

Methodological Approaches to Working with Open Data in Data Journalism

Olzhas Abdikhalyk¹, Aliya Beldibekova², Manchuk Mukasheva¹, Hulya Celik³

¹Faculty of Journalism, Al-Farabi Kazakh National University, Almaty, Kazakhstan

²Scientific Center «Yasaitanu», National Historical and Cultural Museum-Reserve «Aziret-Sultan», Turkestan, Kazakhstan

³Faculty of Education, Sakarya University, Hendek, Sakarya, Turkey

Correspondence: Aliya Beldibekova, Scientific Center «Yasaitanu», National Historical and Cultural Museum-Reserve «Aziret-Sultan», Turkestan, Amir Temir St., 53A, Kazakhstan.

Received: December 18, 2025

Accepted: April 7, 2026

Online Published: April 12, 2026

doi:10.11114/smc.v14i3.8300

URL: <https://doi.org/10.11114/smc.v14i3.8300>

Abstract

Journalism, both as an academic discipline and professional practice, is fundamentally grounded in data. The increasing prominence of data journalism underscores the field's adaptation to digital transformation, the growing availability of data and shifting audience expectations. This article explores methodological approaches to working with open data in data journalism, with particular focus on the practical experience of Kazakhstan. The study tackles a central conceptual question: if journalism itself is inherently data-driven, what rationale exists for distinguishing data journalism as a separate methodological domain?

The article analyzes the development of data journalism in Kazakhstan within the context of the country's open data policy and digital governance initiatives. Methodologically, the research focuses on main stages of working with open data in journalism: data discovery, evaluation of data quality, cleaning and structuring datasets, analytical interpretation, and visual presentation.

The findings highlight that effective data journalism in Kazakhstan depends on interdisciplinary collaboration between journalists, data analysts, and IT specialists, as well as on the development of data literacy within newsrooms. The article argues that open data serves as a methodological foundation for data journalism, enabling evidence-based reporting and strengthening the democratic function of the media. By contextualizing global theoretical perspectives within Kazakhstan's media environment, this study contributes to a deeper understanding of data journalism as both a methodological practice and a socially significant form of journalism.

Keywords: data journalism, visualization, big data, infographics, open data

1. Introduction

In recent years, the rapid development of artificial intelligence and algorithmic technologies has significantly transformed journalistic practices and approaches to information production. Data has become a central resource in the contemporary media ecosystem, shaping not only how information is collected and processed, but also how it is interpreted and presented to the public. In this context, data-driven journalism has gained increasing importance as a methodological response to the growing volume, complexity, and accessibility of digital information.

The widespread adoption of artificial intelligence tools, automated data analysis systems, and machine learning algorithms has expanded journalists' capacity to work with large-scale datasets. These technologies enable the detection of hidden patterns, trends, and correlations within data that would be difficult or impossible to identify through traditional reporting methods alone. As a result, data journalism increasingly relies on structured data, open data platforms, and computational tools to produce evidence-based reporting that supports transparency, accountability, and informed public discourse.

At its core, both traditional journalism and data journalism share a common objective: the dissemination of reliable information in public interest. The key distinction lies not in purpose, but in the methods of knowledge production and representation. While traditional journalism to emphasize immediacy, narrative framing and emotional engagement, data journalism prioritizes analytical rigor, verification and visual interpretation of complex datasets. Experts and researchers

assert that data journalism often minimizes subjective interpretation by grounding narratives in verifiable data, frequently utilizing visualization techniques in place of purely textual storytelling. However, this distinction should not be regarded as absolute, as effective journalism frequently integrates narrative, data analysis and visualization to enhance audience comprehension. Journalists working in data journalism do not necessarily view their work as a distinct branch of professional journalism. This perspective is well illustrated by American web developer Adrian Holovaty's response to the question "Is this data journalism?" - to which he replied, "Who cares?" (Holovaty, 2009). As a result, various terms have emerged and are often used interchangeably within this field.

In the 1990s, the concept of Computer-Assisted Reporting (CAR) was well-defined and widely understood among professionals and researchers (Coddington, 2015). Over time, however, the field has fragmented into different methodologies under various names. Researchers describe it using terms such as computational journalism (Terry, Spurgeon, Daniel & Swift, 2012), programmer journalism (Sylvain & Dagiral, 2013), open-source journalism (Lewis & Usher, 2013), and data journalism (Appelgren & Nygren G, 2014), among others.

For researchers, precise definitions of these terms are crucial. Clear definitions are essential for understanding the professional and cultural significance of these practices. Without them, conducting comprehensive scientific research in this field becomes challenging. Current studies mostly attempt to define this branch of journalism, but the similarities between various definitions often create confusion and hinder the establishment of a solid theoretical foundation.

As data-driven journalism becomes increasingly significant within professional journalism, it is imperative to distinguish between these terms rather than using them synonymously. Each of these fields has its unique characteristics and analyzing how they influence changes in journalism is of great importance.

Although the term "data journalism" is less commonly used in academic discourse, it is frequently employed in the news industry to describe journalism that relies on data analysis and its presentation. According to Jonathan Stray's research, data journalism can be defined as "the collection, investigation, processing, and publication of data in the public interest" (Stray, 2011).

Some researchers consider data journalism as an intersection of multiple fields and methodologies, incorporating elements of statistical analysis, computer science, visualization, web design, and traditional reporting (Bell, 2012). Additionally, Astrid Gynnild argues that data journalism is closely associated with open data and open-source tools (Gynnild, 2014).

Data journalism began to gain momentum in the late 2000s. Prior to this, data analysis was predominantly conducted within the framework of Computer-Assisted Reporting (CAR) or by news organizations focused in financial reporting. While data journalism is now widely adopted across the field, the number of journalists specializing in this area remains relatively limited.

With globalization and the continuous advancement of information technologies (IT), journalism is evolving to align with contemporary demands, giving rise to new approaches to information dissemination. As a result, data journalism holds a unique place in the Information Age. The need to study the genres of data journalism and its methods of dissemination has become increasingly relevant today.

2. Method

This study adopts a qualitative-dominant mixed methodological approach to explain how open data is methodologically employed in data journalism, with a particular focus on the Kazakhstani media environment. The research design integrates document analysis, case study methodology, and comparative analytical techniques in order to capture both the structural and practical dimensions of working with open data in journalistic practice.

2.1 Dataset

The empirical dataset for this study was compiled using non-intrusive data collection methods and consists solely of publicly available materials. The primary data sources include:

National open data platforms of Kazakhstan, particularly datasets published on the official Open Data Portal of the Republic of Kazakhstan (data.egov.kz), including datasets related to public finance, healthcare, social policy, education, and governance.

Official statistical databases, such as publications and downloadable datasets from the Bureau of National Statistics of Kazakhstan.

Publicly accessible governmental reports and analytical documents, released by ministries, akimats, and national agencies.

Data-driven journalistic publications, produced by Kazakhstani media outlets and independent fact-checking platforms (including Factcheck.kz and Stopfake.kz), where open data was explicitly used as the basis for reporting.

Supplementary social media data, used selectively to contextualize journalistic investigations, particularly in cases related to misinformation and public discourse.

The dataset covers materials published between 2016 and 2026, corresponding to the period of active development of data journalism in Kazakhstan. In total, over 20 data-driven journalistic projects and 30 open datasets were selected for in-depth analysis based on their relevance, completeness, and demonstrable use of open data methodologies.

2.2 Materials

The materials analyzed in this research include structured datasets (CSV, XLSX), statistical tables, interactive dashboards, visualizations (maps, charts, infographics), and accompanying explanatory texts published by media organizations. Data collection and examination were conducted using widely adopted analytical and visualization tools commonly employed in journalistic practice, including Microsoft Excel, Tableau, Python-based data analysis libraries, and GIS tools for spatial data interpretation.

In addition, methodological guidelines, newsroom manuals, and training materials related to data journalism and open data usage were reviewed to contextualize professional standards and workflows. The analytical process was informed by thematic analysis principles (Braun & Clarke, 2006, 2019), adapted to the study of methodological practices rather than audience discourse.

2.3 Procedure

The research procedure was carried out in several sequential stages. First, an exploratory review of the availability and accessibility of open data in Kazakhstan was conducted. This stage involved identifying key national data repositories, assessing their structure, update frequency and usability for journalistic purposes. Attention was given to metadata quality, licensing conditions and data formats, as these factors directly impact the usability of data for journalistic work.

Second, a purposive sampling strategy was employed to select data journalism cases that explicitly relied on open data. Selected projects were required to (a) reference open data sources, (b) demonstrate analytical processing of datasets, and (c) include some form of data visualization or structured presentation. Projects that merely cited numerical data without evidence of data processing or analysis were excluded.

Third, each selected case was analyzed using a standardized analytical framework, focusing on the core stages of data journalism methodology:

- data discovery and sourcing,
- evaluation of data quality and reliability,
- data cleaning and structuring,
- analytical interpretation, and
- visual representation and storytelling.

During this stage, recurring methodological patterns, tools, and challenges were systematically documented. The role of interdisciplinary collaboration between journalists, data analysts, and IT specialists was also assessed based on project descriptions and workflow transparency.

Fourth, a comparative analysis was conducted between Kazakhstani practices and internationally recognized data journalism standards described in academic literature. This allowed for the identification of both context-specific characteristics and methodological gaps.

2.4 Analytical Framework

The analysis was guided by a methodological model that conceptualizes data journalism as an integrated process combining journalistic inquiry with computational data practices. Open data was treated not merely as a source of information but as a methodological foundation shaping the entire reporting process.

Analytical attention was directed toward:

- the extent to which open data enables evidence-based reporting,
- the balance between transparency and interpretative storytelling,
- the limitations imposed by incomplete, outdated, or poorly structured datasets,
- and the implications of data-driven reporting for public accountability and anti-disinformation efforts.

Rather than focusing solely on the aesthetic quality of visualizations alone, the study emphasized their epistemic function—how effectively they communicate verified insights and support journalistic claims. This approach enabled a systematic understanding of data journalism in Kazakhstan, not only as a methodological practice but also as a socially significant form of journalism. Against the backdrop of the rapid advancement of modern technologies, the scope of data journalism has expanded significantly. The accessibility of data has markedly improved in comparison to previous decades, enabling journalists to harness vast amounts of information more efficiently than ever before.

The emergence of various computer programs has reduced the excessive workload of journalists in collecting and processing information. As a result, journalists can now access data from multiple sources while saving time in their search for truth. Importantly, professionals engaged in data journalism must have high computer literacy, as data can only be analyzed when presented in a format that computers can process.

The collaborative efforts of creative teams contribute to an increase in the number of reliable materials, as multiple sources require collective effort to gather information on a single topic. In this process, a journalist must ensure that the information remains relevant to the subject, carefully filtering all data, distinguishing between essential and excessive details, and identifying key insights.

In reality, data journalism, which is rapidly developing in the United States, does not merely present numbers but reveals stories hidden within diverse datasets and delivers them in a multimedia format. According to researcher Sherstyukova, data journalism represents a new genre and direction, providing the public with open information through maps and infographics. Meanwhile, journalists Skorobogatko and Ruderman argue that the foundation of this term lies in infographics. Thus, data journalism can be described as a compelling presentation that integrates numbers and illustrations (Erkimbay, 2017).

The increasing speed of information dissemination has driven the evolution of journalism in a new direction. Since data journalism is based on facts and figures, a strong emphasis is placed on the accuracy of information. Today data journalism is recognized as a discipline that involves the collection, filtering and visualizing billions of vast quantities numbers and datasets. This field is characterized by its to present complex information in a timely and easily comprehensible format, thereby serving the public interest by enhancing transparency, accountability, and informed decision-making. Currently, data journalism relies on new methods, including electronic spreadsheets and specialized programs, to organize, filter, and process information. These approaches enable journalists to uncover new perspectives and insights within datasets.

Data journalism has its roots in investigative journalism, with its primary objective being to analyze vast amounts of online data and explain the essence of events. A modern journalist must not only conduct analysis and evaluations but also extract compelling stories from complex numerical data. To achieve this, they primarily rely on computer-assisted tools.

Moreover, data journalism has become an increasingly complex and significant field due to advancements in digital technologies. Therefore, it is impossible to separate data journalism from digital journalism. While each has distinct features, they complement each other and form a unified whole. In today's digital world, all essential data is available.

New media, including the internet, mobile phones, and social networks, have gained greater prominence over traditional mass media. The main task of journalists and news editors is to adapt to and understand technological changes.

Well-designed visual elements enhance the communicative power of text and, in some cases, can even replace it entirely. Visual elements are often more effective than words in conveying the intended message, as they not only provide information but also add additional semantic meaning. These elements must be carefully planned and closely integrated with the text. While it is not necessary to be a professional designer to use visual representations effectively, understanding fundamental design principles is crucial.

Commonly used charts and graphs in journalism have limitations. They can only display easily explainable parts of data, whereas more complex relationships require data visualization. Data visualization is not just about explaining information but also about highlighting crucial insights within large volumes of data (Popova, 2009).

Interactive applications allow users to analyze datasets, leading to new insights or the retrieval of necessary information. Visualization tools are widely utilized in mobile applications as well. Infographics serve as a simple yet effective means of presenting information, offering all the possibilities of storytelling. They can take various forms, ranging from charts and tables to structured reports.

The development of data journalism is closely tied to the growing demand from audiences for more transparent and accurate information. In Kazakhstan, this demand is steadily increasing, with the audience for data-driven reporting expanding as well. A key factor driving this shift is the rising level of public awareness concerning the authenticity and reliability of data. This heightened scrutiny has accelerated the adoption of data journalism. Furthermore, data visualization techniques, such as charts and tables, play a crucial role in presenting information in a clear and effective manner, ensuring that complex data is accessible and engaging for diverse audiences.

Leading universities and journalistic organizations in Kazakhstan are contributing to the development of data journalism by offering specialized courses and programs. As a result, this field is expected to expand significantly in the future, providing journalists with new opportunities.

3. Result

Let us now undertake a detailed analysis of open data sources in Kazakhstan, recognizing the crucial role the internet

plays in this process. One of the key mistakes journalists often make when gathering information is their inability to effectively locate primary data sources. This oversight can have serious implications, especially when investigating critical issues such as budget allocation or criminal activities. By failing to access and utilize these primary sources, journalists risk compromising the accuracy and depth of their reporting. When conducting research on important topics, journalists have the opportunity to find, process, and analyze key data from open internet platforms, which provide a wealth of reliable and publicly accessible information. The ability to navigate and extract valuable insights from these sources is, in itself, a skill that requires expertise. Therefore, by examining Kazakhstan's open data sources, we can identify the most efficient and rapid methods for obtaining relevant data across a range of topics. This exploration will highlight the tools and techniques necessary to leverage open data effectively, empowering journalists to produce more informed and impactful stories.

Table 1. Key Data Sources for Data Journalists

Entity	Documents of Interest
Government	
Ministry of National Economy of the Republic of Kazakhstan, "Economic Research Institute" JSC	Press releases, comments, reports and studies, forecasts on socio-economic research, minister's reports to Parliament and Government, monitoring of state programs
Ministry of Finance and its committees	Draft national budget, reports on allocation and execution of budget funds, minister's report on budget execution, press releases from ministries and committees
Central Government Bodies	Ministers' reports at Government and Parliament meetings, press releases, annual reports on program activities
Representative / Legislative Bodies	
Committees on Finance and Budget of the Mazhilis and Senate of the Parliament of the Republic of Kazakhstan	Committee conclusions on the national budget, opinions of committee chairs and members, press releases
Profile Committees of the Mazhilis and Senate of the Parliament of the Republic of Kazakhstan	Committee conclusions on various sectors financed from national and local budgets, press releases, decisions on adoption of draft laws requiring state funding
Finance and Budget Committees of Maslikhats	Committee conclusions on local (regional, city, and other) budgets
Profile Committees of Maslikhats	Committee conclusions on activities of various sectors financed from national and local budgets
External Oversight Bodies	
Accounts Committee for Control over Execution of the National Budget	Annual and quarterly reports of the committee, committee conclusions on national budget execution, committee chair's reports at Parliament meetings, press releases
Maslikhat Audit Commissions	Annual and quarterly reports of the audit commission, press releases, and reports from the audit commission chair
Non-Governmental Organizations and Citizens	
NGOs and Budget Networks	Reports on monitoring of national and local budget programs, analysis of relevant budgets, press releases, expert opinions
Public Councils	Comments
Research Institutes	Reports, monographs, dissertations, press releases, expert opinions
Independent Experts	Comments
Citizens	Comments and facts regarding the provision/non-provision of public services under budget programs

The Guardian newspaper launched a project in 2009–2010 to investigate the expenses of members of the UK Parliament. They published 460,000 pages of expense reports and asked readers to help identify suspicious expenditures. As a result, several parliament members were forced to review their expenses and return certain funds. This project became a prime example of data journalism, as it made open data accessible to the public and enabled readers to participate in data analysis and reporting.

According to Mark Coddington, it is common in data journalism to openly present datasets to the audience or create tools for their exploration. Sometimes, the processing and analysis of these datasets are delegated to the audience itself (crowdsourcing). For instance, by allowing people to verify data, journalists aim to involve the public in the process rather than managing it entirely on their own.

Editorial choices remain crucial in data journalism. Journalists analyze data and present it in a comprehensible manner, but audience participation in this process is also increasing. Thus, data journalism represents a collaboration between professional and citizen journalism.

There is no doubt that data journalism has a significant impact on research, analysis, and the fight against misinformation. In Kazakhstan today, several initiatives are being developed to deliver data-driven information, attracting public attention. One such initiative is factcheck.kz, founded by journalist Adil Jalilov. Ukrainian journalist Oleksandr Horokhovskiy describes Factcheck as "a modern research tool with a unique concept and technical specifications, aimed at exposing populism and manipulations in the media, as well as identifying questionable statements made by speakers at various levels" (Gorokhovskiy, 2017).

Indeed, data journalism is a powerful weapon against misinformation and disinformation. Fake news is often created to undermine public trust in socially significant issues, political campaigns, or prominent figures and is deliberately spread online. Additionally, data journalism plays an essential role in monitoring the fulfillment of promises made by government officials, the president, and other public figures, ensuring accountability over time. Misinformation remains a primary adversary of journalism, and in combating it, fact-checking, verifying various informational resources, and uncovering the truth through data analysis play a vital role.

In essence, data journalism thrives in an open society where government departments and business entities do not restrict access to data. Since 2008, there has been a global movement advocating for open data. Many countries have established open data portals that provide diverse datasets, both in terms of quantity and quality. In 2008, U.S. President Barack Obama signed three presidential memoranda on transparency and open government. Open data refers to publicly available government data hosted on platforms for further use and reproduction by citizens and businesses. Major international organizations such as the UN, EU, OSCE, World Bank, and OGP (Open Government Partnership) actively support open data initiatives. Government open portals have been established in 23 countries, including the U.S., UK, Singapore, France, Canada, Australia, Denmark, and others. The leading platforms include the UK's data.gov.uk, the U.S.'s data.gov, and Singapore's data.gov.sg. European nations have demonstrated successful implementations of open data projects. In Canada, access to open data facilitated an investigation into fraudulent activities within U.S. charitable foundations, uncovering losses amounting to \$3.2 billion.

When conducting journalistic investigations on broad and complex topics such as Kazakhstan's role on the global political stage, environmental protection or healthcare issues, open data repositories worldwide become valuable resources. While national open data sources offer a wealth of information, certain data concerning Kazakhstan or its citizens may not be readily available within domestic platforms. In such cases, international open data platforms provide a crucial alternative for journalists seeking comprehensive and diverse datasets.

Global open data holds significant importance in data journalism, as it allows journalists to access a broader spectrum of information that may not be captured within local sources. These international platforms offer a wealth of comparative data, making it possible to analyze and contrast Kazakhstan's performance and policies with those of other countries. Utilizing these international databases is both straightforward and convenient, enabling journalists to conduct comparative studies and draw more nuanced conclusions about global trends, issues, and best practices. In this way, global open data enhances the depth and scope of investigative journalism, facilitating a more informed and globally contextualized approach to important issues.

Table 2. International Databases

Database Name	Useful Links	Offered Data
World Bank Open Data	data.worldbank.org	Global demographic indicators, environment, poverty levels, education statistics, economic growth, and financial trends
United Nations Open Data	data.un.org	Global economic and social indicators, environmental statistics, infrastructure, health, and development trends
Stockholm International Peace Research Institute (SIPRI)	sipri.org/databases	Arms trade data: which country buys/sells weapons, weapon types, purchase dates, and financial expenditures
World Trade Statistics	wto.org/statistics	International trade flows, export-import balances, tariffs, and market access data
U.S. Federal Procurement Data System	fpds.gov	Government procurement data, contract awards, and expenditure tracking for federal agencies
UNESCO Institute for Statistics (UIS)	uis.unesco.org	Data on education, scientific research, cultural heritage, communication, and media development
UNODC Crime and Justice Statistics	dataunodc.un.org	Global crime statistics, including homicide rates, human trafficking, drug-related offenses, and prison populations
Offshore Leaks Database	offshoreleaks.icij.org	Information on offshore accounts, tax haven structures, and international financial secrecy
UK Companies House	find-and-update.company-information.service.gov.uk	Public registry of UK businesses, corporate filings, and financial reports
New York Department of Finance (Open Corporates, Investigative Dashboard)	opencorporates.com	Global corporate ownership, business transparency, and investigative tools for financial crime tracking
Miami-Dade Property Appraiser	miamidade.gov/pa	Public property records, ownership history, valuation data, and real estate transactions
International Budget Partnership	internationalbudget.org	Transparency rankings, budget reports, and government expenditure analysis
International Trade Centre (ITC)	intracen.org	Trade statistics, import/export data, and market access information
UN Comtrade	comtrade.un.org	Global trade data, bilateral trade flows, and commodity trade breakdowns
World Bank Enterprise Surveys	enterprisesurveys.org	Business environment analysis, firm performance data, and economic development indicators
World Bank Doing Business	doingbusiness.org	Ease of doing business rankings, regulatory frameworks, and market entry conditions
World Trade Organization (WTO) Database	stat.wto.org	Tariff structures, trade policies, global market trends, and trade agreement compliance

Fact-checking is a crucial component of data journalism and plays a vital role in ensuring information security and authenticity. Verifying the accuracy and correctness of information is conducted to ensure that news is delivered to the audience in a precise and objective manner. This process is particularly important in research based on statements made by public figures or official sources.

For instance, platforms like Factcheck.kz in Kazakhstan enable the analysis of various types of information. The core principle of this resource is verifying the accuracy of facts presented in news reports and statements. A notable example is the article titled “Is Kazakhstan Ranked Third in the World for Abortion Rates?” by writer Medet Esimkhan, published on April 11, 2021. The article investigates the accuracy of data shared on Instagram, which claimed that in 2021, Kazakhstan had an abortion rate of 35 per 1,000 women, positioning the country third globally in abortion prevalence

(Yesimkhanov, 2021). This information was subsequently published by major news portals in Kazakhstan, such as [nur.kz](#), [times.kz](#), and [newtimes.kz](#), expanding public discourse on abortion statistics in the country. However, a [Factcheck.kz](#) investigation revealed that this information was outdated and that actual statistics had changed. The global demographic review website, which was cited as the primary source, contained outdated abortion rate data that had not been updated since 2008.

In this case, the author highlighted discrepancies between different sources. For example, the National Bureau of Statistics of Kazakhstan reported that in 2019, the abortion rate in the country was 16.8 per 1,000 women. This statistic reflects a declining trend over the past 20 years. Using these figures, the author clarified that previous information was significantly inaccurate and that the actual rate was much lower. This underscores the necessity of verifying news accuracy, as distorted or outdated data can shape public perception in a misleading way.

When discussing the significance of truthful information and the principle of responsibility, fact-checking emerges as one of the most critical steps (Erkimbay, 2019). Media theorists like Ekaterina Yakushina identify fact verification as a defining characteristic of reliable information (Yakushina, 2013). She emphasizes that relying on a single source can be dangerous, as information reliability should not be confirmed by just one data provider; rather, it should be corroborated by multiple trustworthy sources. Thus, when determining the completeness and accuracy of information, each source must be carefully examined. Media experts stress the importance of comparing data from two or more credible sources when verifying the authenticity of information.

David Brewer, in his article, discusses the "double-check" method in fact-checking. This classic technique involves verifying information through at least two independent sources, providing additional assurance of its authenticity (Brewer, 2018). Furthermore, Majid Rafizadeh suggests multiple strategies for information verification. He advocates questioning the original source of the information when its accuracy is uncertain and seeking expert opinions in the relevant field (Rafizadeh, 2014). Additionally, specialized services such as PolitiFact and FactCheck offer effective methods for verifying information accuracy online ([www.politifact.com](#)).

In Kazakhstan, a vast amount of information is verified through [Factcheck.kz](#), a project launched by the MediaNet International Journalism School with the support of the Soros-Kazakhstan Foundation. This platform scrutinizes various types of information, exposing manipulation and false claims. For example, the website has played a key role in debunking certain false narratives and determining the credibility of claims circulating in the media.

Ensuring the accuracy of information is of paramount importance in journalism, as false or misleading data can significantly impact public opinion and even harm individuals. Therefore, data journalism is not only about disseminating information but also about verifying facts and ensuring the delivery of reliable content. Accuracy and authenticity in reporting are fundamental responsibilities of journalists, aligning with their professional ethics.

Differences of Data Journalism

1. Data journalism is not limited to research; it is applied across various fields of journalism.
2. It places special emphasis on visualization, meaning that presenting data graphically plays a crucial role.
3. It treats readers as collaborative partners in understanding information and making ethical decisions.

Data journalism is not solely focused on investigative projects; it has found broader applications. Moreover, it considers readers an essential part of the process, allowing them to play an active role in data analysis and decision-making.

Data journalism preserves the core direction of CAR (Computer-Assisted Reporting), prioritizing the integration of data into narratives and stories. The main goal is to tell a story through data. In his research "The Art and Science of Data-Driven Journalism", Alexander Howard states: "Data journalism is about telling stories through numbers or finding a story within those numbers" (Howard, 2014).

A key feature of data journalism is that professional methods are not always required to uncover a story. For example, conducting interviews is not mandatory; simply analyzing data may be sufficient. This approach has made working with data accessible not only to journalists but also to anyone who understands data interpretation. Media researcher Liliana Bounegru describes this as follows:

"Data journalism allows anyone to explore data, find meaningful information, verify facts, and challenge commonly accepted assumptions. Previously, these methods were only available to professionals, but now they are accessible to everyone" (Bounegru & eds., 2012).

Principles of Data Visualization and Infographics

Before visualizing data, it is essential to determine responsibility for it.

- If the data contains unclear terms, we must request explanations or metadata from relevant organizations or experts.
- If there are difficulties or doubts in understanding numerical values, consulting industry experts is necessary.
- If the data describes events over time, it is best to present trends and changes.
- If the data includes maximum, minimum, or extreme values, bar charts are the most effective visualization method.
- When dealing with geographic data, cartographic visualization techniques should be used.
- Data is not limited to numerical indicators; in some cases, comparing past and present situations through visual materials (e.g., old and new photographs) is beneficial.
- When publishing data, it is crucial to cite its original sources.

Infographics are an essential tool in data journalism, serving to simplify and communicate complex information in a visually engaging format. Consequently, acquiring proficiency in infographic design at the early stages of journalistic training is highly beneficial. In this context, it is pertinent to emphasize the practical experience gained through the regional newspaper Zhana Turkistan, which has integrated infographics into its journalistic practices. This hands-on experience provides valuable insights into the design and application of data visualization, enhancing the ability to effectively convey information to a broad audience.



Figure 1. Zhana Turkistan newspaper. Infographic

As illustrated in Figure 1, the editorial team of Zhana Turkistan actively used infographic techniques to present epidemiological data during the COVID-19 pandemic. The newspaper regularly published visualized information on the spread of coronavirus infection in Kazakhstan, with a particular focus on the regional context of the Turkistan region. Statistical data on daily infection rates, recoveries, and other epidemiological indicators were systematically transformed into infographic formats and presented to readers on a daily basis.

This approach made it possible to simplify complex statistical information and present it in a clear and accessible visual form. By using charts, icons, and graphic elements, the newspaper ensured that readers could easily interpret current epidemiological trends at the local level. As a result, infographics served not only as a tool for visualizing numerical data but also as an effective means of communicating socially significant information during a public health crisis.

The experience of Zhana Turkistan demonstrates that the integration of data visualization into journalistic practice significantly enhances the accessibility and communicative effectiveness of media content, especially when reporting on rapidly changing and data-intensive topics such as the COVID-19 pandemic.

4. Discussion

Data journalism is aimed at informing society about important issues, but the role of the public in this process varies significantly. The primary objective of data journalism is to enable society to independently analyze and interpret data. In this context, the role of the data journalist is to access and present data on behalf of the public. Thus, the process of data journalism is structured to be beneficial to the user.

When developing data journalism products, such as data visualizations or web applications, ensuring their usefulness to the audience is of paramount importance (Stray, 2010). The public plays a direct role in data journalism because the goal of this process is to provide the public with tools that facilitate understanding and independent evaluation of societal issues. However, as researchers Fink and Anderson highlight, despite many data journalists aiming to serve an engaged public, their conceptualization of the audience is often depersonalized and assessed anonymously through online metrics (Fink & Anderson, 2014). In other words, within the journalistic process, the audience is seldom regarded as an individual entity or an interactive participant.

According to the study by Boyles and Meyer, data journalism offers several advantages over traditional journalism. By incorporating big data, digital technologies, visualization techniques, and computational methods, data journalism can achieve greater depth, objectivity, and accuracy. Additionally, it expands opportunities for audience engagement, allowing for more active participation in news production, distribution, and consumption (Boyles & Meyer, 2016). Data journalism seeks to establish a deeper connection between media and readers by actively involving the audience. However, audience participation in the content creation process can sometimes pose challenges to the professional credibility of journalism. Researchers Anderson and Borges-Rey attribute this concern to the growing influence of readers, which may weaken journalistic control over content (Anderson & Borges-Rey, 2019). Nonetheless, this approach has several notable advantages:

1. Expansion of sources: Increased audience engagement facilitates access to new and significant information for journalists.
2. Enhanced transparency: Audience participation helps journalists present their work in an open and comprehensible manner, thereby strengthening credibility.
3. Trust and integrity: Close interaction with the audience promotes the fair and honest representation of socially relevant issues.
4. Reinforcement of journalistic reputation: These advantages enable journalism not only to maintain but also to strengthen its credibility in the digital age.

Data journalism is a relatively new form of journalism, but a comprehensive understanding of it requires more than just viewing it as a technological tool. Anderson's perspective is particularly relevant here: data journalism should not merely be studied as a tool or methodology but as an integral part of the broader field of journalism (Anderson, 2013).

This implies that it is necessary to examine how data journalism interacts with other domains, such as traditional journalism or computer science, and to analyze the connections and contradictions that arise between them. Additionally, studying its role within societal and cultural contexts is essential.

Data journalism not only intersects with other disciplines but also has the potential to influence the broader field of journalism. For instance, by incorporating new technologies and analytical methods, it can enhance traditional journalistic practices or introduce innovative formats. Therefore, it should be considered within the broader framework of journalism rather than as a standalone discipline.

The development of data journalism in Kazakhstan has gained significant momentum over the past decade. This progress is directly linked to the advancement of digital technologies and the country's prioritization of digital transformation. Moreover, Kazakhstan places considerable emphasis on the IT sector. If this growth trend continues, it is likely that within the next five to ten years, data accessibility will significantly improve for information seekers.

Furthermore, Kazakhstan does not impose restrictions or barriers on the dissemination of information through social media, news websites, or other communication platforms, provided that the content complies with national legislation. This means that the scope for data discovery is broad for those seeking information. However, access to crucial datasets, such as "Public Procurement" records and governmental reports, remains challenging due to inadequate adherence to transparency obligations by data holders. Consequently, strict enforcement of Kazakhstan's "Law on Access to Information" is necessary. Additionally, information seekers must actively assert their rights to access public data.

At present, Kazakhstani higher education institutions lack educational resources authored by local scholars for training data journalists. While it is possible to educate data journalists using foreign instructional materials, Kazakhstan's approach to digital systems, transparency policies, and governmental operations possesses unique characteristics that warrant locally developed educational materials.

A key driver of the advancement of data journalism is the growing demand from audiences. In Kazakhstan, this audience base is expanding, driven by a broader societal shift towards heightened awareness and discernment in assessing the credibility of information. Furthermore, there is an increasing inclination among the public to approach information with a critical mindset, fostering a demand for more transparent, data-driven reporting that can withstand rigorous scrutiny.

With the rise of data journalism, data visualization has become crucial in capturing audience attention. Empirical evidence

suggests that presenting logically structured diagrams, charts, and tables enables the effective communication of information that would otherwise require two or three lengthy paragraphs. Based on existing research, data journalism provides new opportunities for reporters, such as collecting and integrating information from multiple sources, identifying relationships between institutional documents (with computational support), and deriving key insights from extensive datasets. Furthermore, it enables the creation of socially significant projects that assist the public in understanding various social issues.

This field is defined by two key characteristics. First, it involves working with databases to retrieve and analyze the necessary information. Second, it integrates textual content with infographics – such as charts, tables, and visually compelling graphics – which not only enhances the credibility of the information but also improves clarity, making complex data more accessible and comprehensible to the audience. In the context of Kazakhstani media, there is a prevailing reluctance to present raw data without proper contextualization. Rather than offering data in its raw form, media outlets focus on presenting information in a manner that is both engaging and visually appealing, aiming to captivate the audience while ensuring the content remains informative and meaningful. Thus, data journalism in Kazakhstan is becoming increasingly relevant due to its capacity to uncover intricate details regarding various issues, verify or refute official statements from government officials, politicians, and business leaders, and analyze corporate reports, financial data, and social indicators (such as healthcare, education, and housing quality).

5. Conclusion

This study has identified several challenges and obstacles hindering the development of data journalism in Kazakhstan. However, despite these barriers, the field is steadily attracting a growing number of journalists. This trend is evidenced by the increasing availability of specialized courses and training programs offered by prominent universities and journalistic organizations across the country. As a result, data journalism is poised for significant growth in the near future, driven by a new generation of journalists equipped with the skills and knowledge necessary to navigate this evolving landscape.

This study contributes to the growing field of data journalism by examining methodological approaches to working with open data in the contemporary media environment. Focusing on the development of data-driven journalism practices, the study highlights how open data sources, analytical tools, and visualization techniques are increasingly influencing the creation of scientifically sound media content. Integrating approaches from data analysis, verification methods, and visual storytelling, the findings demonstrate that the effective use of open data requires not only technical expertise but also a high level of editorial skill and methodological rigor.

The study demonstrates that open data journalism functions as both an analytical process and a communicative practice. While access to publicly available datasets expands opportunities for investigative reporting and transparency, the quality of journalism depends largely on journalists' ability to critically interpret data, apply fact-checking procedures, and translate complex datasets into accessible visual narratives. In the context of Kazakhstan's evolving digital media landscape, these practices are increasingly important in strengthening public accountability, increasing information transparency, and supporting evidence-based public discourse. Overall, the study expands on existing scholarly work on data journalism by emphasizing methodological approaches that integrate data collection, verification, interpretation, and visualization. The results highlight the importance of developing interdisciplinary competencies that link journalism, data science, and digital storytelling. In the context of the rapid digitalization of media systems, enhancing journalists' open data literacy and integrating data-driven methodologies into editorial practices will be crucial for the future development of investigative and analytical journalism. As the media landscape continues to evolve, journalists must be equipped with the skills to effectively utilize open data, enabling them to conduct more comprehensive and informed investigations. The incorporation of data-driven approaches into editorial processes will not only enhance the depth and accuracy of reporting but also ensure that journalism remains responsive to the growing demand for transparency and accountability in the digital age.

Acknowledgments

Not applicable.

Author contributions

PhD student Olzhas Abdikhalyk was responsible for conceptualisation, study design, methodology development, data collection, and original draft preparation. Associate Professor Aliya Beldibekova contributed to methodology development, data analysis, and validation. Candidate of Philological Sciences Manchuk Mukasheva participated in data collection, empirical analysis, and interpretation of the findings. Associate Professor Hulya Celik was responsible for manuscript review and editing, refinement of the literature, and ensuring academic quality and consistency.

All authors contributed to the interpretation of the results, critically reviewed the manuscript, and approved the final

version for publication.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Redfame Publishing.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Copyrights

Copyright for this article is retained by the authors, with first publication rights granted to the journal.

References

- Anderson B., & Borges-Rey, E. (2019). Encoding the UX: User Interface as a Site of Encounter Between Data Journalists and Their Constructed Audiences. *Digital Journalism*, 7(9), 1253–1269. <https://doi.org/10.1080/21670811.2019.1607520>
- Anderson, C. W. (2013). Towards a Sociology of Computational and Algorithmic Journalism. *New Media and Society*, 15, 1005–1021. <https://doi.org/10.1177/1461444812465137>
- Appelgren E., & Nygren G. (2014). Data Journalism in Sweden: Introducing New Methods and Genres of Journalism into 'Old' Organizations. *Digital Journalism*, 2, 394–405. <https://doi.org/10.1080/21670811.2014.884344>
- Bell, E. (2012). Journalism by Numbers. *Columbia Journalism Review*. [Online]. http://www.cjr.org/cover_story/journalism_by_numbers.php?page=all
- Boyles, J., & Meyer, E. (2016). Letting the Data Speak. *Digital Journalism*, 4(7), 944–954. <https://doi.org/10.1080/21670811.2016.1166063>
- Brewer, D. (2018). *Fact-checking Tips for Journalists*. <https://ijnet.org/en/story/fact-checking-tips-journalists>
- Coddington, M. (2015). Clarifying Journalism's Quantitative Turn. *Digital Journalism*, 3(3), 331–348. <https://doi.org/10.1080/21670811.2014.976400>
- Erkimbay, A. (2017). Practical Journalism in Kazakhstan. MediaNet International Journalism Center. *Data Journalism*. 89–94. ISBN: 978-601-06-4251-5.
- Erkimbay, A. (2019). Business Journalism: A Practical Guide. *Minber Journalists Support Center, Almaty*, 118–122.
- Fink, K., & Anderson, C. W. (2014). Data Journalism in the United States: Beyond the "Usual Suspects". *Journalism Studies*. <https://doi.org/10.1080/1461670X.2014.939852>
- Gorokhovskiy, A. (2017). *Fact-checking as a Trend in Investigative Journalism: Opportunities and Future Prospects*. Almaty. ISBN: 978-601-7222-09-3, p. 11.
- Gray, J., Bounegru, L., & Chambers, L., eds. (2012). *Data Journalism in Perspective*. The Data Journalism Handbook.

Available at: <http://datajournalismhandbook.org/1.0/en/index.html>

- Gynnild, A. (2014). Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets. *Journalism*, 15, 713–730. <https://doi.org/10.1177/1464884913486393>
- Holovaty, A. (2009). *The Definitive, Two-Part Answer to 'Is Data Journalism?'*. <http://www.holovaty.com/writing/data-is-journalism/>
- Howard, A. (2014). *The Art and Science of Data-driven Journalism*. New York: Tow Center for Digital Journalism. <http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf>
- Lewis, S. C., & Usher, N. (2013). Open Source and Journalism: Toward New Frameworks for Imagining News Innovation. *Media, Culture and Society*, 35, 602–619. <https://doi.org/10.1177/0163443713485494>
- Popova, M. (2009). *Data Visualization: Stories for the Information Age*. BusinessWeek, August 12. www.businessweek.com
- Rafizadeh, M. (2014). *The Challenge and Dilemma: Fact-checking in Journalism and Research*. HuffPost. https://www.huffpost.com/entry/the-challenge-and-dillema_b_4806347
- Stray, J. (2010). How the Guardian Pioneered Data Journalism with Free Tools? *Nieman Journalism Lab*. <http://www.niemanlab.org/2010/08/how-the-guardianis-pioneering-data-journalism-with-free-tools/>
- Stray, J. (2011). *A Computational Journalism Reading List*. [Online]. <http://jonathanstray.com/a-computational-journalism-reading-list>.
- Sylvain, P., & Dagiral, E. (2013). Data-driven Journalism and the Public Good: Computer-Assisted Reporters' and 'Programmer-Journalists' in Chicago. *New Media and Society*, 15, 853–871. <https://doi.org/10.1177/1461444812463345>
- Terry, F., Spurgeon, C., Daniel, A., & Swift, A. (2012). The Promise of Computational Journalism. *Journalism Practice*, 6, 157–171. <https://doi.org/10.1080/17512786.2011.616655>
- Yakushina, E. (2013). *Media. Information. Communication*, 6, 6–8.
- Yesimkhanov, M. (2021). *Is Kazakhstan Ranked Third in the World for Abortions?* FactCheck.kz, April 11. <https://factcheck.kz/kaz/basty-bet1/kazakstan-abort-zhasaudan-alemde-3-orynda-ma/>