DATA JOURNALISM
PROFESSIONAL ACADEMIC NOTES
ORDINARY DIPLOMA IN JOURNALISM

PART I: DATA JOURNALISM CONCEPTS

Data refers to facts, figures, observations, measurements, records, or other pieces of information that can be collected and organized for analysis. Data may appear as numbers, words, dates, categories, geographic information, survey responses, photographs, records, or digital files.

Examples of data include population figures, school enrolment records, hospital attendance figures, rainfall measurements, census results, prices of goods, crime statistics, business records, and government budget figures.

In journalism, data becomes useful when a journalist examines it to identify facts, patterns, comparisons, changes, relationships, or trends that can help explain a story.

1.2 Meaning of Data Journalism

Data journalism is a form of journalism in which data is used as an important part of discovering, investigating, verifying, analysing, and communicating news and feature stories.

Data journalism does not mean simply putting numbers into a news story. The journalist must identify relevant data, assess its quality, organize it, analyse it, interpret what it shows, and communicate the findings clearly to the audience.

In practical terms, data journalism connects three activities: obtaining reliable data, finding meaningful information within the data, and turning those findings into understandable journalism.

A data-driven story may use a spreadsheet, government report, survey dataset, public database, research report, or records obtained through reporting. The journalist then asks questions about what the data shows and why the findings matter.

1.3 Data Journalism and Traditional Journalism

Traditional journalism often begins with a person, event, statement, interview, observation, or document. Data journalism can begin with a dataset or a numerical pattern. The two approaches can work together. Data can strengthen traditional reporting, while interviews and field reporting can provide context for data.

For example, a journalist may discover from education statistics that school attendance has declined in a particular area. Interviews with teachers, parents, learners, and education officials can then help explain why the change occurred.

1.4 Data Journalism in News Writing

In news writing, data can help a journalist report measurable developments precisely. Data may provide evidence for the scale, frequency, location, cost, or trend associated with an event.

For example, instead of writing only that "food prices have increased," a journalist can examine price records and report specific changes over a defined period. The figures provide evidence, while interviews can explain how the changes affect consumers and traders.

1.5 Data Journalism in Feature Writing

Feature writing allows more space for explanation, context, background, human experiences, and interpretation. Data can therefore establish the scale of an issue and support a deeper narrative.

For example, a feature about youth unemployment may combine labour-force statistics with interviews with young job seekers, employers, economists, training institutions, and government officials.

1.6 Advantages of Data Journalism

Evidence-based reporting

Data can provide measurable evidence that supports a journalistic claim rather than relying only on opinions.

Identifying trends

Data can reveal changes over time and help journalists compare periods, locations, groups, or categories.

Revealing patterns

Large datasets may contain relationships or differences that are not immediately visible.

Supporting investigative journalism

Data can help journalists examine public spending, contracts, budgets, service delivery, business activities, environmental records, and other documentary evidence.

Improving precision

Specific, verified figures can make reporting more precise.

Comparing situations

Data allows journalists to compare locations, periods, groups, or institutions using consistent measures.

Explaining complex issues

Well-selected data can simplify complex subjects by showing their scale and structure.

Supporting accountability

Records and datasets can help journalists examine institutional performance and use of resources.

1.7 Challenges of Data Journalism

Limited access to data
Useful datasets may be difficult to obtain, inaccessible, incomplete, or available only in difficult formats.

Poor data quality
Data may contain missing values, errors, duplicates, inconsistent definitions, or outdated information.

Limited technical skills
Data journalism may require spreadsheet, data-cleaning, calculation, database, visualization, and digital research skills.

Misinterpretation of statistics
Figures can be misleading if journalists use inappropriate comparisons or confuse correlation with causation.

Large and complex datasets
Large datasets require appropriate methods to filter, organize, and analyse relevant information.

Time constraints
Data collection and analysis can take longer than ordinary reporting.

Data bias
The method and definitions used to collect data can influence the findings.

Difficulty explaining findings
Technical results must be translated into clear language for a general audience.

PART II: PRINCIPLES AND TECHNIQUES OF DATA JOURNALISM

2.1 Principles of Data Journalism

Accuracy
Figures, dates, units, calculations, labels, and sources should be checked carefully.

Verification
Journalists should establish where data came from, how it was collected, when it was collected, and whether it is appropriate.

Transparency
Where relevant, explain the source, important limitations, definitions, and methods.

Relevance
Selected data should help answer the journalistic question.

Context
Figures should be explained in relation to time, population, location, and definitions.

Fairness
Data should not be selectively presented to create an unjustified impression.

Independence
Journalists should analyse data professionally rather than manipulating findings to support a predetermined conclusion.

Clarity
Findings should be communicated in language the intended audience can understand.

Accountability
Journalists should be prepared to explain their methods and correct significant errors.

2.2 Data Journalism Techniques
- Data collection: obtaining relevant data from appropriate sources.
- Data cleaning: identifying and handling errors, duplicates, missing values, and inconsistent entries.
- Data organization: arranging information into a usable structure such as rows and columns.
- Data filtering: selecting records relevant to the journalistic question.
- Sorting and ranking: arranging values to identify highest, lowest, earliest, latest, or other meaningful categories.
- Calculation: using appropriate arithmetic or statistical calculations.
- Comparison: examining differences between groups, locations, periods, or categories.
- Trend analysis: examining changes over time.
- Data visualization: presenting findings through charts, graphs, maps, or other suitable visual forms.
- Cross-checking: comparing data with other reliable sources.
- Interpretation: explaining what the analysed data means in a journalistic context.

2.3 Organizing Data for a News Story
Once data has been collected, a journalist should organize it around a clear reporting question. The journalist should avoid collecting large amounts of information without knowing how it will answer the story's central question.

1. Define the story question.
2. Identify the data needed to answer the question.
3. Obtain the data from appropriate sources.
4. Check the source, date, definitions, and quality of the dataset.
5. Clean and organize the data.
6. Filter information relevant to the story.
7. Calculate or compare figures where necessary.
8. Identify significant findings or patterns.

9. Cross-check important findings.
10. Find human and expert sources to provide context.
11. Write the story clearly and accurately.
12. Use an appropriate table, chart, or other visual form when necessary.

2.4 Selecting an Appropriate Data Technique

You want to know whether a value has increased over five years. — Appropriate approach: Use trend analysis and comparison over time.

You want to identify which district has the highest number of cases. — Appropriate approach: Sort or rank the data and compare districts.

You want to show the distribution of values among categories. — Appropriate approach: Use a suitable table or chart.

You want to identify missing or repeated records. — Appropriate approach: Use data cleaning and quality checks.

You want to compare two groups. — Appropriate approach: Use comparable measures and clearly define the groups.

PART III: SOURCES OF DATA FOR NEWS STORIES AND FEATURES

3.1 Meaning of a Data Source

A data source is the person, institution, document, database, record, survey, observation, or system from which data is obtained. The supplied curriculum specifically requires learners to identify sources of data for writing news stories and features for print media and to differentiate primary from secondary sources.

3.2 Common Sources of Data for Journalism

- Government ministries, departments, and agencies.
- National statistical and census publications.
- Local government records.
- Hospitals and health institutions.
- Schools, colleges, and universities.
- Research institutions and academic publications.
- Non-governmental and civil-society organizations.
- Courts and official legal records.
- Company and business records where accessible and appropriate.
- Public budgets, reports, procurement records, and official documents.
- Surveys and questionnaires conducted by journalists or researchers.
- Interviews and observations.
- Public databases and official websites.
- International organizations and research bodies.

3.3 Primary Data Sources

Primary data is data collected directly for a specific research or reporting purpose. The journalist or researcher obtains the information first-hand rather than relying solely on data previously collected by another organization.

- A journalist conducting a survey of residents.
- Interviews conducted specifically for a story.
- Direct observation and recording of an event.
- A questionnaire administered to a selected group.
- Measurements collected directly in the field.

3.4 Secondary Data Sources

Secondary data is data that has already been collected, processed, or published by another person, institution, or organization and is then used by the journalist for another purpose.

- Government statistical reports.
- Published census results.
- Research reports.
- Academic studies.
- Annual reports.
- Previously published surveys.
- Institutional databases and official records.

3.5 Difference Between Primary and Secondary Data

Aspect	Primary Data	Secondary Data
Source	Collected directly by the journalist/researcher or for the specific study.	Previously collected by another person or institution.
Purpose	Collected for a particular reporting or research question.	Originally collected for another purpose, although it may be useful for new reporting.
Examples	Interviews, questionnaires, direct observation, field measurements.	Census reports, published studies, government reports, institutional records.
Control	The collector has greater control over how the data is collected.	The journalist has less control over the original collection process.
Limitation	May require more time and resources to collect.	May not exactly match the journalist's question or may

		contain limitations inherited from the original collection.

3.6 Evaluating Data Sources

- Who collected the data?
- Why was the data collected?
- When was it collected?
- How was it collected?
- What population or area does it represent?
- What definitions were used?
- Are there missing records?
- Are there possible errors or biases?
- Can the figures be independently checked?
- Is the data current enough for the story?

PART IV: USING DATA SOURCES TO WRITE NEWS AND FEATURES FOR PRINT MEDIA

4.1 From Data to a News Story

A journalist should transform data into a clear story rather than simply reproduce a dataset. The first step is to identify the central finding and then provide evidence, context, and relevant human perspectives.

- Headline: communicate the main verified finding clearly.
- Lead: present the most important finding.
- Evidence: provide the key figures supporting the finding.
- Context: explain what the figures mean and why they matter.
- Sources: identify the origin of the data and relevant human sources.
- Analysis: explain significant comparisons or trends without exaggeration.
- Human perspective: where appropriate, include relevant experiences or voices.
- Response: seek relevant institutional or individual responses where the story involves claims or criticism.
- Conclusion: provide the final relevant fact, context, or implication.

4.2 Example of Data-Based News Writing

Suppose a local education dataset shows that enrolment in a group of schools increased over a three-year period.

Weak approach: “School enrolment is doing very well.”

Stronger approach: “Enrolment in the surveyed schools increased from X learners to Y learners between the two reporting years, according to the education records.”

The journalist should then explain the source of the figures, verify the calculation, investigate reasons for the change, and obtain relevant responses from education stakeholders. Actual figures should only be inserted after verification.

4.3 Data in Feature Stories

A feature can use data to establish the scale of an issue while using human stories to show how the issue affects individuals.

For example, a feature about water access may use official statistics to show the proportion of households with access to clean water, then use interviews with residents and service providers to explain the daily experience behind the figures.

The journalist should avoid using statistics merely to make a feature appear academic. Every figure should contribute to the story's explanation.

4.4 Presenting Data in Print Media

- Use a table when exact values across several categories need to be compared.
- Use a line chart when showing change over time is the main purpose.
- Use a bar chart when comparing categories or groups.
- Use a map when geographic location is central to the story.
- Use an infographic when several related findings need to be communicated visually.

4.5 Common Errors in Data-Based Reporting

- Publishing figures without checking the original source.
- Using outdated data without explaining its date.
- Comparing figures produced using different definitions.
- Confusing percentage with percentage-point change.
- Ignoring missing data.
- Treating correlation as proof that one factor caused another.
- Selecting only figures that support a preferred narrative.
- Using a graph that exaggerates differences.
- Failing to identify the population or period covered by the data.
- Writing a conclusion that goes beyond what the data can support.

PART V: PRACTICAL DATA JOURNALISM WORKFLOW

1. Identify a newsworthy question

Start with a clear question that can be investigated using evidence.

2. Identify relevant data

Determine what figures, records, or measurements are needed.

3. Find appropriate sources

Locate reliable primary or secondary data sources.

4. Obtain and document the data
Keep records of where the data came from and when it was obtained.

5. Examine data quality
Check definitions, dates, missing information, errors, and limitations.

6. Clean and organize
Prepare the data for analysis.

7. Analyse
Calculate, compare, filter, sort, and examine patterns as appropriate.

8. Verify findings
Cross-check important findings and investigate unexpected results.

9. Report and contextualize
Combine data with relevant interviews, documents, observations, and expert explanation.

10. Write and edit
Present findings clearly and accurately for the intended audience.

11. Visualize where useful
Use an appropriate table, chart, map, or infographic when it improves understanding.

12. Review and correct
Check the final story and correct significant errors when identified.

Essay and Discussion Questions

13. Define data journalism and explain its role in news and feature writing.
14. Discuss the advantages of using data journalism in news reporting.
15. Explain the major challenges faced by journalists when using data.
16. Discuss the principles that should guide journalists when gathering and analysing data.
17. Explain the process a journalist should follow when turning a dataset into a news story.
18. Differentiate primary and secondary data and give relevant journalism examples.
19. Discuss major sources of data available to journalists writing news and feature stories for print media.
20. Explain how data can be combined with interviews and human-interest information in feature writing.
21. Discuss common mistakes journalists should avoid when reporting statistics.
22. Explain how tables, charts, maps, and infographics can improve data-based journalism.