

MEDIA FOR
TRANSPARENCY



DATA JOURNALISM IN THE NEWSROOM

A DIGITAL GUIDE

DATA JOURNALISM IN THE NEWSROOM: A DIGITAL GUIDE

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

We would like to thank the following organisations and individuals for their work on data journalism research and pedagogy, which helped us in putting together this guide: Datajournalism.com (An initiative of the European Journalism Centre); the Global Investigative Journalism Network; Numbers in the Newsroom: Using Math and Statistics in News by Sarah Cohen; Data Journalism Handbook edited by Liliana Bounegru, Lucy Chambers, and Jonathan Gray; Online Journalism Blog by Paul Bradshaw; Internews Pakistan Data Journalism Training Manual: Thinking Like A Data Journalist by Eva Constantaras; The Online Journalism Handbook: Skills to Survive and Thrive in the Digital Age by Paul Bradshaw and Liisa Rohumaa; and, The Investigative Reporter's Handbook by Brant Houston and Investigative Reporters and Editors (IRE).

TABLE OF CONTENTS

Section	Page
Setting up a data team	1
Data management	3
Planning data-driven investigations	5
Data journalism for broadcast and multimedia reporting	8

SETTING UP A DATA TEAM

Data journalism can often be a lone effort for journalists in Pakistan, with individual reporters or editors trying to use data as evidence for their work. A more disciplined and systematic approach might be required if news organisations want to increase the frequency, quality, and impact of their data-driven news reports. The example of international news organisations shows that one of the best strategies is to set up a dedicated data team in the newsroom. News outlets such as The New York Times and the Washington Post in the US, The Guardian in the UK, and Ojo Publico in South America have followed this model successfully and produced public-interest data stories in the recent past. The model is now being replicated at news organisations around the world.

The following tips can help you set up a data team at your news organisation.

1. THINK ABOUT SIZE AND DIVERSITY

- Do not worry that you will have to set up a large team with a lot of staff members. You can start small.
- The size of a data team can be relative to your news output needs, and it can be adapted to suit the funding and resources available at your media organisation.
- More important than the size of the team is the diversity of its members.
- You need to ensure that journalists with different skills and newsroom roles are included in the team. It should not consist of only reporters, but can and should also include editors, data analysts, illustrators, and web desk staff.

THE DIVERSITY OF A DATA TEAM IS MORE IMPORTANT THAN ITS SIZE.

So who should be in your data team?

2. TEAM COMPOSITION

Successful data journalism teams around the world show the following roles should be reflected in a data team:

- Reporter - Start with one or two, but you can increase the number based on need.
- Editor - At least one editor to supervise the planning and production.
- Database manager - This could be a reporter with good data journalism skills or a data analyst or programmer with an understanding of journalism, but you need at

least one member to create, maintain, and analyse databases for your various projects.

- Data visualisation specialist - To help with the visual treatment of data; you can work with existing human resources in the newsroom such as illustrators or graphics designers, but you may need someone that can handle interactive visualisations for web publishing.
- Developer - Someone with IT skills to help you with presentation of your team's data stories especially on the Internet. The web or applications developer can add crucial interactive components to the data story, which could benefit your storytelling and reader experience.

3. COLLABORATE

The work of the data team is based on the principle of collaboration.

- Start with planning meetings. Team members should collectively decide which stories they want to pursue together. Begin with medium-term or long-term data investigations to manage time.
- Once the team takes on a project, ensure that every member knows their role.
- Figure out a schedule and deadlines because team members might have other daily duties in the newsroom too. Set realistic and achievable targets.
- Get the team members to talk about their data problems and how to solve them. Through these discussions, the team members will also transfer their skills to each other.
- Encourage the data team members to help other journalists around the newsroom with daily stories related to current affairs that can use a data-driven treatment.

4. DO NOT WORRY ABOUT THE TOOLS

Start with basic tools and learn to use spreadsheets. Always try and connect your use of a data journalism skill with a specific news problem or reporting question.

5. SET STANDARDS

The La Nacion data team randomly checks data samples at least three times. Decide how your data team will verify the data and how will you share your methodology with the readers.

6. GET TRAINING AND READ

- Nominate your data team members for local or international training opportunities. Look for online resources to build your team's capacity.
- Review the data-driven projects published by news outlets around the world and think about their story ideas, reporting, use of data, and use of visualisations. Apply the learning to your work.

DATA MANAGEMENT

Pakistani journalists use lots of data points and documents in their stories on a daily basis but these are typically one-off reports and there is little effort to store and retain the data for future use in follow-up or trend analysis reporting. Data journalists should try to build their own databases to avoid this issue. This will help them in cases where there is no publicly available data or when they have to combine multiple data sets or if they are working with descriptive documents.

Here's how journalists can go about setting up simple data sets:

- As an example, consider that a crime reporter wants to record incidents of car thefts in the city. They can start a new Excel sheet to log any car theft crime reports they come across. They can define what information should be included in the data set, such as location of car theft, date of incident, police station. They can then populate the data set over six months or a year and do a trend analysis at the end of the period.
- When creating a database, we have to think about variables. A variable is any attribute which will take on different values within an acceptable or pre-defined range. For example, the “date” variable can take on a different value each day of the year. But you cannot store the “registration number” of a stolen vehicle in the date field. For this, you will have to create a separate variable. Variables most commonly appear as column labels in a data table.
- News stories often have recurring attributes that can be quantified as variables. In the example above, the date of complaint, location of incident, and name of police station are all variables that can have different values for different incidents.
- You also have to think about records. One record is usually one complete incident described by a set of variables. For example, if the car reported stolen from Blue Area in the jurisdiction of Margalla Police Station on 1 January 2020 is one incident and should be documented as one record, but it is composed of values for three variables: Date = 1 January 2020; Location of incident = Blue Area, and Police Station = Margalla. Records are stored as rows in a data table.
- While choosing variables, journalists need to think about the importance of the information stored in the variable for their reporting. Some variables will be essential to a data set. For example, in a data set about dengue patients in Punjab districts, the “number of reported dengue cases” is an essential variable. The names of the patients might or might

MICROSOFT EXCEL, GOOGLE FORMS & SHEETS, AND GITHUB CAN BE USED BY DATA JOURNALISTS TO STORE AND SHARE DATA SETS.

not be available and can be considered non-essential variable for this particular data set.

- It is often a good practice to limit the range of values a variable can assume. Some variables might take on infinite values but most variables can be limited to a few values or a binary option such as yes and no. Limiting the range of values a variable can take will make analysis easier.
- Data journalists must consider how they are going to record nil or missing responses for a data variable. Many times information will not be available against a particular variable for a particular record. Usually “N/A” or blank cells are used but this might affect analysis so journalists have to be careful about the convention they follow.
- One rookie mistake while quantifying data is the desire to store aggregates. Variables should be created to store attributes for individual records. Aggregates can be calculated using spreadsheet formulas during analysis. For example, the car theft data set should have one entry for each incident rather than a variable that counts the total number of car thefts. The latter would only provide a single data point for reporting and would not allow in-depth analysis of the data.

But which software can you use?

1. USE EXCEL

Microsoft Excel is a popular and convenient spreadsheet software used by data journalists. Journalists can use it to record new data sets that they can return to later for analysis and follow-up reporting. The “Fundamentals of Data Journalism” online course in Urdu offers video tutorials about Excel use. You can check out the course at <http://pakrtidata.org>.

2. USE GOOGLE FORMS AND GOOGLE SHEETS

Journalists can create a Google Form for a database if they want to make their manual data entry process a bit more convenient. The Google Form is linked to a Google Sheet, a spreadsheet similar to Excel. The Google Sheet file can be used at any time for data analysis. Video tutorials about using Google Sheets and Google Forms are available in the “Fundamentals of Data Journalism” online course in Urdu by Media Matters for Democracy.

3. USE GITHUB

Github is a platform that allows users to publish programming code and data sets for public use. It allows users to collaboratively edit code and share changes in data sets over time. It has become a popular portal for data journalists to share their data analysis work with readers and other journalists in the interest of transparency. Data journalists can also use github to publish their data and maintain it over time. You can check out how to use Github at this link: <https://guides.github.com/activities/hello-world/>

PLANNING DATA-DRIVEN INVESTIGATIONS

Producing quality data-driven investigations that serve the public interest and create impact on policy is the best way to incentivise and encourage data journalism within a newsroom. Data teams provide evidence of their necessity through such projects. It is important that new data teams apply themselves wisely and look towards systemic issues of public importance that can be addressed with the help of data. The following model for planning a data-driven investigation is based on the process described in the Investigative Reporter's Handbook and the six-step data story process shared by Eva Constantaras in the Internews training manual for data journalism.

1. CHOOSING AN INVESTIGATION

- Story ideas for an investigation can be derived from perpetual or systemic issues, such as corruption accusations and malpractice allegations.
- News tips, follow-up on beat stories, and daily reporting can also help with investigative story ideas.
- Always keep the public interest in mind when choosing an investigation idea.

2. HYPOTHESIS

This is the claim that you can test with the help of data to confirm or reject it. It offers clarity and focus for the investigation and is a guaranteed way of checking if the data story is feasible or not. Specific hypotheses that can be tested with data will help you identify data sets, maintain realistic goals, and ensure you either get a story or save you from wasting time when there is no story.

THE DOCUMENT STATE OF MIND SUGGESTS THAT INVESTIGATIVE REPORTERS SHOULD BELIEVE “A DOCUMENT EXISTS SOMEWHERE TO EXPLORE, CONTRADICT OR CONFIRM EACH POINT OF AN INVESTIGATION”

Eva Constantaras suggests that a strong hypothesis has the following characteristics:

- It presents a claim that can be confirmed or refuted with data.
- It is specific about what is being measured.
- It speaks about the data is available.

- It connects with a topic is important to the public.

3. BACKGROUND RESEARCH

Read up as much as you can on your chosen topic so you can develop an understanding of the issue.

Good background research helps to:

- Identify sources
- Identify secondary documents
- Identify unanswered questions

4. DOCUMENTS STATE OF MIND

Data journalists must never think that data does not exist on the issue they are reporting. They need to adopt a documents state of mind, used by investigative reporters, which states that “a document exists somewhere to explore, contradict or confirm each point of an investigation”.

- The kind of documents journalists should look for depends upon whether they are investigating individuals, institutions or issues.
- For individuals, property records, licenses and court proceedings can be useful.
- For institutions, trust deeds, employee information, registration documents, tax returns, audit reports, mortgages, and loans can provide clues.
- Documents can help journalists develop chronologies of incidents and reveal important facts or details that were buried under the paperwork.

5. ANALYSIS

Data analysis serves a two-fold purpose: It can provide answers to your initial questions and throw up new questions for further reporting.

6. INTERVIEWING

- Most investigative reporters start interviewing early on in the investigative process but sometimes documents can lead you to important questions.
- Data journalists should try to find insiders, former employees who could have a deep understanding of the system, whistleblowers, and outside experts.
- But most importantly, journalists have to speak with ordinary people affected by the issue.

7. ORGANISATION

Data-driven investigations can take time and if you do not regularly organise your notes you might have a hard time when you sit down to write at the end. Use an internal filing system and document the analysis steps so you can re-trace them later.

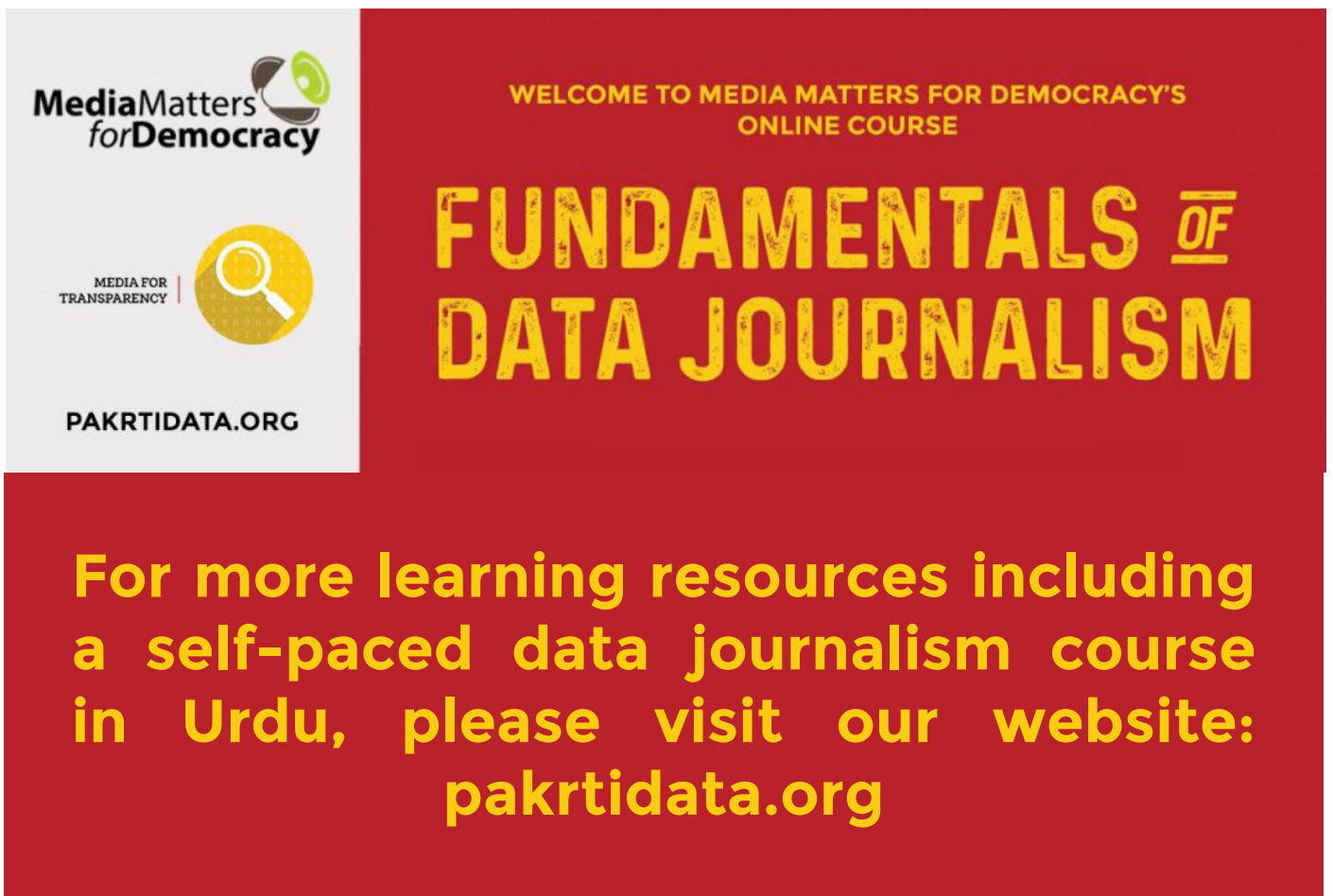
For writing or producing news reports, data journalists can use various story structures.

8. VISUALISATION

Data-driven news investigations should have a visual component to communicate the findings. If the visualisation is interactive, it can increase engagement. The principles of visualisation should be followed but above all the information should be communicated clearly and accurately. More guidance on visualisation is available in Media Matters for Democracy’s “Fundamentals of Data Journalism” online course in Urdu.

And a final tip:

Before publishing, data journalists must go back and check all the facts and data findings in the story to ensure there are no errors.



MediaMatters
forDemocracy

MEDIA FOR
TRANSPARENCY |

PAKRTIDATA.ORG

WELCOME TO MEDIA MATTERS FOR DEMOCRACY'S
ONLINE COURSE

**FUNDAMENTALS OF
DATA JOURNALISM**

**For more learning resources including
a self-paced data journalism course
in Urdu, please visit our website:
pakrtidata.org**

DATA JOURNALISM FOR BROADCAST AND MULTIMEDIA REPORTING

Broadcast journalists often complain about the newsroom pressures and time constraints that keep them from working on data stories. Additionally, one concern among broadcast producers is that the audience will not understand the use of data and statistics in news packages.

At the same time, Pakistani news organisations and freelance journalists are now catering to audiences on the Internet through their web content and social media platforms. This section will share tips and techniques for broadcast and digital journalists for working on data stories.

1. DO NOT SETTLE FOR BEEPERS AND TICKERS

Many times when broadcast journalists get their hands on data or statistics, they share them as forgettable beepers or tickers. A bit of additional reporting especially interviews with affected communities and some context can easily transform a series of tickers into an impactful data-driven news package.

2. LOCALISING INTERNATIONAL STUDIES

When an international organisation launches a research report with statistics about Pakistan, many broadcast journalists are tempted to publish the numbers as they were shared. Think about what the numbers mean for the situation on ground and how you can locally contextualise the international report by speaking with community sources and comparing with official data.

3. CASE STUDIES TO EXPOSE SYSTEMIC ISSUES

Many times we forget that numbers are generated due to the real issues faced by real human beings. Whenever you come across a set of data, always think about the phenomenon behind it and how it would have affected people. For example, if the data shows that the education system is in crisis, speak with teachers and parents of school children to get a sense of the impact.

4. EXPERIMENT WITH DIGITAL FORMATS

Animated videos, maps, visualisations, explainer videos, and audio embeds are all ways you can add to your data storytelling techniques. Do not use a format or visualisation to show off, but think about how it can help you to tell the story in a more effective way. Think about your readers and how it may help them understand the issue better.

5. START WITH QUICK DATA STORIES

Broadcast journalists, who feel they do not have time for a big data project or are required to turn in daily stories, should start with quick data stories. One way of going about using data in daily news reports is to identify and keep track of sources where data can be accessed quickly or data is generated, compiled or published regularly. For example, the accidents data recorded by Rescue 1122 emergency services.

6. LEARN SIMPLE CALCULATIONS

For daily stories, broadcast journalists can rely on number comparisons to contextualise and verify claims made by politicians and government departments.

- The common comparisons can be with the help of percentages or per capita, rates of change, simple difference or calculating averages.
- In order to do this, the broadcast journalists can learn some simple calculations using spreadsheets.
- For example, if the police department publishes murder statistics, a broadcast journalist can divide the number by the population of the city and multiply by 1,000 to determine how many murders were committed per 1,000 people. The simple calculation can add an impactful headline or data point to the story and help the journalist rise above their news competitors.

AMANDA FARNSWORTH, THE HEAD OF VISUAL AND DATA JOURNALISM AT BBC, SUGGESTS THAT DATA JOURNALISTS AND DATA TEAMS AT BROADCAST NEWS ORGANISATIONS CAN BE GUIDED BY THE PRINCIPLES OF DISTINCTIVENESS, LIVELINESS, AND UNDERSTANDING. SHE RECOMMENDS THAT VISUAL GRAPHICS ARE ONE WAY TO HELP BROADCAST DATA JOURNALISM STAND OUT AND THESE ALSO RESONATE WITH AUDIENCES THAT ARE CONSUMING CONTENT ON MOBILE DEVICES.

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