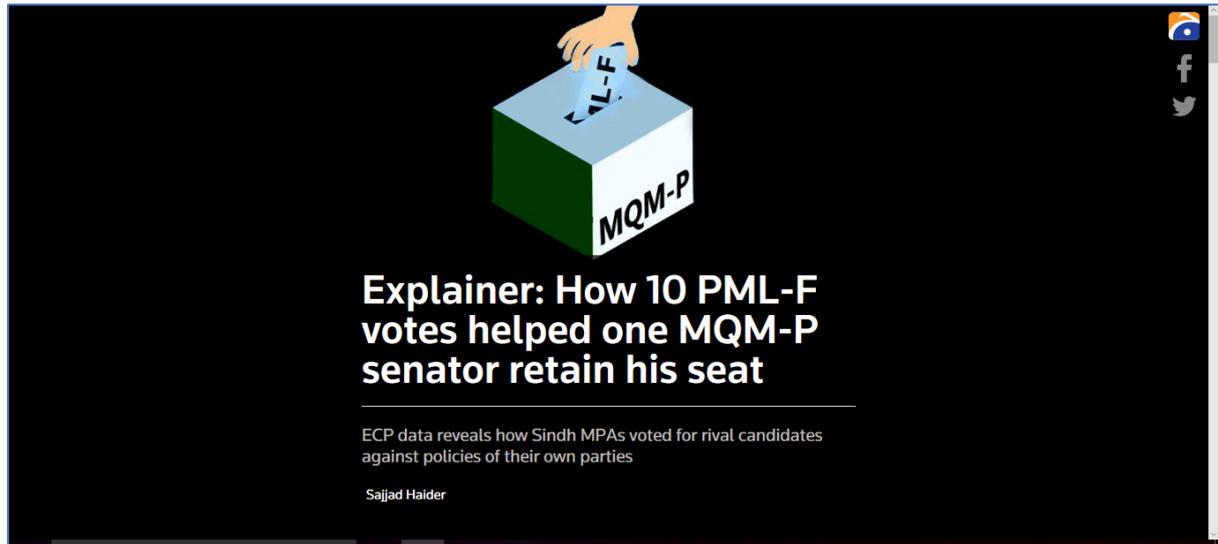
⁴³ The story can be accessed at <https://www.dawn.com/news/1223313>

⁴⁴ Email interview on November 22, 2017

⁴⁵ Accessed at <http://labs1.tribune.com.pk/for-peshawar/>

“There wasn’t a lot of data. The story is based on just a single page of numbers from the Election Commission of Pakistan [website] and is about how votes were transferred from one candidate to another for the Sindh Senate polls.”

— Sajjad Haider⁴⁶, writer of the data story for Geo TV of Jang group.



c) *Covering topics to rebuild trust in media*

Worldwide public trust in media is at an all-time low⁴⁷ and the situation in Pakistan is no different⁴⁸. One of the ways journalists are winning back trust is with the help of data journalism⁴⁹. Similarly in Pakistan, journalists are covering important subjects with the help of data to gain back the trust of the public.

“I see data journalism in the Pakistani context as an instrument to help news organisations build trust with the public. The local media, especially broadcast, are facing a crisis of credibility, which is not too dissimilar from the trends of media trustworthiness being recorded worldwide. People routinely accuse the Pakistani media of bias and sensationalism. Data journalism techniques can provide the media with the ability to report on issues of public importance factually and accurately, transcend the perception of bias and create a real impact on policy through their reporting.”

—Waqas Naeem

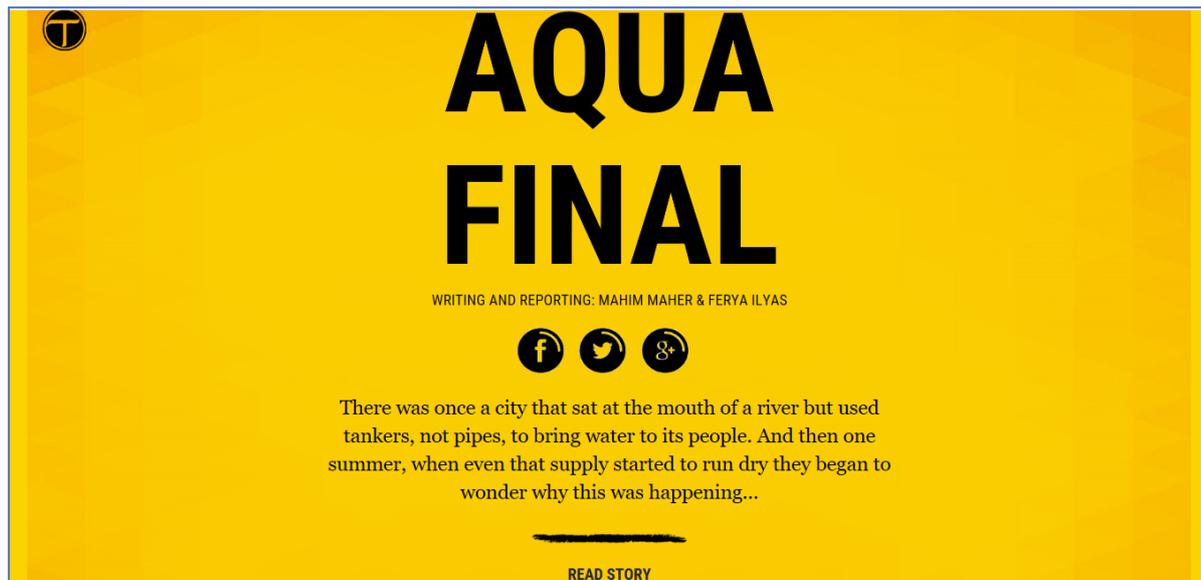
⁴⁶ Facebook Messenger interview on March 2, 2018

⁴⁷ ‘Trust the news? Most people don’t, social media even more suspect – study by Belinda Goldsmith, Reuters, accessed at <https://www.reuters.com/article/us-media-news-survey/trust-the-news-most-people-dont-social-media-even-more-suspect-study-idUSKBN19D015>

⁴⁸ ‘Examining Citizen’s Confidence in Institutions of Pakistan: An Analysis of Citizen’s Trust’ by Naila Yosuf and Binish Nauman, accessed at http://hrmars.com/hrmars_papers/Examining_Citizen%E2%80%99s_Confidence_in_Institutions_of_Pakistan_An_Analysis_of_Citizen%E2%80%99s_Trust.pdf

⁴⁹ <https://www.poynter.org/news/study-shows-data-reporting-gaining-hold-newsrooms>

One way to win public trust in journalism is by covering important but less well-known subjects,⁵⁰ for instance, water scarcity. According to a report by the World Resources Institute, Pakistan is on track to become the most water-stressed country in the region and 23rd in the world by 2040⁵¹. “Aqua Final”⁵² by Tribune Labs is a notable example on the subject which was written during a heatwave in Karachi when many parts of the city were without a water supply. This led to protests and a breakdown in law and order.



Mahim Maher⁵³, one of the writers of the story, explained to me how it was produced:

“My editor, Kamal Siddiqi, told me to work on the water crisis that took place in Karachi that summer in 2014. He wanted me to find out what had caused it. It took me three months of constant daily work to get this story done. Ferya Ilyas worked with me on getting all the moving parts together. I did not start out with the idea of looking for data but at the back of my head I’m always looking for data to support the story. The main data in this story came from the documents submitted to the Supreme Court on the number of FIRs [First Information Reports]⁵⁴ or police cases against illegal hydrants and water theft.

I took great pleasure in working with the younger more tech-savvy journalists who took an immense burden off me. For if I were going to focus on the content, they needed to pick up the slack on the data and visualisation.”

⁵⁰ Why data journalism is important? Page 139 of *Data Journalism Mapping the Future*

⁵¹ Water insecurity by Neil Buhne, accessed at <https://www.dawn.com/news/1370550>

⁵² Accessed at <http://labs1.tribune.com.pk/aqua-final/>

⁵³ Email interview on February 5, 2018

⁵⁴ First Information Report is a complaint lodged with the police by the victim of a cognizable offence or by someone on his/her behalf, First Information Report (FIR) A Guide for Citizens, accessed at <http://www.shehri.org/FIR.pdf>

“Pregnant & fired: a Pakistani woman’s workplace dilemma”⁵⁵ by *Dawn* highlighted discrimination against working mothers, married women and maternal policies in medium- and small-sized organisations, an issue rarely brought up in newspapers.



d) *Investigative reporting*

Data journalism has become an integral part of a global strategy to support investigative journalism that holds the most powerful institutions and entities in the world accountable⁵⁶. Some great examples are “Legal Drugs, Deadly Outcomes”⁵⁷ about the epidemic in deaths from prescription drugs from the *Los Angeles Times* and “Toxic Waters”⁵⁸, a series by The Upshot, *The New York Times*, on the increasing pollution of waters in the US.

Pakistani news organisations, also, value investigative reporting that holds powerful institutions accountable for their actions and policies.

“Regardless of the back-breaking work pressures of daily print reporting and the even more gruelling 24/7 broadcast news cycle, most Pakistani news outlets still revere the idea of an investigative story that shows truth to power. Data makes it much easier for journalists to work on investigative news reports.”

—Waqas Naeem

One excellent example is the months-long investigative story by *Dawn*, “Rao Anwar and the killing fields of Karachi”⁵⁹. This report asserted that a police officer, Rao Anwar, was

⁵⁵ Accessed at <http://labs1.tribune.com.pk/aqua-final/>

⁵⁶ Page 14 of *The Art and Science of Data-Driven Journalism* accessed at <https://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf>

⁵⁷ Accessed at <http://graphics.latimes.com/prescription-drugs-part-one/>

⁵⁸ Accessed at <https://www.nytimes.com/interactive/projects/toxic-waters/index.html>

⁵⁹ Accessed at <https://www.dawn.com/news/1389761>

notorious for staging 'encounters'⁶⁰ and drew attention to the failure of law enforcement agencies to arrest him for a recent alleged extrajudicial killing of an ethnic Pashtun young man in January 2018⁶¹. The killing received nationwide attention and *Dawn's* story, which appeared on February 16, 2018, built pressure to arrest Anwar. The outcome was that he was finally arrested on March 21, 2018 after eluding law enforcement agencies for over a month⁶².



e) *Utilising data scraping software & data visualisation programs*

When Alexander Benjamin Howard wrote the report *The Art and Science of Data-Driven Journalism* for the Columbia Journalism School in 2014, he highlighted how journalists from around the world at the time were telling stories by tapping into vast amounts of data with the help of free online tools and software to collect, clean and publish data in interactive features. "As data journalists grow in skill and craft, they move from using basic statistics in their reporting to working in spreadsheets to more complex data analysis and visualisations, finally arriving at computational journalism⁶³."

Except for computational journalism, it was exciting and heartening to find stories from Pakistan that not only utilised basic statistics but spreadsheets, interactive features, infographics and complex visualisations.

"Fortunately, the cadre of Pakistani digital and multimedia journalists that's been building up its ranks over the past five years or so is attuned to the global trends in new media and digital journalism. They

⁶⁰ Encounter in the South Asian context is a violent incident in which a suspected criminal is killed by a member of the police, especially under controversial circumstances, definition in Oxford Dictionaries, accessed at <https://en.oxforddictionaries.com/definition/encounter>

⁶¹ "Pakistan police killing of a Pashtun youth fuels anger over 'encounters'" by Syed Raza Hasan for Reuters, accessed at <https://uk.reuters.com/article/uk-pakistan-killing/pakistan-police-killing-of-a-pashtun-youth-fuels-anger-over-encounters-idUKKBN1FG05C>

⁶² "Rao Anwar arrested after finally appearing before SC in Naqeebullah murder case", *Dawn*, accessed at <https://www.dawn.com/news/1396657>

⁶³ Page 8 of *The Art and Science of Data-Driven Journalism*, accessed at <https://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf>

read up on international best practices and challenges; they are experimenting with new data visualisation tools as well as immersive storytelling experiences with the help of new equipment, coding and web development.”

—Waqas Naeem

Here arguably are the best examples of data visualisations and interactive features produced by major media companies in Pakistan, thus, setting an example for others to emulate:

❖ ‘WHY TERRORISTS ATTACK EDUCATION’

“This story [“Why terrorists attack education”, 2015⁶⁴, see image below] is about the number of attacks on educational institutions across Pakistan. We used a map comparing Pakistan in this regard to other countries. Data indicated that Pakistan has suffered the most fatalities in attacks on schools. We used the information and data to make an interactive map and a heat map showing every single attack on an educational institution in Pakistan. The story also included a chart listing the terrorist organisations responsible for the attacks, the kind of attacks and a separate map for attacks conducted by the TTP [Tehrik-i-Taliban Pakistan].”

— Shayan Naveed, former editor of The Tribune Labs, Express Tribune

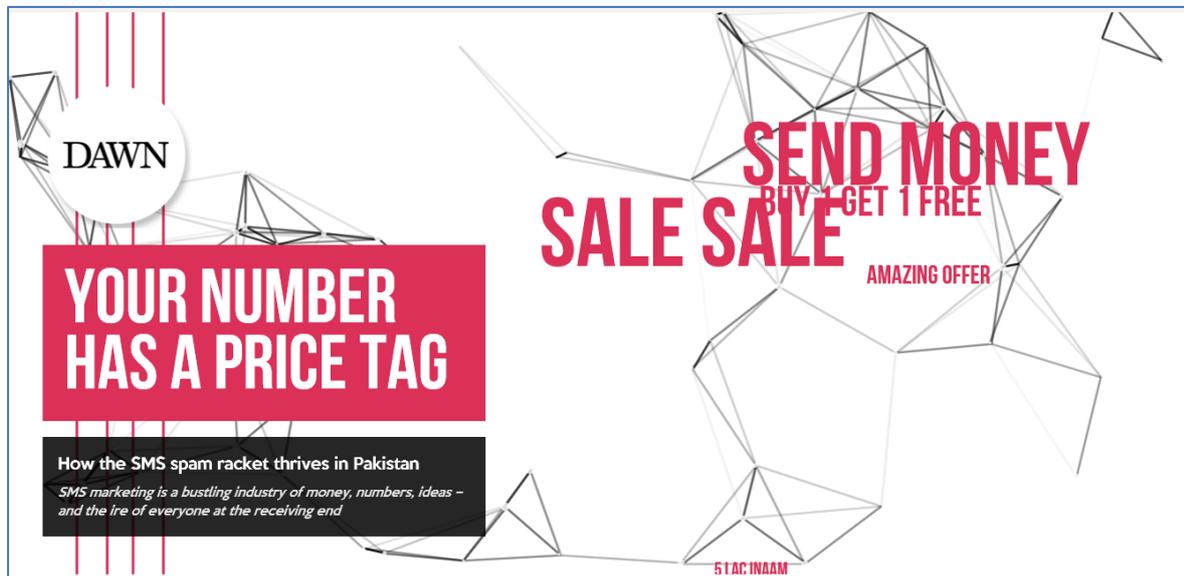


❖ ‘YOUR NUMBER HAS A PRICE TAG’

Dawn teamed up with International Centre for Journalists (ICFJ) fellow, Shaheryar Popalzai, in 2017 to produce an interactive data-driven investigative feature “Your number has a price tag” which is about the spam marketing text messages industry concealed under the term ‘SMS marketing’. Utilising data from a survey taken by 2,057 visitors to *Dawn*’s website, it showed nearly half the respondents received three to six spam messages daily from

⁶⁴ Accessed at <http://labs1.tribune.com.pk/attack-on-education/>

unknown sources. It features audio interviews with SMS marketers in Urdu along with English translations, pie charts and infographics⁶⁵.



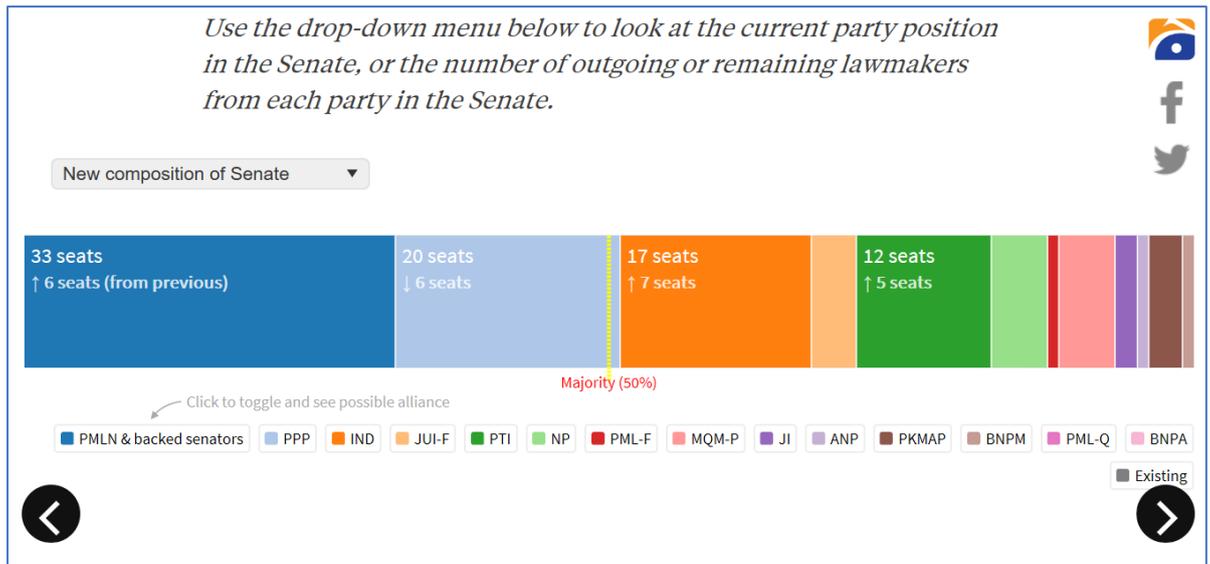
❖ '2018 SENATE ELECTIONS: WHAT'S UP FOR GRABS?'

For the "2018 Senate elections: What's up for grabs?"⁶⁶ by Geo TV, Sajjad Haider told me what web-based tools he had used for data collection, manipulation, analysis and visualisation:

"The Election Commission of Pakistan released data for the candidates in the form of image-based PDFs which meant the data had to be OCRed. So, I used Tesseract for that. I, then cleaned and manipulated the data using Excel/R. I visualised the story with the help of d3js-based visualisation running on top of Flourish.studio [free data visualisation site]. I also used Infogram and Canva for making infographics."

⁶⁵ <https://www.dawn.com/in-depth/spam-sms/#/>

⁶⁶ Accessed at <https://www.geo.tv/latest/183840-2018-senate-elections-whats-up-for-grabs>



f) Producing long-form data journalism

The last couple of years have seen a boom in long-form stories on the web produced by news media companies such as *The New York Times*, *the Guardian* and *Forbes*. Pakistani journalists and news organisations have also embraced this narrative format in the digital space and produced in-depth pieces such as:

- ❖ “Aqua Final” (mentioned earlier) had two versions. Short-form for the print edition with 1,500 words, and long-form for the online edition with 8,700 words⁶⁷,
- ❖ “PIA’s seventy-year-long descent”⁶⁸ by Tribune Labs is a 10,000-word story chronicling the decline of Pakistan’s national airline,
- ❖ And, “Pakistan’s step-children”⁶⁹ by *Dawn* is a 4,000-word story on the forced repatriation of Afghan refugees.



⁶⁷ Accessed at <http://labs1.tribune.com.pk/aqua-final/>

⁶⁸ Accessed at <http://labs1.tribune.com.pk/pia-descent/>

⁶⁹ Accessed at <https://www.dawn.com/in-depth/afghan-refugees/>

5.2 CHALLENGES

WHILE data journalism has improved the means and shown innovative ways to tell a story and is becoming an integral part of digital journalism in Pakistan, interviews with data journalists and trainers highlighted some of the obstacles to be removed for data journalism to fully thrive.

a) *Language*

Pakistan is quite diverse linguistically. According to the 1998 census, Punjabi is the mother tongue of 44 per cent of the population followed by other languages such as Pashto, Sindhi, Seraiki, Balochi and Urdu⁷⁰. Data journalism practitioners in English realise they are not reaching out to a large segment of the population because of the language barrier.

“All Labs stories are in English. Ideally, I would have liked all of them to be translated in Urdu and published on our Express Urdu website. However, most of our stories on Labs, especially data stories, were developed from scratch. Translating them in Urdu and presenting the same visuals would have taken twice the time.”

— Shayan Naveed

It is important to point out here, in Pakistan, the role of English is especially complex. Because of the language policy during British colonial history when English became the language of power and prestige, it still held the same position post-partition. It is the language of government, military and higher education⁷¹. It also plays an important gate-

⁷⁰ “Profile of Pakistan”, Chapter 2, *Language and Education in Pakistan: Recommendations for policy and practice* by Hywel Coleman and Tony Capstick, 2012, British Council, accessed at <https://www.teachingenglish.org.uk/sites/teacheng/files/Language%20In%20Education%20in%20Pakistan.pdf>

⁷¹ *Language and Education in Pakistan: Recommendations for policy and practice* by Hywel Coleman and Tony Capstick, 2012

keeping role i.e. restricting access to higher education, desirable fields of employment and high social status⁷².

Besides a language barrier, data journalism in Urdu and regional languages is compounded by the management styles in their newsrooms and by old school thinking.

“The Urdu and regional language press are far behind the English print and digital news publications in the use of data journalism techniques. This is due to the style of Urdu and regional language print journalism. These papers do not usually carry graphics and the word count they allocate to each story is often too limited for a good-quality data journalism treatment. Also, based on conversations with journalists, the people who are in-charge of the editorial desks in Urdu newspapers and sometimes broadcast newsrooms are quite resistant to the idea of change. They want to stick to their style. The few examples of Urdu data stories on the web are also from newsrooms that have good English-language web teams.”

—Waqas Naeem

This is not unique to Pakistan. In Afghanistan, language barriers are also one of the main challenges journalists face when working with data. For instance, with most data scraping tools the menus are in English, data about Afghanistan is available in English and when data stories are published from Afghanistan, they are more likely to appear in English, targeted at an international audience⁷³.

b) Right to Information laws

Right to Information (RTI) laws are present in three of four provinces of Pakistan and at the federal level. Earlier known as the Freedom of Information Ordinance, this right of citizens to access government records and information for better governance, was first promulgated in the country in 1997 with Pakistan having the distinction of becoming the first South Asian country to do so⁷⁴.

Yet, journalists need to jump several hurdles while working with these laws to access government data, to hold the state accountable and to audit their activities. Apart from operational challenges, the bureaucratic mindset and traditional culture of secrecy among public officials is a serious challenge⁷⁵.

⁷² “Findings”, Chapter 4, *Language and Education in Pakistan: Recommendations for policy and practice* by Hywel Coleman and Tony Capstick, 2012, British Council, accessed at

<https://www.teachingenglish.org.uk/sites/teacheng/files/Language%20In%20Education%20in%20Pakistan.pdf>

⁷³ “Reporting open data in Afghanistan: key challenges” by Catalina Albeanu, accessed at

<https://www.journalism.co.uk/news/building-a-reporting-community-around-open-data-in-afghanistan/s2/a564158/>

⁷⁴ Freedom of Information Pakistan Network accessed at <http://www.foiapakistan.com/right-to-information.html>

⁷⁵ “Status and implementation of RTI laws”, Chapter 2, *Pakistan Media Legal Review 2017* by Muhammad Aftab Alam for the Institute for Research, Advocacy and Development, accessed at <http://irada.org.pk/wp-content/uploads/2018/01/Pakistan-Media-Legal-Review-2017-IRADA.pdf>

Italian journalists also experience the same. There is a deep-rooted culture of non-disclosure in Italian public administration which makes the work of data journalists pretty hard⁷⁶.

Here are two viewpoints from Pakistani data journalism practitioners on the challenges when working with RTI laws.

“Data journalism remains a factor of media’s interest in investigative journalism and official transparency, which remains a problem with authorities slow to respond to RTI requests. This is mainly because of the official culture. Even if RTI exists, it cannot ensure a culture of data processing within government departments. That, then, is more a problem of internal communications and processes that needs to be streamlined to respond promptly to information requests. With no culture of proactive information processing for sharing with citizens, data availability becomes a reactive process and therefore not very efficient.”

– Aurangzaib Khan⁷⁷, Peshawar-based data journalism trainer and journalist

“The data that is available out there in the public domain is inconsistent and in a sorry state. That’s because of larger factors of accountability and corruption in our bureaucracy. The actual data collection that is done by the government is problematic at best and there is a reason for that. So that ‘aap ki khichai na ho’/you are not held accountable. It’s not that there is no data. It is there some of it is there such as Umar Cheema [of Panama Papers fame] with FBR [the Federal Board of Revenue]⁷⁸ has shown us⁷⁹.

– Rahma Mian⁸⁰, Karachi-based media strategist, lecturer and journalist

c) Innumerate journalists

When Aron Pilhofer, former associate managing editor for digital strategy at *The New York Times*, was asked⁸¹ why he thought people with coding backgrounds were moving into journalism and not vice versa. He said he knew journalists who took pride in not knowing maths. “Journalism is one of the few professions that not only tolerates general innumeracy but celebrates it. It’s hard to get a journalist to open a spreadsheet. It’s [numeracy] not just something that they, in general, think is held to be an important skill.” He pointed out it was a cultural problem as there was far too much tolerance for anecdotal evidence as material for news stories⁸².

This mindset is more or less the same in Pakistan, resulting in little understanding of how data could be utilised in stories. With government bodies releasing troves of data in the public domain, they are useless without journalists possessing numeracy skills and digital literacy to analyse them.

⁷⁶ “Working without Freedom of Information laws”, Page 39, *The Art and Science of Data-Driven Journalism*, Tow Center for Digital Journalism.

⁷⁷ Email interview on January 22, 2018

⁷⁸ Federal tax collection agency

⁷⁹ ‘Nearly 70 per cent of Pakistani lawmakers don’t file taxes’, accessed at <https://in.reuters.com/article/pakistan-tax/nearly-70-percent-of-pakistani-lawmakers-dont-file-taxes-group-idINDEE8BB08X20121212>

⁸⁰ WhatsApp interview on February 25, 2018

⁸¹ This query was put forth to him by Alex Howard for the 2014 report *The Art and Science of Data Driven Journalism*.

⁸² Page 46 of report *The Art and Science of Data Driven Journalism*.

“Journalists are wont to including data (as facts) in the stories as mere inclusion of them, but they need to analyse them for scope, depth and comparisons to suggest progress, its absence or deterioration.”

– Aurangzaib Khan, trainer and journalist

“There is an assumption here [Pakistan] that data stories take longer than other stories. It boils down to understanding what data is and the ways it can be used.”

– Shaheryar Popalzai⁸³, trainer and senior digital media resource

d) Threats

Journalists in Pakistan face multiple threats⁸⁴ and sometimes face physical violence, making for difficult working conditions⁸⁵. Data journalists, too, face comparable dangers.

“The risks for reporting are real. Controls of information, editorial policies, and ownership agendas will hinder the publication of data-driven stories just the same as all other news items. Pakistani news organisations tend not to worry too much about stories on government corruption and accountability. They are more cautious about the other groups so data stories about the judiciary, the military and the militants would potentially be risky both in terms of reporting and publication.”

– Waqas Naeem

⁸³ Email interview December 4, 2017

⁸⁴ See Chapter 3 Media Influencers of this paper

⁸⁵ For in-depth report on the kind of threats Pakistani journalists face, please see Reuters Fellowship paper: “Caught between Five Extremes: Reporting Pakistan” by Razeshta Sethna, accessed at <http://reutersinstitute.politics.ox.ac.uk/our-research/caught-between-five-extremes-reporting-pakistan>

5.3 RECOMMENDATIONS

a) Data journalism teams

Specialised teams within newsrooms would further boost data journalism in Pakistan. It doesn't have to be a large team but can be a lean one to produce data-driven stories. For instance, the data desk of *The Los Angeles Times* for years did data-driven projects with only two dedicated staffers⁸⁶.

But, we needn't look further as we can learn from the experience of our next-door neighbour India where at least three mainstream media organisations (*The Hindu*, *The Times of India* and *Hindustan Times*) have in-house data journalist teams.

In an interview by *Storybench* of Northeastern University's School of journalism, Gurman Bhatia and Samarth Bansal, data journalists at the *Hindustan Times*, based in New Delhi, India, talk about how the data team started in their newsroom:

"It all started last year [2016]. It was the vision and initiative of Nic Dawes, [formerly] chief editorial and content officer of the Hindustan Times, to kind of think of newer ways of storytelling. It started with the air quality map which first kicked off in late 2015 ... And with a few successes, they thought this was something we should do more. So, we started doing data interactive process in full swing sometime in January last year. January 2016 was also when Harry Stevens, one of the three data

⁸⁶Page 113 of *The Art and Science of Data-Driven Journalism*.

journalists on the team, joined HT and that's when we started doing things on the web. So, you could say management efforts and some early [project] successes led to the beginning of it.⁸⁷

Rukmini S, former Data Editor at *The Hindu*, India's second-most circulated English-language newspaper, discusses the utility of having an in-house data team within their organisation and the necessity of hiring data journalists:

"On any given day we [The Hindu] are able to respond very fast to things that happen. We are able to use data in our regular stories. We are able to get other journalists who don't necessarily do data stuff to incorporate it into their reporting. We're hiring more people. This is something we're building. So, I do think that organisations that don't do that are losing out on talent. They'll get special things that these people [independent data journalists] do for them, but they're not going to get sustained data integration into their journalism⁸⁸."

b) *Potential topics*

There is tremendous scope for data journalism in Pakistan as there are several topics that are not being looked into. In fact, only the surface has been skimmed even in the subjects that are covered, such as politics, economy and sports.

"We are open to data journalism but am unaware of it. We are willing to experiment. And want to do more politics and economy stories. We are certainly equipped to do it."

– Ali Khizar⁸⁹, Lahore-based editor of the research wing of Business Recorder, a financial daily.

"All issues of civic importance and public interest. Politics, governance, health, education, crime, the state of the economy, and other civic issues could be reported on with data. Themes of accountability and transparency in governance can be highlighted through data journalism. But even stories about culture and lifestyle could use data-driven storytelling. There's really no limit to it."

– Waqas Naeem

⁸⁷ "How the Hindustan Times is introducing Indian readers to interactive stories" by Shraddha Kakade, accessed at <http://www.storybench.org/hindustan-times-introducing-indian-readers-interactive-stories/>

⁸⁸ "Rukmini S discusses data journalism in India", interview by Sam Solomon for Polling One Billion, accessed at <https://pollingonebillion.wordpress.com/2015/09/28/rukmini-s-discusses-data-journalism-in-india-the-challenges-of-reporting-on-opinion-surveys-and-the-most-recent-census-data/>

⁸⁹ Telephone interview on March 3, 2018.

5.4 FUTURE

The future of data journalism in Pakistan will see the likelihood of the emergence of digital publications in languages other than English, stories that engage with readers through interactive data, investment in open government data, collaboration of large media organisations with small media advocacy groups and research think tanks specialising in data journalism, and more and more journalists undertaking training in data journalism.

But the future of data journalism is already happening in Pakistan, as observed by industry insiders.

“For short-term trainings, it was a mixed group of participants from all over Pakistan, from both print and broadcast media. Some of them are working for bigger organisations, such as Geo, Express Tribune, Dawn but some are from smaller newspapers and channels. The language medium is usually English and Urdu although we have had regional languages represented in other trainings.”

– Hira Siddiqui of The Centre for Excellence in Journalism⁹⁰ on other language journalists training in data journalism

⁹⁰ Email interview on October 13, 2017

“My data stories are written in English for News Lens⁹¹ and are translated in Urdu and Pashto.”

– Abdul Salam Afridi⁹², Peshawar-based data journalist, on translations of his stories in other languages

“The Urdu digital publications, I think, will evolve especially because they might begin to see audiences respond to interactive, data-driven content online and on social networks.”

– Waqas Naeem on future of other language digital publications

“Government data collection and digitisation has grown by leaps and bounds over the last decade [especially in Khyber Pakhtunkhwa, Fata and Punjab]. The two will come together soon.”

– Mahim Maher on digitisation of government data

“Newsrooms are more likely to collaborate with data start-ups. PakVotes⁹³ collaborated with Dawn and The Express Tribune in 2013. The ICFJ⁹⁴ media fellowship was a good collaboration between the fellows (Rahma Mian, Sheheryar Popalzai) with three newsrooms separately. Dawn’s website is currently running short videos and stories in collaboration with the Justice Project Pakistan⁹⁵, which is an advocacy concern.”

– Waqas Naeem on collaboration

“I am the news editor for The Friday Times⁹⁶ right now. I have no team as such. But I just got a friend, Shahzeb Ahmed of the Centre for Excellence in Journalism⁹⁷, to crunch, scrape and help visualise the data I got on encounters in Malir under Rao Anwar. I then asked a graphics designer Armeen Tinwalla to specially make me a map. You need a lot of support to do data.”

– Mahim Maher on collaboration

“Local reporters and editors are more likely to receive trainings on data journalism skills now than ever before. There are many online opportunities too and journalists are aware of Multiple Open Online Courses[MOOC] that are offered by international organisations.”

– Waqas Naeem on training

⁹¹ Joint project of Media Foundation 360, based in Lahore and Near Media LLC, an Atlanta-based media advising company

⁹² Email interview on March 5, 2018

⁹³ An election project of Bytes for All, a communication technologies think tank based in Pakistan

⁹⁴ The International Center for Journalists based in Washington.

⁹⁵ Legal action charity based in Lahore

⁹⁶ A Pakistani English-language newsweekly based in Lahore.

⁹⁷ Karachi-based institute for the professional training of journalists.

6. APPENDIX

Pakistan at a glance

Population: 207.8 million⁹⁸

GDP per capita: \$1410⁹⁹

Literacy rate: 60 per cent¹⁰⁰

Main languages: Urdu, Sindhi, Pashto, Balochi, Punjabi and English¹⁰¹

⁹⁸ These are provisional census results, but they have become controversial with some provinces refusing to accept the results. Accessed at <https://www.dawn.com/news/1353867>

⁹⁹ United Nations data accessed at <http://data.un.org/en/iso/pk.html>

¹⁰⁰ Economic Survey of Pakistan 2015-2016 accessed at http://www.finance.gov.pk/survey/chapters_16/10_Education.pdf

¹⁰¹ According to the 1973 Constitution of Pakistan Urdu is the official language but English can be used for official purposes, accessed at http://www.na.gov.pk/uploads/documents/1333523681_951.pdf

Mobile phone users: 146 million¹⁰²

Internet users: 34 million¹⁰³

No of TV stations: 89 satellite channels¹⁰⁴, eight state television channels¹⁰⁵

No of newspapers and periodicals: 464¹⁰⁶

No of radio stations: 143 FM channels¹⁰⁷

Ranking in World Press Freedom Index: 139¹⁰⁸

¹⁰² Pakistan Telecommunications Authority (PTA) accessed at <https://pta.gov.pk/en/telecom-indicators>

¹⁰³ Internet Live Stats Pakistan accessed at <http://www.internetlifestats.com/internet-users/pakistan/>

¹⁰⁴ List of TV Licence Issued by Pemra, accessed at http://58.65.182.183/pemra/pemgov/wp-content/uploads/2015/09/list_of_licences_issued_stv.pdf

¹⁰⁵ <https://www.ptv.com.pk/public/userView/ptvFamily>

¹⁰⁶ Member publications of All Pakistan Newspapers Society, accessed at http://www.apns.com.pk/member_publication/index.php

¹⁰⁷ List of commercial FM radio licences issued by Pemra, accessed at http://58.65.182.183/pemra/pemgov/wp-content/uploads/2015/09/List_Of_Commercial_FM_Radio.pdf

¹⁰⁸ Reporters Without Borders accessed at <https://rsf.org/en/pakistan>