DATA JOURNALISM IN PAKISTAN
CHALLENGES & PROSPECTS

MEDIA FOR
TRANSPARENCY

MediaMatters
*for***Democracy**

Policy Research & Advocacy Initiative

DATA JOURNALISM IN PAKISTAN: CHALLENGES & PROSPECTS

Writing and Design:

Waqas Naeem

Published by Media Matters for Democracy in 2020.

Available for use under Creative Commons licence (CC BY-SA 4.0)

About Media for Transparency

This digital guide is part of the learning resources created by Media Matters for Democracy for Pakistani journalists and news organisations under its Media for Transparency project. Since 2017, the project has attempted to build the capacity of local journalists to use data journalism skills and investigative reporting techniques in connection with the country's right to information laws to produce data-driven stories in the public interest.

As part of the project, Media Matters for Democracy has trained around 200 journalists, created a six-week online course on data journalism in Urdu, filed records requests and shared data in open formats online, produced in-depth data stories, and initiated an awards contests for data journalism in Pakistan. For learning resources and other information, please visit our website: http://pakrtidata.org

Acknowledgements

TABLE OF CONTENTS

Section	Page

INTRODUCTION

Pakistani news organisations and journalists routinely use data and public records in the course of their reporting. In fact, the amount of publicly available data, especially online data, has increased in recent years and many reporters have learnt to access it through their reliable networks of sources or by searching the relevant government websites. Pakistani journalists regularly rely on statistics in their news coverage, regardless of the news beats they cover. Sometimes these are budgetary figures; other times, these are records of enrollment or registration. On many occasions, the data deals with service delivery failures or achievements: the number of ghost schools, for example, or the number of new basic health schemes initiated by the government. Day in, day out, beat reporters quote statistics in their stories or use data to supply evidence for the issues under consideration.

Yet, many Pakistani journalists are also fascinated by the idea of 'data journalism' as something alien or inaccessible. Some are simply afraid of numbers – innumeracy is unfortunately the general trend in the local media, not the exception. Others do not consider their work as data journalism; they might not be entirely correct, but there is a kernel of truth to the lack of self-confidence. Even though journalists use data, including statistics and documentary evidence, in their reporting, they most often do not follow any disciplined approach or systematic process to crunch numbers. As a result, they are mostly dependent on analysis or information fed by sources, official documents, and research studies. They publish it as it is, and often fail to ask the most important questions or find hidden patterns in the data themselves. As a result, data-driven investigations of public interest issues remain scarce. Isolated stories of daily current affairs that use quoted statistics dominate the coverage.

This report offers recommendations and resources about data-driven journalism for journalists, news organisations, and media development organisatons.

This trend persists despite an increase in the availability of training opportunities related to data journalism during the past decade or so, with many national and international media development non-profit organisations chipping in to provide free training and resources on data-driven news reporting for local journalists. The training sessions mostly attract individual reporters, who apply and attend these sessions often at their personal initiative. Occasionally news organisations also collaborate with non-profits to conduct data journalism training of their staff. Training of trainers, including training of journalism educators and students, has also taken place in the country in the past. These skill-development opportunities are limited in scope and the impact of their outputs has to contend with a set of challenges.

Some of the challenges journalists speak of often when it comes to their ability to deploy data journalism skills in Pakistani newsrooms are concerned with the structural issues of the national and local news media. The daily reporting routines and deadline pressures, the political economy of news media that takes away from the public interest role of journalism, and capacity constraints regarding data analysis are some of the most prominent problems. There is, however, the potential for innovation among news outlets and some hope about the appetite among news audiences for data-driven reporting. Local journalists realise the importance of data in telling public-interest stories and often do their best to work with data and public records. Whether or not they could do it competently and consistently in order to make their reporting impactful is the subject of an ongoing struggle, in which not giving up would perhaps be the most important advice.

With this context, 'Data Journalism in Pakistan: Challenges and Prospects' briefly describes some of the routines that impede the production of data-driven stories in the Pakistani news media. The report also shares examples from national print, broadcast, and digital media to highlight recent trends that demonstrate the importance of data journalism for public-interest news reporting. The report relies on interviews with journalists, who are familiar with data journalism concepts, to provide evidence about the key challenges to data-driven journalism in Pakistan. The interviews also offer insight about strategies that could prove useful for local journalists and news organisations to adopt and use data journalism skills.

The report offers recommendations for individual journalists, news organisatons, and media development non-profits regarding the promotion and practice of data-driven journalism. Finally, it also provides access to local and international resources for Pakistani journalists who are interested in developing their data journalism skills.

KEY ISSUES

The 2018 Reuters Institute Fellowship Paper "Showing the way: How big Pakistani media groups are leading the surge in data journalism" mentions language, innumeracy, and the flawed implementation of right to information laws as some of the obstacles to data journalism in the country.

Most of the data-driven journalism is in the English language but the majority of the public watches and reads the news in Urdu. The Urdu news media is reluctant to pick up on data skills, even though some digital publications are now trying to change this attitude. Innumeracy is a big problem and local journalists do require training to work with data. The right-to-information laws were expected to make it easier for citizens and journalists to access public records, but in many jurisdictions journalists have found information requests to be a lengthy and tedious process that is useless if they are working on tight deadlines.

In addition to these, the following technical issues prevent journalists from effectively using or presenting data in their reports when they do manage to get access to data and public records.

DATA DISCIPLINE

Most Pakistani journalists do not create or maintain their own data sets or databases. This means they lose opportunities to conduct data-driven follow-ups on an issue. They are also unable to analyse trends on their beats. Organising and maintaining the data they used in one story could help them later on to ask new questions about persistent issues or find new and better ways of telling stories about systemic problems.

NO CALCULATIONS, NO COMPARISONS

Even those journalists that are comfortable with using data in their stories to supply evidence often do not make the effort to run some basic calculations or compare data over time. As a result, their stories end up regurgitating the numbers without sufficient context. They miss the bigger picture.

> SURPRISINGLY, ACCESS TO DATA AND STATISTICS MAKES THE REPORTERS IGNORE THE MOST IMPORTANT PART OF THEIR JOBS: REPORTING.

OVERSTUFFING NUMBERS

Frequently journalists get access to a report or budget summary on their beats. In these cases, they tend to publish all the data points or statistics available to them. The stories get overstuffed with data and become almost unreadable. Surprisingly, access to data and statistics makes the reporters ignore the most important part of their jobs: reporting. They often fail to speak with human sources

about the impact of the reality the numbers represent or the reason why the situation presented by the numbers exists.

TYPICAL BROADCAST NEWS FORMATS

Broadcast news media is the most prominent source of news and information in Pakistan, but broadcast journalists have a difficult relationship with rich data despite their visual format. Often when broadcast journalists get access to data, they tend to present it as live beepers or tickers. This takes away from the potential of working on the story with original analysis and reporting and presenting it as a cohesive data-driven report. Instead, the data is presented as successive lines of text or a brief Q&A with a newscaster. This treatment means that the data use will never have the kind of longevity or impact that a news package based on the data could have.

CAUSE FOR OPTIMISM

Some national and local news organisations have continued to use data journalism skills to tell stories of public importance. These examples provide some cause for optimism and also present approaches for other journalists to follow if they want to produce data-driven news reports.

USE OF DOCUMENTS AND VISUALISATIONS FOR A RANGE OF STORIES

Samaa Digital's multi-part series on a megaproject to supply water to Karachi is a great mix of data-driven investigation, multimedia elements, and explanatory reporting. The story uses documents and data to trace back the origins of the project, describes the technical and bureaucratic aspects in accessible language, and discusses the solutions and responsibility for the future of the project. In other stories, the same outlet has used short videos and animated visualisations to present data in a clear and meaningful way. The range of data-driven projects by Samaa Digital offers a way for journalists to use data skills adaptively: they can focus on small stories regularly and work on a major investigation from time to time.

COLLABORATION WITH EXPERTS

Dawn's PRISM section started a new monthly series for evidence-based articles on policy issues in 2019. The Analytical Angle is published by Dawn.com in collaboration with the Lahore-based policy institute Centre for Economic Research in Pakistan (CERP) and the Harvard Kennedy School's Evidence for Policy Design research initiative. In the first article for the series, CERP researchers Maroof Ali Syed and Maha Rehman wrote about Pakistan's trade imbalance. The experts used a variety of data sources to show that the country's exports are less than a quarter of its total trade volume and its exports have only grown by half in the decade until 2017 while regional competitors, such as Bangladesh and Vietnam, have tripled their exports during the same period.

> JOURNALISTS CAN USE DATA SKILLS ADAPTIVELY: THEY CAN FOCUS ON SMALL STORIES REGULARLY AND WORK ON A MAJOR INVESTIGATION FROM TIME TO TIME.

While the Prism series is written by subject experts exclusively and does not include reporting, it could be a good model for news organisations to get their journalists to collaborate with experts on data-driven reports about policy, especially about governance issues. The experts can provide the relevant data analysis and the journalists can do the reporting based on the analysis.

USE OF DATA IN POLITICAL TALK SHOWS

On 5 November 2019, the Geo News talk show Aaj Shahzeb Khanzada Kay Saath ran a segment on a press conference held earlier that day by the Federal Minister for Economic Affairs, Hammad Azhar. Azhar had presented the economic performance under the ruling Pakistan Tehreek-e Insaf (PTI) government, including information related to the fiscal deficit and tax revenue. For the indicators, the minister had made comparisons either with the previous governments' initial years or his own government's first year in power.

The talk show host, Shahzeb Khanzada, used year-on-year data for the same indicators for the past few years to argue that Azhar put spin on the economic performance. Using bar charts and percentage difference calculations, Khanzada illustrated why the minister might have changed the comparison time-period for every indicator rather than use standard year-on-year comparisons. If the comparisons had been made consistently, Khanzada argued, the economic performance would probably be much worse today than what the minister claimed.

A video clip of the programme was viewed thousands of times on Twitter and generated a discourse on social media around the economic performance. The popularity of the segment led to a follow-up programme where the minister tried to present a rebuttal of the issues raised in the previous show. This is a good example for television talk show hosts and their research teams to use data to test official claims. Even though the programme did not discuss the factors behind economic decline, the host did call out the government representative for a lack of transparency about the methodology for comparative analysis.

VIEWS FROM THE FIELD

In order to get a clear picture of what the challenges and prospects are for data journalism in Pakistan, we spoke with three journalists who have worked on data-driven news stories at their respective organisations and are currently in positions where they could encourage more journalists to use data journalism skills in their reporting.

Mahim Maher, Editor Digital Properties Samaa, shared the strategies she and her team use for pursuing data stories.

"When I do my own stories, they always have a data component," Maher said, attributing this to the fact that writing about Karachi infrastructure and the issues of the metropolis require some measure of quantification. "I come from this school of thought (that you have to) measure to manage."

She said finding data is the number one problem. Her strategy is to team up with beat reporters.

"(Most beat reporters have) never worked in data frankly, but they know where data is in their institutions and they get it for you," Maher said. "The tragedy is that the beat reporters do not know how to crunch the data and use it to amplify the issues (in their reporting)."

The beat reporters might bring in a small data set or an Excel sheet, but most often they would not have the Excel skills to work on it. Maher said, in such situations, she often looks toward a network of people with data skills. She has built this network over the years. It is composed of private university graduates with data or statistics skills who ended up in newsrooms where Maher worked for short periods of time. Or else, she said, she would clean and analyse the data herself, which takes much longer.

"When the data is in some shape, I have to trim the fat and make it a one-point agenda," Maher said. She then works with the graphics team. "I know what I want (from a certain data story), and I will put an example, (usually) an international example and say, 'here are the variables, now make it look like this international example'."

It is a simple strategy but it pays off.

Maher said she pushes her team to work on small and quick data-driven interventions regarding current affairs. Usually it begins with a basic news question. She gave the example of the train accident in October 2019. The question in this case was: how many train accidents have taken place under the watch of the current federal railways minister?

"We got an editor to scrape the web for past incidents and put the dates and number of fatalities in a table," Maher said. "We cross-checked with the railways beat reporter and vetted the data; I sat with an animator and drew it on paper and said, 'these are the numbers and then I want them to look like this sketch'."

A simple video animation shared on Samaa's social media accounts the same day the accident took place showed the numbers tallied up in a table with each successive train accident. Maher

characterised this approach as "get the formula right and hit the news cycle". From experience, she said, such short data pieces do well with the online audiences.

Maher said the key to making data use sustainable in newsrooms is if the public can see it and understand an issue clearly. The average news desk in Pakistan, she said, is very far from data-driven journalism because they do not have the capacity or the interest and often journalists are scared of data. She said Samaa Money's reporter Farooq Baloch uses a lot of data in his stories and it works because he uses it to explain and contextualise the issues. On the other hand, Maher said, rookie reporters sometimes get excited about tools but when they do one data story with a new tool, it often does not look good on mobile or does not make the issue accessible to the public.

In the Samaa newsroom, she said, there is a culture of discussing story ideas but that is not necessarily true for other news organisations. "The culture of talking in the newsroom has to be created by example," Maher said. "My reporters talk with me or our web editor; they will ask, 'how do I do this?' and I will re-route them to the relevant resources."

She said if reporters can get trained in using Excel and the art department can help them with visualisations, more data stories can be produced. However, she recommended that journalists need to start with small data projects.

> MAHER SAID THE KEY TO MAKING DATA USE SUSTAINABLE IN NEWSROOMS IS IF THE PUBLIC CAN SEE IT AND UNDERSTAND AN ISSUE CLEARLY.

Farooq Tirmizi, the Managing Editor of Profit Magazine, has a pessimistic view of the potential for data journalism in Pakistan. But his cynicism offers a reality check.

"I'll be honest with you," he said. "There is not a single print organisation (in Pakistan) with the budget to pull off this (kind of data journalism work). Some of the TV people have money for it, but they are blipping idiots with no intention to do this."

He said TV does not lend itself to the kind of data-driven reporting we see around the world and very few people in Pakistan's broadcast news media are doing interesting work anyway.

"So some organisations might try to do something of this nature with their digital teams but I do not know how serious they will be," Tirmizi said. "It is a low priority; absolutely no one talks about it because no one cares."

He said the business decisions in Pakistan's news industry are not based on good journalism rather these are about what can be sold to advertisers. He said even a Pakistani business trade publication that has years of unique data sets in its archives does not have a dedicated data team, which goes on to show how poor the situation regarding data-driven journalism is in the country.

Tirmizi also spoke about the lack of capacity among journalists to think innovatively. This keeps

them from effectively using data for their stories. He said if you show specific examples to local reporters then you can get them to not repeat their mistakes and they might get it at, say, the seventh attempt. But it does not seem to be their news instinct to originally see a phenomenon and question what happened and why, he said.

"Thinking quantitatively but telling the story qualitatively, it just does not come to them," Tirmizi said.

He said there is a desire among journalists to break a story but there is a corresponding lack of curiousity.

"There is a difference between breaking a story and the desire to fully and completely understand what happened," Tirmizi said. "It's the second desire that you need to have that will make you good at (data-driven journalism)."

As an example of how bad the situation is, he said not a single Pakistani newspaper can tell you how much the electricity rates have gone up in the past year. "There is a reason government can get away with atrocious planning, because there are no follow up questions," Tirmizi said.

In other countries, he said, we would notice think tanks and research institutions putting out numbers about different issues. The vast majority of such infrastructure does not exist in Pakistan so it is difficult to have this kind of journalism without it.

"Data journalism cannot exist in a vacuum," he said.

Gibran Ashraf, a Senior Sub-editor at The Express Tribune, said journalists are bogged down with daily work so they do not get sufficient time to work on data stories. Another issue is hurdles at the newsroom level where even if some reporters want to work on data stories, they might be surrounded by colleagues who do not understand the value or advantages that data journalism brings. He said the primary purpose of local news media is to serve their own interests rather than the public interest, which takes away from the data journalism initiative.

Apart from that, Ashraf said, those journalists who have worked with data have not been able to produce enough compelling stories to convince others. He said data journalism requires a little more effort and skills than everyday news coverage. Collecting and analysing data can be a long and intimidating process, and therefore reporters often do not make the effort to do their own calculations and analysis. They would rather pick statistics from a report and print them, he said.

Ashraf said we do not see any dedicated data teams at any local news organisation. This would require investment in public-interest journalism, and the nature of the media market in Pakistan is such that this kind of investment in building journalism resources is unlikely to take place. About collaborative data projects, he said, local reporters feel that if they share their work with someone else they will not get credit for it.

"They don't realise the power of collaboration," Ashraf said.

He said news organisations need to build audience awareness and think about how their audiences respond to data-driven news. This can help them push data-driven stories to diverse audiences in

different formats. At the organisational level, Ashraf said, more training opportunities, especially skill-building interventions over longer periods of time, can help to get journalists to practise data journalism skills. More human interest stories that use data components, greater experimentation with digital formats, and community-building around data stories so reporters get to work with experts, data analysts, and other professionals, can also help encourage more data-driven journalism locally, Ashraf said.

RECOMMENDATIONS

These recommendations can help Pakistani journalists who want to use data journalism skills in their work more regularly.

LEARN SPREADSHEET SKILLS

The first step for any Pakistani journalist who wants to make better use of data in their news reporting is to learn basic spreadsheet skills. Microsoft Excel is a great place to start. It can help journalists understand data management procedures, perform basic calculations, and find patterns in the data.

LEARN BASIC DATA VISUALISATION PRINCIPLES

Many local journalists are interested in presenting their data, especially statistics, on colourful charts and maps. But they often rush to learn online tools that generate quick graphs based on the data input. The result is that often these journalists end up making elementary mistakes in visualising their data. These could include calculation errors, baseline problems, and proportionality mistakes, among others. Sometimes they stuff their charts with numbers and colours, making the final product extremely complicated and difficult to understand. This is why journalists must familiarise themselves with the principles of data visualisation first and learn how to prepare their data for visual treatment before they use any online or offline design tools.

START PRACTISING WITH SMALL STORIES AND BASIC QUESTIONS

It is always good to build your confidence in data skills over time by starting with small current affairs stories and moving gradually to bigger, more complex investigations. Journalists should ask basic questions about the issues they are reporting on their beats and think, "How can the data or documents available on my beat help me answer these questions?" or "What kind of records do I need to find the answers to these questions?" This is where spreadsheet skills will also come in handy.

******

News organisations can also support the practice of data journalism by providing their journalists with the enabling environment, skills, and resources to work on data-driven investigations.

SET UP DATA TEAMS

Around the world, news organisations have benefitted from setting up small data units – teams of journalists that work on data-driven investigative projects. These teams are composed of editors, reporters, database managers, visualisation experts, and developers. Together, they support each other's work to report efficiently and effectively on systemic problems of public importance. The concept of a data team may appear to be expensive and resource-intensive, but news outlets can be innovative about how they approach it given the available financial and human resources at their disposal.

CREATE AND MANAGE YOUR OWN DATABASES

A lot of data goes through a typical newsroom in Pakistan every day. Most news organisations cover a variety of issues, including politics, health, education, sports, and the economy. Beat reporters rely on data nuggets and statistics for many of their regular daily stories and most of their features and special investigations. However, there is very little effort – outside of individual initiative – to keep track of the data on an organisational level. Even if news organisations cannot afford to set up a data team or are simply not interested, they could do better to organise and archive the data their reporters are bringing in. Data management skills training could help beat reporters and editors stay on top of key developments and trends analyses for the issues and areas they cover.

EXPERIMENT WITH DATA PRESENTATION

The best data stories capture the attention of the readers and help them understand complex issues that affect the public. The impact and, in the case of digital publications, the readers' engagement with data-driven stories can provide newsroom managers with solid evidence to push their reporters for more data journalism. But with digital media, publishers also need to consider what kinds of formats their regular audiences are most receptive to and think about how they can present stories with data in these formats. The answers to these questions may vary from news outlet to news outlet and from platform to platform, but these answers can help the journalists determine the best way to present their data-driven news reporting to the public.

******

Local and international media development organisations often deliver basic skills-development training for Pakistani journalists. However, there is a need to assess the effectiveness of these training programmes and think about revising the design or delivery of the skill-building training based on the assessment. The following recommendations could be helpful for media development non-profit organisations engaged in data journalism capacity-building work in the country.

SPEAK WITH LOCAL MEDIA OUTLETS ABOUT THEIR TRAINING NEEDS

Journalists and news organisations might be more responsive to data journalism techniques if media development organisations can in turn demonstrate how these data skills will actually help the journalists and news organisations in their daily routines and news operations. For this, the media development organisations need to hear directly from the outlets and training participants and tailor their general learning modules and training material to fit the needs of the trainees.

DESIGN ACCESSIBLE TRAINING MATERIAl

English training materials might not be best suited for local journalists. Media development organisations could think about the accessibility of the training materials, whether it is in terms of languages or delivery formats. Translation to local languages could be a good first step. Short videos might work. Listening to local journalists and newsroom managers could prove beneficial in this regard as well.

OFFER ALTERNATIVES THROUGH DATA JOURNALISM

One of the best ways to convince journalists to adopt data journalism skills and techniques is to show them examples of impactful, high quality data-driven journalism. While many international news examples are available, local journalists might not be able to relate to them. Media development organisations should develop case studies of good local examples where journalists used data or documents to investigate and report on an issue of public importance. In the absence of such examples, the next best thing could be to take a story that showed potential and demonstrate how it could have been improved through the use of a data-based hypothesis, data analysis, and visualisations.

RESOURCES

- **MEDIA FOR TRANSPARENCY**

Media Matters for Democracy's Media for Transparency project offers many resources for Pakistani journalists to learn data journalism skills. The resources include a self-paced online course with video tutorials, exercises, reading material, and assignments. The course is available in Urdu. Please visit: http://pakrtidata.org/

- **DATAJOURNALISM**

The European Journalism Centre's project Datajournalism.com is a treasure trove for learning resources on data journalism, including courses, video lectures, guidebooks, and informative articles. Please visit: https://datajournalism.com/

- **INTERNEWS PAKISTAN DATA JOURNALISM TRAINING MANUAL**

The training manual, developed by data journalism trainer Eva Constantaras, for Internews Pakistan provides training modules and lab exercises for journalists and journalism educators to learn and teach data journalism skills. The manual is available on Github here: https://github.com/evaconstantaras/internewspakistanmanual

- **ONLINE JOURNALISM BLOG BY PAUL BRADSHAW**

Data Journalism educator Paul Bradshaw writes extensively about data journalism trends and techniques on his blog. You can follow or subscribe to his posts here to remain up to date about latest conversations from the world of data-driven journalism: https://onlinejournalismblog.com/

ABOUT MEDIA MATTERS FOR DEMOCRACY

Media Matters for Democracy is a Pakistan based not-for-profit working to defend freedom of expression, media, Internet, and communications in Pakistan. Our activities include policy research, advocacy, training, legal aid and support and public interest litigations. The organisation was founded by a group of journalists who believe in free expression and are working to ensure that the media and public alike have the tools and an enabling environment to exercise their fundamental rights. Our core objective is to ensure that rights to free expression, association, access to information and related freedoms are protected in Pakistan, in policy and practice.

You can follow our work at mediamatters.pk