

Effective Governance for Economic Development Pillar 2: Civil Society & Accountability



Report Mapping Open Data in Tajikistan

February – March, 2022

Acknowledgments

This report Commissioned by Internews and prepared by Zerkalo-Analytics under the EGED Pillar 2: Civil Society for Accountability project. This project is supported by funding from the UK Government and implemented in cooperation with ACTED.

Many individuals have contributed to this report. The inception and research reports were authored by Ion Marandici (PhD). The database was put together by a team consisting of Kahramon Bakozoda (Director), Berik Kozhomkulov (Coordinator), Ion Marandici, Cholpon Beishekeeva, Sayfullo Sadykov, Rukhshona Dzhalolova, Iskandar Razhabov, Farrukh Sadykov and other colleagues from Zerkalo. The author of the report wishes to thank the following reviewers from Internews for taking the time to provide helpful comments on the previous drafts of this study: Saida Manieva, Isabelle Schläpfer, Aizada Toktogulova, Asel Omuralieva, Rachel Caywood, and Zarina Ergasheva.

This material has been funded by the UK aid from the UK government; however, the views expressed do not necessarily reflect the UK government's official policies.

Table of Contents

1. Executive Summary	4
2. Introduction	7
3. A Brief Overview of the State of Data Journalism in Tajikistan	10
4. Mapping Open Data in Tajikistan	13
5. Challenges for Data Journalists in Tajikistan	20
6. Recommendations	22
Annex 1: Instructions for Journalists on How to Use the Database	24
Annex 2: Checklist on How to Work with Open Data	26
Annex 3: Data Request Letter (Original)	28
Annex 4: Top-3 Visualization Tools	29
Annex 5: Recommendations on PDF Conversion	30

Executive Summary

Context: New data sources related to socio-economic development should improve journalistic reporting. However, in the absence of a centralized open data repository in Tajikistan, journalists may face delays in getting quick access to important datasets.

Goal: To address this challenge, Zerkalo conducted a mapping study, identifying the open data sources in Tajikistan and putting together a comprehensive database accessible on [Airtable](#) ([link](#)) and [Google Spreadsheets](#) ([link](#)). By mapping most of the open data, the current study facilitates the access to development-related information for journalists working on socio-economic issues.

Method: Relying on a two-stage desk research approach, Zerkalo constructed a database containing most of the open data sources in Tajikistan. The data collection phase for the mapping study was implemented from February 14 to March 10, 2022. Overall, the research team examined 1032 hyperlinks to 68 websites, holding relevant datasets. As a result, Zerkalo identified 294 datasets on socio-economic themes and included them in the database.

Results:

- The mapping revealed that most datasets are held by the following organizations: the National Bank of the Republic of Tajikistan (48), the Statistical Agency under the President of Tajikistan (46), World Bank (32), Dushanbe City Administration (21), the University of Central Asia (15), the International Monetary Fund (14), Tax Agency (12) and the Microfinancing Association of Tajikistan (11).
- Another finding concerns the format of the datasets: PDF (107), webpages (86), Excel (64) etc. This result suggests that many of the datasets are embedded in PDF files, which are not as suitable as other files for data analysis.
- The mapping shows that there are more datasets on particular topics such as Banks and Finances (52), Income (19), Trade (18), Taxes (18) and Regional Development (13).
- In terms of publisher, datasets originate from state agencies (171), international organizations (52), local NGOs (29) and academic institutions (16).
- As regards language, the datasets are accessible in Russian (159), English (64) and Tajik (45).
- Instructions on how to navigate the database on [Airtable](#) (click [here](#)) are provided in Annex 1.

Conclusion: Taken together, our report finds that there is potential for the emergence of a strong data journalism community in Tajikistan, proposing several areas in which improvements could be made.

Recommendations

To Media Organizations:

- Journalists should take advantage of the opportunity to explore the open data landscape in Tajikistan by accessing the database ([link](#)).
- To encourage the development of data journalism, media organizations should set aside sufficient resources for reporters engaging in data analysis.
- The study suggests that journalists could benefit from workshops on data visualization.
- Media organizations should eliminate technical obstacles preventing data journalists from embedding data visualizations in their stories (Annex 4).
- To ensure data reliability, journalists working with data could follow the guidelines in Annex 2.
- Journalists should ask for updated datasets from authorities. To that end, the study includes a template data request letter (Annex 3).

To Government:

- As most open data are linked to certain themes, one recommendation is to prompt officials from other agencies to release data.
- Government should invest more in its capacity to consistently collect and publish data.
- It should increase awareness on the role and importance of open data for socio-economic development providing data in an easily and safely accessible machine-readable format.
- Government officials tasked with data distribution in each agency/ministry should respond in a timely manner to journalistic requests for data updates.

To Donor Organizations:

- Civil society organizations and international donors supporting the development of a data community in Tajikistan need to engage the governmental agencies to persuade them of the importance and usefulness of making data in an appropriate format.
- At the same time, employees of state institutions could be provided with hands-on training on how to publish data in a machine-readable format rather than as PDF documents. This would make the job of data journalists easier.

- To address the problem of weak journalism capacity, targeted data literacy courses could be offered for local media organizations.
- A pilot open data project for the city of Dushanbe should be implemented.

Introduction

Citizens around the world receive their information from the mass media (radio, TV, newspapers), online content creators and social media. Journalists fulfill a key function in modern societies as they inform citizens of new policy initiatives and often monitor the way such policies are implemented and their effects on people's lives, connecting policymakers to the public and holding them accountable. This mechanism works relatively well in advanced democracies with strong and pluralist media structures. Moreover, existing studies confirm that strong media structures have a positive impact on economic phenomena such as poverty, wealth inequality, and corruption.¹ By contrast, developing countries with fledgling media markets are more prone to suffer from a lack of reliable and unbiased information about governmental activities.

Data journalists play a special role in such information ecosystems. In a world in which data has become increasingly abundant, journalists working with large datasets can often better fulfill their mission of holding officials accountable and informing the public of policy implementation by delivering data-driven stories about the latest governance-related policies in a particular context. Of course, when improperly trained or not observing professional standards and ethics, journalists and the media can contribute to biased information environments, enhancing disinformation and public opinion manipulation. In this sense, data often may be used to sway the public in favor of the policy preferred by the authorities. By contrast, when used correctly data can provide important evidence and combat disinformation and manipulation attempts.

Unlike their peers reporting on socio-economic issues in a traditional manner, journalists working with data are often overlooked by development organizations supporting civil society consolidation. Such neglect is probably linked to the relative recency of data journalism as a distinct specialization. The underlying assumption of this study is that strengthening the work of data journalists will contribute to their efforts to hold officials accountable and empower the public through evidence-based reporting.

For data journalists to emerge as a community, collaborative work and accessible data are crucial. Accessible and reliable open data enable journalists to offer high-quality reports on socio-economic issues. For the purpose of this study, we define open data as “data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike.”² Specifically, this study focuses on open data originating from public institutions as they have most data. Another aspect concerns the machine-readability of open data. A 2017 World Bank study mentioned that “for open data to have the greatest value, machine-readability must become a key goal for data providers.”³ In the context of this study, machine-

¹ Sanghamitra Bandyopadhyay, *Knowledge-based Economic Development: Mass Media and the Weightless Economy*. DARP (74), Suntory and Toyota International Centres for Economics and Related Disciplines, UK: London, 2005..

² See *The Open Data Handbook*, Open Knowledge Foundation, <https://opendatahandbook.org/guide/en/what-is-open-data/>

³ Audrey Arriss, Machine-readable Open Data: How it's Applicable to Developing Countries, Data Blog post, March 2, 2017. <https://blogs.worldbank.org/opendata/machine-readable-open-data-how-it-s-applicable-developing-countries>

readable data refer to structured data that can be exported as tabular data (Excel or CSV files) for further processing by computers to calculate various statistics or change the structure of a dataset. While in developed countries there is a wealth of data sources that journalists can retrieve to write their stories, the situation is rather dire in developing countries.

Four problems stand out:

- First, data are scarce as developing states usually lack the capacity to produce high quality data. By high quality data, we refer to data that are disaggregated, time series, micro-level and granular data with geolocation information. A time series is a dataset including data points collected over longer periods of time. Examples include GDP figures from the last fifty years or the World Value Surveys. Micro-level refers here to unit-level data such as census information (i.e. individuals, households), while granularity pertains to the way the data are classified.
- Second, such dearth of data means that journalists lack opportunities to consult, practice with and get better at incorporating data in order to tell a compelling story.
- Third, journalists in developing countries rarely have the skills and know-how to use the latest data tools as it may be time-consuming to acquire such technical competences. After all, learning how to do proper data analysis requires training, which can be costly.
- Fourth, data journalism itself is a laborious process. Often media organizations cannot afford data stories as editors lack the resources to adequately reward such work. For example, the usual pay in editorial offices is on average 150-200 somoni (\$11-\$15) per article. Accordingly, even trained data journalists may not be motivated to work on a data story and implement the acquired knowledge.

In other contexts, data-driven reporting, which comes close to investigative journalism, may pose personal risks to journalists and their families, discouraging them from following this path. A worldwide survey of data journalists revealed that access to quality data (56%), time pressure (49%), lack of adequate knowledge in data analysis (44%) and ensuring data reliability (39%) were among the top difficulties encountered by data journalists.⁴

Successful data journalism produces public interest stories. In this sense, the work of data journalists is situated at the intersection of several distinct areas: working to find data and construct databases, develop stories centering on data, use advanced data visualization tools, and apply new innovative technologies to

⁴ See Andrea Abellan, The State of Data Journalism Survey, 2021, <https://datajournalism.com/read/blog/data-journalism-survey-2021>

process data.⁵ In this sense, the “Mapping of Open Data” study contributes to the first area by putting together a rich database, which facilitates access to reliable open data from disparate sources.

The present document provides details about the results of the study. The structure of the final report is as follows. Section 2 will discuss the state of data journalism in Tajikistan, highlighting the existing legal and institutional framework. Section 3 describes the research methodology and key results. Section 4 points out potential challenges for data journalists, suggesting possible solutions. Building on our results, the last section will include key recommendations.

⁵ Ángel Arrese (2022) “In the Beginning Were the Data”: Economic Journalism as/and Data Journalism, *Journalism Studies*, DOI: [10.1080/1461670X.2022.2032803](https://doi.org/10.1080/1461670X.2022.2032803)

A Brief Overview of the State of Data Journalism in Tajikistan

Tajikistan has a growing media community largely concentrated in its capital city – Dushanbe. According to a 2019 study of media preferences, 376 registered newspapers (112 - state-owned and 264 - private), 245 journals (114 - state-owned and 131 - private), 10 news agencies (1 - state-owned and 9 - private), 34 TV stations (8 - national and 26 - regional), and 33 radio stations (4 with national coverage) were active in the country.⁶ Of all registered newspapers, the government-owned newspapers are the most distributed. Dzhumhuriyat, for instance, the only daily newspaper in Tajikistan, has a circulation of 240 000. Omuzgor (the publication of the Ministry of Education), Minbari Halk (a party newspaper), and Sadoj Mardum (the parliamentary weekly) are widely circulated too. The most popular private newspapers are three tabloids publishing entertainment news followed by the publications Tajikistan and Asia-Plus, which offer more analytical reporting. Much as in the rest of the world, circulation figures confirm that the number of copies sold is declining each year.⁷ On top of that there is a growing community of Internet content creators and online portals in Tajikistan. Still, despite this apparent abundance of information sources, outlets do not have dedicated departments that use data to create content, and when data is used, it is usually used in a cursory manner without understanding of how it can strengthen a story. To support data journalists in Tajikistan, Internews has conducted several data training and mentorship sessions, launching a data journalism program in 2021.

The underdevelopment of data journalism as a separate field in Tajikistan does not mean that journalists do not use data in their reporting on various themes. Several examples selected after the review of several Tajik media outlets will illustrate this point. One major online resource is the Asia Plus portal, which publishes a variety of content, including economic and business news.⁸ Asia Plus has a special section dedicated to the economy, featuring longer stories. For instance, an investigation into the causes of the significant debt incurred by Barqi Tojik, Tajikistan’s national power utility company, relies on a wealth of comparative data about energy consumption to explain the recurring power supply problems.⁹ Some reports utilize data to tell fascinating stories about the effect of global warming on the melting of the glaciers in Tajikistan, diverging trade statistics in Central Asia, delayed wage payments, crediting, remittances, and budgetary spending. In another example, a report on the Radio Ozodi website discusses economic growth without offering any visual representations.¹⁰ In this case, apparently, the website of the media organization

⁶ Media preferences of the population of Tajikistan: TV, radio, print media, sites, social networks and messengers, Media-Consulting and Zerkalo, 2019. http://zerkalo.tj/uploads/files/2020/MediaPreff-2019_Final_RUS.pdf

⁷ Razhab Mirzo, How are the Tajik newspapers sold?, Newreporter.org, August 26, 2019. <https://newreporter.org/2019/08/26/kak-prodayutsya-tadzhikskie-gazety/>

⁸ For instance, one media project consisted of interviews with successful entrepreneurs: <https://asiaplustj.info/projects/Probizness> .

⁹ Pairav Chorshanбиеv, Three Reasons behind Barqi Tojik’s “Poverty”, Asia-Plus, March 1, 2022.

<https://asiaplustj.info/ru/news/tajikistan/economic/20220301/tri-prichini-bednosti-barki-tochik>

¹⁰ Abdullo Ashurov, “7.3% рушди иктисоди Тоҷикистон. Аммо чаро муҳоҷирати корӣ идома дорад?,” Radio Ozodi, April 13, 2019. <https://www.ozodi.org/a/29876593.html>

does not allow the integration of certain visualization frames. Despite the fact that some journalists received data-related training, the brief overview of media content reveals that most journalists use data in a conventional way without relying on the latest data tools.

In other cases, journalists use infographics and data visuals prepared by international economic organizations rather than developing their own. For instance, an extended report on poverty in Tajikistan published by Your.tj, a recently established news portal, embedded an infographic generated by the World Bank.¹¹ This practice suggests that not all editorial offices can afford hiring an individual who would create infographics. While it is possible that some key publications may have been overlooked here, these examples indicate that data journalism is still a developing field in Tajikistan. Hence, there is much room for improvement in terms of data availability and literacy.

It is important to mention that there have been several projects aiming to introduce open data policies and data journalism. In 2015, an open data readiness report was published with support from the Open Society Institute, Open Knowledge Foundation, and the World Bank. Among its key recommendations were the formation of an Open Data Expert Group that would design a National Open Data Strategy as well as strong collaboration between government agencies, civil society, and developer community in conjunction with the removal of the practice of denying access to data with reference to the law on state secrets. The paper also recommended the design of pilot projects and a program for capacity building for government officials.¹² Likewise, the Khoma project supported by FCDO and the Open Society Institute from 2016 to 2020 educated journalists on financial literacy, including a data journalism component. A key result of this project was a specialized website dedicated to economic and data journalism.¹³ Cabar.asia conducted workshops in data journalism within the framework of the School of Young Analysts with some online training tools for data journalists available to everyone.¹⁴

In terms of data access in Tajikistan, it is important to point out that roughly half of all the state institutions have a web contact form for the public to inquire about unpublished data. This is a positive finding as it signals a certain degree of openness toward information requests. However, state institutions vary in terms of their response time. A study by the National Association of Independent Media found out that the websites of government agencies do not always contain the entire amount of data, often they are not in the proper format, are not updated on time, and there is often a risk that data can be deleted. The same study suggests that many government agencies respond reluctantly to inquiries, months late, answering only a few questions from a list. Moreover, request letters are sometimes completely lost or left without an answer for

¹¹ “Tajikistan’s Economy Grows, but Poverty Persists,” Your.tj, <https://your.tj/jekonomika-tadzhikistana-rastet-no-naselenie-vse-ravno-bedneet/>

¹² See https://web.archive.org/web/20161017073358/http://ispa.tj/wp-content/uploads/2015/09/En_ODRATajikistan_20151.pdf

¹³ See <https://www.khoma.tj/datajournalism/>

¹⁴ See <https://school.cabar.asia/en/kursy-2/>

months.¹⁵

Our key assumption of this study is that by locating and facilitating the access to data from multiple venues, the current data mapping study contributes to effective governance. It does so by consolidating access to data sources to assist local media organizations with low data analysis capacity. Journalists will write better stories once they can easily find and analyze the required data. They will be aware of how to use data in their reporting as they familiarize themselves with the instructions accompanying the database (see below). In doing so, they will help citizens monitor the effectiveness of public policies in Tajikistan. At the same time, by supporting journalists' access to economic development data, the data mapping initiative may also encourage civil society actors to demand more transparency with regards to data production and distribution.

¹⁵ See The National Association of Independent Media, Report on Access to Information Practices in Tajikistan, June 20, 2019. <http://a2i.nansmit.tj/2019/06/20/> For the exact response time see: <http://a2i.nansmit.tj/%d0%bf%d0%be-%d0%b0%d0%b4%d1%80%d0%b5%d1%81%d0%b0%d1%82%d0%b0%d0%bc/>

Mapping Open Data in Tajikistan

The “Open Data Mapping in Tajikistan” team put together a database containing available open data sources on socio-economic themes. Carried out over the February-March 2022 period by an international and multicultural team, comprising specialists from Kyrgyzstan, Tajikistan, and Canada, the study focused on open data in the area of economic development released by the state.\

Private sector actors traditionally feature as beneficiaries of open data initiatives as they may use such data to identify new business opportunities.¹⁶ Often companies act as intermediaries. They can, however, generate and own data as well. For example, Twitter provides paid access to its archive of data, while Uber and Airbnb released some of their data as open data. Most private companies hesitate to make their data available as open data for business-related reasons. Regrettably, such private sector data have not been identified in Tajikistan with some companies selling their data instead. This mapping prioritized open data that can be utilized by journalists without any additional costs to them. Since such closed datasets owned by private enterprises cannot be referenced by journalists, they remain unsuitable as a viable source of information for journalists.

To construct the database, Zerkalo adopted a two-stage desk research approach. First, we conducted desk research identifying all the sources containing relevant socio-economic data about Tajikistan. We began with an evaluation of all data published online by state agencies. Z-Analytics operators consulted 1032 webpages on 91 websites of governmental agencies, international organizations, local and regional NGOs. The mapping identified 294 datasets from 54 sources as of March 18, 2022. Once a dataset was located, the operators followed a standard protocol, indexing and recording its characteristics. The last batch of datasets was added by using Google’s dataset search function and within selected data repositories.¹⁷

In cases where governmental data were outdated, the operators proceeded to the second stage. In agreement with Internews, requests were made for data from the 2019-2022 period in order to update the datasets already collected. In most cases, such data were requested from institutions that were less active in terms of data publishing. Requests were not sent to those state institutions, which did not publish any data. The template for such data requests is included in Annex 3. In some instances, the research assistants contacted officials by phone or visited the data-holding divisions within state institutions. At this stage, they recorded whether they received a response and whether authorities complied with the existing access to information legislation.

The main outcome of the project is a “flat-file” database consisting of simple rows and columns as

¹⁶ See for instance Open Data 150, a project implemented in partnership with the Open Data 500 Global Network, listing Canadian companies relying on open data to develop their enterprises: <https://canada.opendata500.com/list-page.html>

¹⁷ A list of open data repositories can be found here: <https://mira.mcmaster.ca/research/open-access-data-repositories> Other potential sources include <http://dataverse.harvard.edu> , data.europe.eu , <https://cyberleninka.ru> and similar repositories.

seen in Figure 1 ([link](#)). The structure of the database is identical to a spreadsheet but adds a layer of complexity by including a series of drop-down lists. Available in English and Russian, it contains information about datasets, structured along 15 basic categories, each corresponding to a column. To make it more user-friendly, the new database has been placed on the data-organization platform Airtable ([link](#)). The dataset of datasets can be easily searched, sorted, and updated. It is free to download, disseminate and use.

Figure 1. Excel format of the database with column names.

Dataset Title	Access Date	Data period	source (Full nam	Publisher	Type	Publication Area	Link	Publication Level	Format	Access	Language	Issuance Cycle	Next Release	ation Main Theatrum	Main Theatrum Theset	Main Theatrum Theset (email or pl)	Comments
Statistics on edu. 02/14/2022.		2017-2018	Ministry of Educ	State	Other		https://maorfi.t	National	PDF.	Free	Tajik	Annually	Not known	Education	Institutions	Students	(992 37) 221-46+ P. 5-9, 12-17, 35
Indicators by nu 02/17/2022.		The first 3 mont	Ministry of Cult	State	Budget		https://vfarhanj	National	PDF.	Free	Tajik	Not known	Not known	Culture	Achievements	Offers	734025, Sharr D P. 16, 29, 39-40
Socio-economic 02/18/2022.		2014-2020; Indu	Statistics Agenc	State	Social protectio		https://stat.wvv	National	PDF.	Free	Tajik	Quarterly	Not known	Food security; P	Prices and tariff	Industrial scena	Tajikistan, Dush. Collection "Foc
Environmental 02/18/2022.		1991-2018	Statistics Agenc	State	Agriculture		https://stat.wvv	National	PDF.	Free	Tajik	Not known	Not known	Environmental (R	Rational use of investments, pr	Tajikistan, Dush. Environmental	
Main indicators 02/18/2022.		2014-2018	Statistics Agenc	State	Other		https://stat.wvv	National	PDF.	Free	Tajik	Not known	Not known	Development o	Indicators and Ir	Health financing	Tajikistan, Dush. Collection "Hea
List of laws on ir 21.02.2022.		2005-2012	World Trade org	international	Or Legislation		http://vto.tj/er	National	Web Page	Free	English	Not known	Not known	World Trade Org	Laws. Trade	Agreements	Ministry of Econ Separate websi
Indicators for in 21.02.2022.		2011-2020	World Intellect; international	Or Ownership			https://www.wi	National	Web Page	Free	Russian	Not known	Not known	Economic indica	Technology	Patents	Contact center + Sale of statistic
Monthly inflati 21.02.2022.		September 2016	National Bank o	State	Prices		https://nbt.tj/ri	National	PDF.	Free	Russian	Monthly	Not known	Banks and Finan	Inflation	Consumer price	Tel.: + (992) 44 6 Link to the page
Inflation Review 21.02.2022.		2016-2021.	National Bank o	State	Prices		https://nbt.tj/ri	National	PDF.	Free	Russian	Monthly	Not known	Banks and Finan	Inflation	Consumer price	Tel.: + (992) 44 6 #VALUE!
Price indices 21.02.2022.		1999-2019	National Bank o	State	Prices		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Banks and Finan	Inflation	Dynamics	Tel.: + (992) 44 6 #VALUE!
Monetary policy 21.02.2022.		2022.	National Bank o	State	Monetary Policy		https://nbt.tj/uj	National	PDF.	Free	Russian	Annually	Perhaps 2022-20	Banks and Finan	Forecasts	Inflation	Tel.: + (992) 44 6 #VALUE!
Credit Investme 21.02.2022.		2000-2021.	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	2022.	Real sector	Loans	Statistics	Tel.: + (992) 44 6 #VALUE!
Residues of savi 21.02.2022.		2002-2021	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	2022.	Real sector	Deposits	Statistics	Tel.: + (992) 44 6 #VALUE!
National Bank v 21.02.2022.		2001-2007	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Monetary secto	Assets	Reserve money	Tel.: + (992) 44 6 #VALUE!
National Bank v 21.02.2022.		2008-2021.	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Monetary secto	Assets	Reserve money	Tel.: + (992) 44 6 #VALUE!
Monetary Overv 22.22.2022.		2001-2007	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Monetary secto	Assets	Loans. Deposits	Tel.: + (992) 44 6 #VALUE!
Monetary Overv 22.22.2022.		2008-2020	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Monetary secto	Assets	Loans. Deposits	Tel.: + (992) 44 6 #VALUE!
Monetary Overv 22.22.2022.		2001-2007	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Assets. Liabilitie	Loans. Deposits	Holders. Cash. S	Tel.: + (992) 44 6 #VALUE!
Monetary Overv 22.22.2022.		2008-2020	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Assets. Liabilitie	Loans. Deposits	Holders. Cash. S	Tel.: + (992) 44 6 #VALUE!
Monetary aggreg 22.22.2022.		2001-2007	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Tajik	Annually	Not known	Wide money rat	Money	Deposits. Secur	Tel.: + (992) 44 6 #VALUE!
Monetary aggreg 22.22.2022.		2008-2020	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Wide money rat	Money	Deposits. Secur	Tel.: + (992) 44 6 #VALUE!
Weighted avera 22.22.2022.		2002-2021	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Interest	Betting	Loan rates	Tel.: + (992) 44 6 #VALUE!
Weighted avera 22.22.2022.		2002-2021	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Interest	Betting	Deposit rates	Tel.: + (992) 44 6 #VALUE!
Information abc 22.22.2022.		1995-2021	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Interest	Betting	Medium bankin	Tel.: + (992) 44 6 #VALUE!
Refinancing rate 22.22.2022.		02/05/2021 to th	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Refinancing rate %	per annum	Validity	Tel.: + (992) 44 6 #VALUE!
Results of aucti 22.22.2022.		2005-2020	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Excel	Free	Russian	Annually	Not known	Securities of the	State treasury b	Trade. Yield	Tel.: + (992) 44 6 #VALUE!
Calendar of the 02.22.2022.		February 2022.	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Web Page	Free	Russian	Monthly	March 2022.	Monetary secto	Securities	Calendar releas	Tel.: + (992) 44 6 #VALUE!
Calendar of the 02.22.2022.		2022.	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Web Page	Free	Russian	Annually	2023.	Monetary secto	Securities	Calendar releas	Tel.: + (992) 44 6 #VALUE!
Rate 02.22.2022.		02.22.2022 (tod	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Web Page	Free	Russian	Daily	Not known	Foreign Econom	Rate	#VALUE!	Tel.: + (992) 44 6 #VALUE!
Information on 02.22.2022.		On 08.28.2015	National Bank o	State	Banks and Finan		https://nbt.tj/ri	National	Web Page	Free	Russian	Not known	Not known	Foreign Econom	Currency quotat	#VALUE!	Tel.: + (992) 44 6 #VALUE!

Journalists can access the database in two ways. They can access it as a Google spreadsheet ([link](#)). As seen in Figure 1, the spreadsheet includes two sections: open and on-demand data. The open data section lists all the accessible data, whereas the on-demand part contains data which can be obtained by contacting the data holders. Both the open data and on-demand data sections include specific contact information in case database users need more details or updated versions of the dataset. To request on-demand data, journalists can use the template included in Annex 3.

The second option includes the user-friendly version of the dataset on Airtable, a practical data-organizing tool allowing for the easy retrieval of data. This is the recommended way of accessing the database since it provides a convenient visualization of the data. In doing so, journalists can group the data along any column of choice, which enables them to easily search and access it. Instructions on how to access the Airtable data can be found in Annex 1.

Below are the major findings based on the analysis of the collected data. Figure 2 shows that most of the datasets are at the national level with some regional and international. The database contains only a few local datasets.

Figure 2. Dataset Level.



Source: Zerkalo Center • Created with Datawrapper

In terms of publisher type, we observe that most of the data are from governmental agencies and international organizations with a few of them originating from local NGOs and academic institutions (Figure 3).

Figure 3. Datasets by Publisher Type.

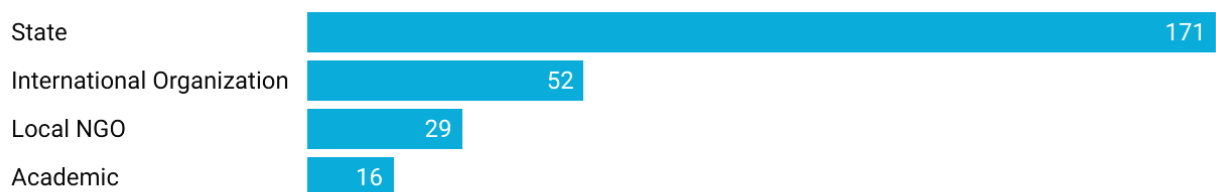


Chart: Ion Marandici • Source: Zerkalo Center • Created with Datawrapper

Figure 4 presents data on the format of the data. It reveals that in Tajikistan open data are usually published as PDF documents and integrated into web pages. This, of course, has implications with regards to how journalists can access data. The fact that many of the datasets are embedded in PDF documents makes it difficult for the media to access and extract data for subsequent analysis. In Annex 5, we provide recommendations on how to extract data from PDF documents.

Figure 4. The Format of the Datasets.



Chart: Ion Marandici • Source: Zerkalo Center • Created with Datawrapper

With regards to language, most of the datasets included in the database are accessible in Russian, followed by English and Tajik. Some of the datasets may be available in multiple languages.

Figure 5. Datasets by language.

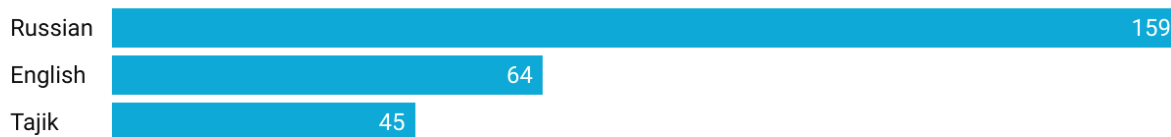


Chart: Ion Marandici • Source: Zerkalo Center • Created with Datawrapper

In terms of coverage, most of the open data are concentrated in the following sectors: banks and finance, trade, tax, revenue and regional development (Figure 6). Datasets on housing, property rights and productivity are relatively rare. One reason for the abundance of open data in the banking and financial sectors may be related to the data-heavy activity of the public institutions in this area, which must inform the public about exchange and interest rates, required reserves, and formulate forecasts based on various macroeconomic indicators. In other words, they may be producing large amounts of data and releasing some of it. Another reason may be linked to certain reporting requirements. The National Bank of Tajikistan (NBT) is bound by certain obligations to its creditors such as the International Monetary Fund, which may demand more transparency as well as the regular publication of data before issuing preferential loans. The NBT also participates in trading on international exchanges and thus needs to comply with regulatory provisions in order to be allowed to participate.

Figure 6. Datasets by Economic Sector.

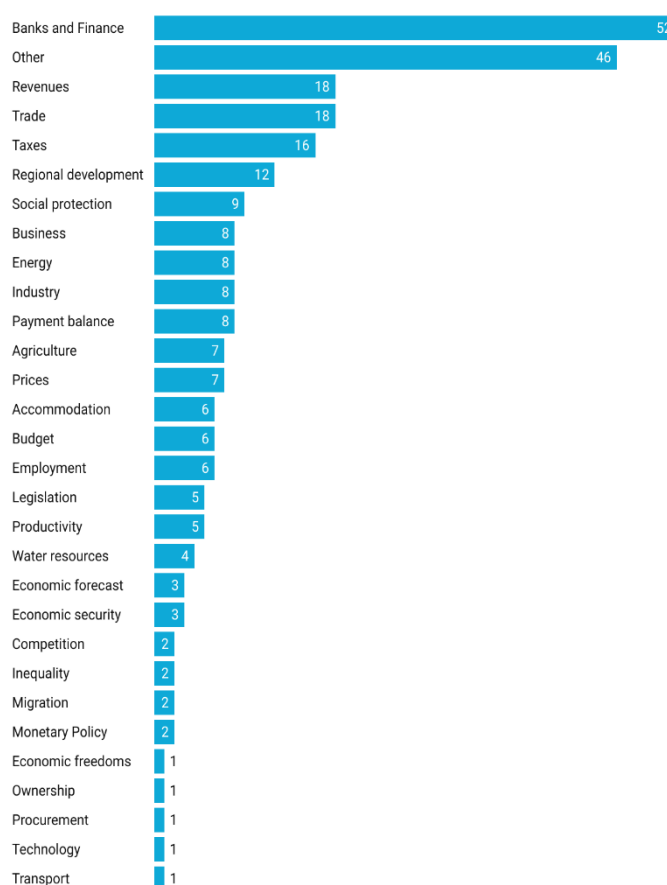


Chart: Ion Marandici • Source: Zerkalo Center • Created with Datawrapper

There is great variation with regards to the amount of open data found on governmental websites. Table 1 lists the number of datasets by domestic sources. It reveals that three institutions - the Statistical Committee of Tajikistan, the National Bank, and the Office of the Mayor in Dushanbe - generate regularly more open data than other institutions. The three aforementioned institutions published more than 20 relevant datasets.

Table 1. Tajik Websites ranked by the Number of Relevant Datasets
More than 20 Datasets
The Statistical Agency subordinated to the President of Tajikistan
The National Bank of Tajikistan
The Mayor’s Office in Dushanbe
From 2 to 5 Datasets
The Agency for Employment and Labour (Ministry of Labour, Migration and Employment)
Analytical center “Navo”
State Committee on Land Management and Geodesy
State Committee on Investments and State Property Management

Ministry of Finance
Ministry of Energy and Water Resources
National Patent Center
Barqi Toqik, the National Energy Company
The Republican Alliance of Consumer Associations – Tajikmatlubot
Customs Agency
From 6 to 20 Datasets
The Association of Microfinancial Organizations
State Enterprise “Dushanbevodokanal”
Ministry of Economic Development and Commerce
Tax Agency (Tajik Government)
The National Association of Small and Medium Enterprises
1 Dataset
General Prosecutor’s Office
Economic and Demographic Institute of the National Academy of Tajikistan
Committee on Women and Family Affairs (Government of Tajikistan)
Ministry of Culture
Ministry of Education and Science
Ministry of Transportation
Tajikistan’s Trade Portal
Center for Strategic Studies (President’s Office)

While the reasons behind NBT’s open data publication have been discussed above, in the case of the Mayor’s Office in Dushanbe, the apparent data richness may be explained by other factors. A close examination of the datasets revealed that the city administration is not publishing high-quality data nor does it regularly release statistical reports. Instead, multiple datasets covering various regional development topics across its districts are often placed on the same webpage. Also, the city is required to publish detailed price lists for the services it provides, which were included in the database too. Still, Dushanbe does not have an open data portal as some major cities, but it could benefit from such an initiative in the future.¹⁸

Public data are often catalogued and sold as brochures and printed volumes. The Statistical Agency of Tajikistan offers such collections of data.¹⁹ To access Tajikistan’s legislation, data journalists may first

¹⁸ See for instance the open data portal of the city of Toronto launched in 2009: <https://open.toronto.ca/>

¹⁹ Journalists interested in purchasing such data can check out the catalogue of available publications here: https://stat.wv.tj/library/tj/2021_catalog.pdf

consult the website of the National Legislative Center, where most legal acts are published with a delay of two-three months.²⁰ Another useful database containing the key legal acts adopted by the Tajik state institutions is Adlia. A project of the Ministry of Justice, USAID and World Bank, Adlia currently provides paid access to all kinds of legislation faster than the National Legislative Center.²¹

Open data are often removed from sites, disappearing from the public domain. At times, this is the result of the technical limitations of various website designs. In other cases, data is deleted from the Internet, because journalistic investigations reveal some inconvenient aspect about a public official. For instance, according to Tajik journalists, information about earlier contracts has been deleted from the open procurement and public tenders portal, which has not been updated for a while.²² Other portals with missing data include nozirot.tj, investcom.tj (Foreign Investment), and moh.tj (Ministry of Public Health). Often websites are updated and previous data ends up being removed. The general plan of Dushanbe has been deleted from the dushanbe.tj website belonging to the City Executive, while data about owners of private enterprises and other entities disappeared from the fiscal registry website at andoz.tj.

To address this issue, Zerkalo has archived all the datasets. Some data were saved on Google Drive, while in other cases copies were stored using the digital archive of the world wide web platform (see Archive.org). Preserving these datasets enables data journalists to access copies of the datasets on their devices as well as various public data archived online in cases where URLs are broken, or data disappear from the web.

To update some of the datasets, on March 18, Zerkalo requested additional data from multiple state institutions: the Ministry of Education, Agency for Patents, Ministry of Energy, Committee for Investments, Committee on Women Affairs, Fiscal Committee, Dushanbe's Executive Committee, Statistical Agency (see [here](#) for a complete list). Out of the ten institutions contacted, nine responded. Two datasets in pdf format were provided by the Agency for Land Management, Geodesy and Cartography. The Committee on Women Affairs redirected the request to the National Statistical Agency, while the other state organs did not provide the requested data, mentioning instead that all the available datasets are published on their official websites. The representative of the State Committee on Investments promised to provide the data if it is not already published online but did not follow through on his promise. These results suggest that obtaining the latest data from state institutions in Tajikistan is a hassle for data journalists as officials express reluctance in taking such steps to avoid potential career consequences.

²⁰ Website of the National Legislative Center at <http://ncz.tj/legislation>

²¹ The tariff plans vary from 100 to 400 USD. See <http://www.adlia.tj/>

²² See <https://zakupki.gov.tj/ref/zakupki/>. A new version of the website lists the last procurement contracts: <https://eprocurement.gov.tj/tj/searchanno>.

The database can be updated by regular users, who may choose to add new datasets. We encourage journalists and media organizations to add new entries via the following form ([link](#)). This way the open access database may become a self-sustaining project in the sense that it would be maintained by the community of data journalists.

Main Challenges for Data Journalists

Data journalism as a distinct form of journalism is developing slowly in Tajikistan for several reasons. There are multiple causes behind this phenomenon.

First, some journalists find it difficult to use data analysis in their stories. This trend is then further reinforced by the hassles encountered in the process of obtaining data from state institutions, limited resources and the relative lack of popularity of such data-stories among readers. In addition, the way journalism as a discipline is taught in universities matters as well.

A second reason may be related to the manner in which the media in the country operates. Specifically, media freedom has been deteriorating in Tajikistan. This tendency certainly has repercussions for journalists investigating sensitive topics and requesting data from officials. Access to information on certain topics may be limited. According to the 2021 World Press Freedom Index compiled by *Reporters without Borders*, Tajikistan ranked 162 out of 180 countries in terms of media freedom.²³ International media organizations also report on how Tajik journalists and the online content creators are intimidated as a result of their work, with the pandemic further aggravating such trends as authorities introduce new restrictive measures.²⁴

Third, the overall scarcity of open data is another challenge faced by Tajik journalists. To some extent the issue is circular as it is unlikely that someone will become a data journalist in a context in which data are scarce. The absence of data journalists also means that open data are not requested as often as they would otherwise be. Tajik officials, thus, may not be accustomed to data requests. But even if there was much demand for public data, our study revealed that the data would most likely be made available in PDF files, which would need additional processing to extract the necessary information for more advanced analysis. To address this challenge, Annex 5 contains some practical recommendations on how to convert such PDF files into machine-readable formats.

Fourth, data may be securitized. Some information of public interest is classified according to Tajikistan's special law as a state secret and hence impossible to obtain.²⁵ In other cases, the law does not

²³ Reporters without Borders, *2021 World Press Freedom Index*, <https://rsf.org/en/tajikistan>

²⁴ International Partnership for Human Rights, *The Price of Silence vs. the Cost of Speaking Out: Media Freedom in Tajikistan*, July 2020. <https://www.iphronline.org/wp-content/uploads/2020/07/ENG-Media-Report-TJ.pdf>

²⁵ Tajikistan's Law on State Secrets, 2014 (amended in 2015 and 2020). http://ncz.tj/system/files/Legislation/1095_ru.pdf

define certain types of data as secret but public institutions refuse to provide them and may offer them only to those journalists attending press conferences.

Fifth, when state agencies update their websites, often data may be removed from the public domain. Such deletions can only be prevented by archiving such websites.

Recommendations to Stakeholders

Recommendations to Media Organizations

- The study identified numerous datasets related to the socio-economic development of Tajikistan. To be incorporated in reports it is recommended that journalists familiarize themselves with the data sources included in the database and seek out other sources of data themselves.
- Given the lack of data-driven reporting, it is reasonable to assume that local journalists may lack the competencies to process data. To that end, data analysis workshops and seminars could be organized to enhance the level of data literacy of the Tajik media community.
- To encourage the development of data journalism, media organizations could set aside resources for data analysts.
- Data journalists may be limited by the technical capacity of their own outlets' websites. Media organizations should eliminate technical obstacles preventing data journalists from embedding various types of visualizations in their data stories.
- An important data-related skill involves data visualization. Zerkalo conducted an overview of the existing data visualization tools for journalists and included in Annex 4 a list of three data visualization tools that could be used by Tajik journalists interested in delivering data-driven stories.
- It is recommended that journalists download the datasets and take screenshots of the source webpage as in many cases data were deleted by authorities. Journalists could also retrieve saved pages from the Internet archive ([link](#)).

Recommendations for Government

- One recommendation concerns data-scarce sectors. The analysis of our database shows that there are numerous open data on banks and finance, but not so much on housing, productivity, and employment. That means that authorities overseeing certain policies are less likely to produce open data. One solution to this problem would involve steps to improve the institutional capacity of various agencies so that more data can be generated.
- A related recommendation concerns the sectors in which data exist, but are not published. Additional research would be needed to identify why certain data of public interest are not made available online.
- Another idea to be considered is for employees of state institutions to be trained to process and release data in formats suitable for analysis. To that end, the government could enter into partnerships with international organizations that could work on changing the way state institutions and academics in Tajikistan publish data by encouraging them to switch from PDF and images to machine-readable formats.

- Building on the experience of other countries, the government could ease access to open data by creating a single open data portal.

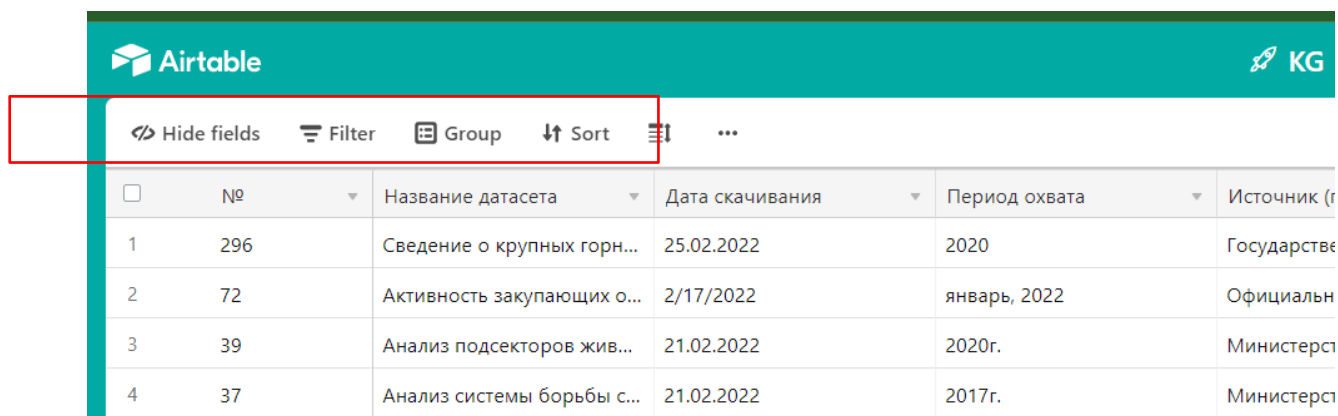
Recommendations to Donor Organizations

- Donor organizations could offer funding for data literacy courses that would increase the data competencies of Tajik journalists. Such training could help create a pool of journalists interested in pursuing more advanced preparation in data analysis.
- Donor organizations could encourage government agencies to produce more machine readable open data.
- Development aid providers could launch an open data initiative in the city of Dushanbe as a pilot project with city businesses and civil society organizations deriving significant benefits from such an initiative.

Annex 1: Instructions for Journalists on How to Use the Database

Airtable’s interface is easy to use and straightforward. Once you access the database on Airtable ([here](#)) follow some simple steps to get the data you need. The key tools for using the database can be found in the upper-left corner and include the following functions:

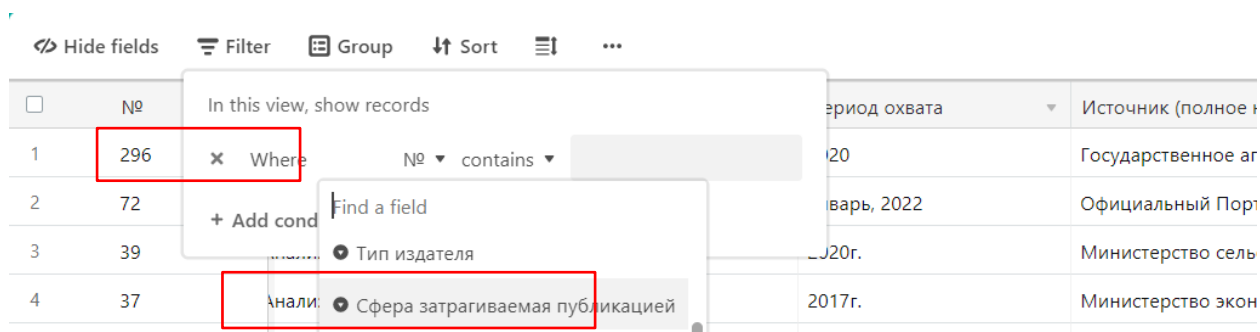
1. Filter
2. Group
3. Sort
4. Row height



№	Название датасета	Дата скачивания	Период охвата	Источник (г...
1	Сведение о крупных горн...	25.02.2022	2020	Государстве
2	Активность закупающих о...	2/17/2022	январь, 2022	Официальн
3	Анализ подсекторов жив...	21.02.2022	2020г.	Министерст
4	Анализ системы борьбы с...	21.02.2022	2017г.	Министерст

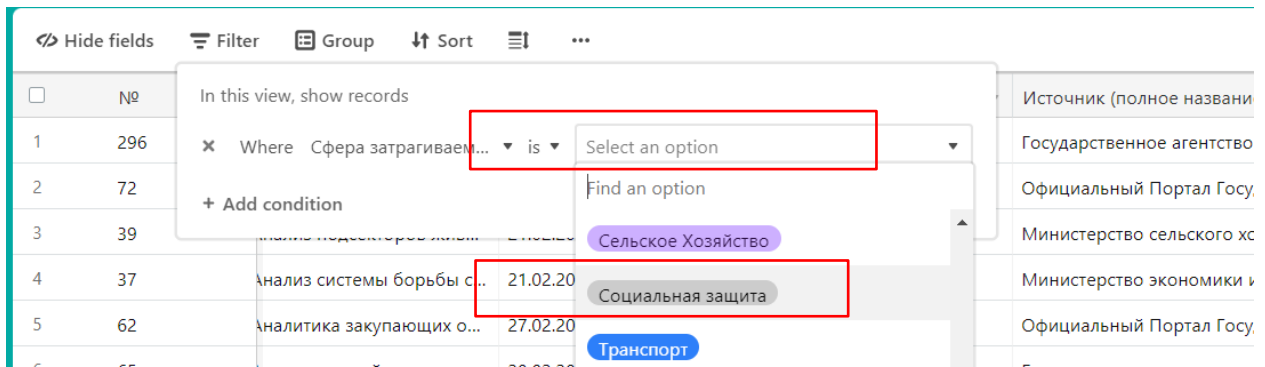
Hide fields – this function allows you to hide selected columns. For instance, when you have too many columns and they do not fit on your screen, Hide Fields allows you to remove from view the unnecessary columns.

Filter – this is a key function for working with data. The user can adjust the settings and find the data of interest. Example: Suppose that you want to see all the data related to social protection. To do that click on “Add Condition” and then under the setting “Where” choose “Publication Area.”



№	Название датасета	Дата скачивания	Период охвата	Источник (полное н...
1	Сведение о крупных горн...	25.02.2022	2020	Государственное аг
2	Активность закупающих о...	2/17/2022	январь, 2022	Официальный Порт
3	Анализ подсекторов жив...	21.02.2022	2020г.	Министерство сель
4	Анализ системы борьбы с...	21.02.2022	2017г.	Министерство экон

Next, select the option “Social protection” from the menu.

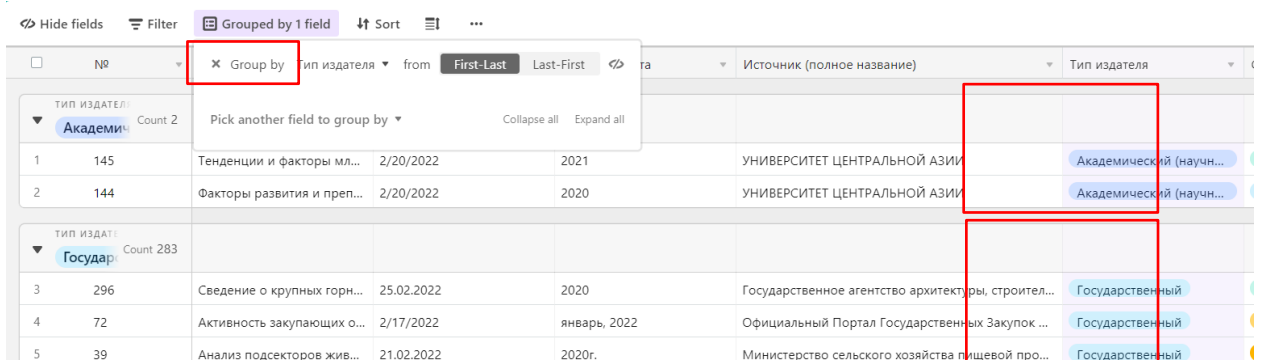


Now, the table will display solely the datasets pertaining to the area of social protection.

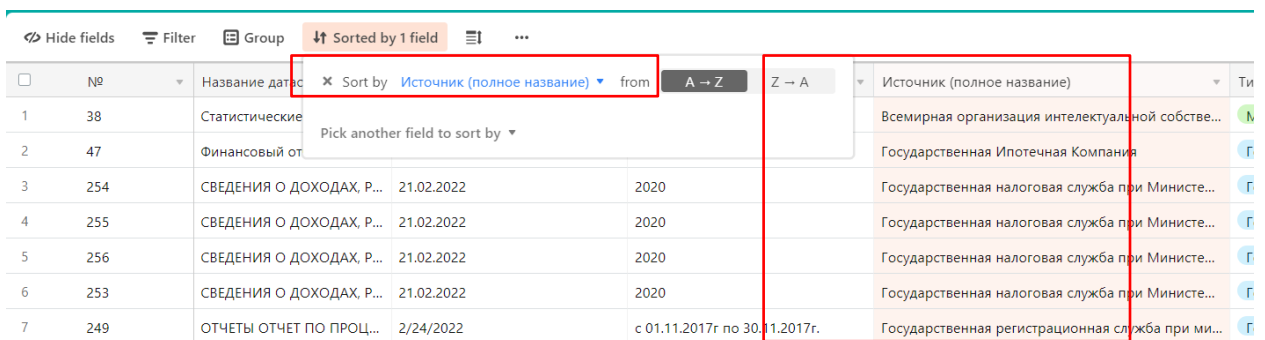
To remove the applied filter, press **x** to the left of the setting *Where*.

Group – the Group function enables the user to arrange the data based on the content of the column.

In the screenshot below, the data are classified by publisher type using the column “Type of Publisher.”



Sort – this option enables the classification of data based on the content of a particular column. In this case, the sorting is done in alphabetical order.



Row height – this option can be used to change the height of each row. One may select the minimum height of a row, so that data are displayed in a compact manner. Setting the maximum of row height increases readability as it provides more space between the text in each row.

Tags column – this option allows users to group datasets by the “Tags” column, which contains keywords. To customize the way information is being displayed in the database, the users should use the group function described above and choose the “Tags” column as the grouping column in order to find data on similar topics.

Annex 2: Checklist for Journalists Working with Data

To avoid bias, journalists should follow a set of ground rules when working with data. One way to accomplish that is by displaying critical thinking and asking the following questions:

Who collected the data? Are the organizations collecting the data trustworthy? Data collection is a complex and costly process. By finding out who collected the data, journalists may gain more confidence that the data are reliable. An organization known for data research is a positive sign, whereas an unknown organization presenting data needs requires additional scrutiny.

How was the data collected? Did the people collecting the data receive special training? Data collection often requires skills, which can be acquired via formal education and training. This does not mean that data enthusiasts cannot produce and publish data, but it does increase the likelihood that the data you are about to use are reliable.

Did the data producer use a rigorous methodology? Is the methodology publicly available for journalists to examine it? To verify this a journalist could ask the data holder to provide the methodology used to collect the data. As a rule of thumb, research methodologies are published together with the data. Similarly, a codebook should be made available by the data provider.

Who distributes the data? When and why? It may happen that the data producer is not the data distributor. For instance, a private company may have ordered a survey about its products. The sociological company conducting the survey provides the data, but the private firm decides to publish only the data that promote its products. In such cases, journalists are advised to ask who funded the survey and why are its results released at a particular time. There is plenty of evidence that in certain countries opinion polls are released before elections to influence the results. That is why certain countries imposed restrictions on the publication of opinion polls on the day of the elections.

How are the data presented and for what purpose? The way data are presented matters a lot. Often data users pick only the information supporting their claim, leaving out data that may disprove their story.

How were the questions phrased? In surveys, the wording of the questions has a great impact on the answer one gets. The respondent's understanding of a term or question may not correspond with the meaning attached to it by the researcher. Did the pollster mislead the respondents by choosing particular words?

How reliable are the data? For surveys: is the sample size large enough? Journalists could use one of the available online sampling calculators. The one offered by SurveyMonkey will do the job ([link](#)).

What is the permissible margin of error in cases of survey data use? Whenever a sample is selected, the margin of error indicates how confident the data provider is that the sample represents the whole population. One rule of thumb is to have 3 percent as the margin of error.

Are there alternative datasets that could confirm official data? Whenever possible, journalists should compare data from one source to data stemming from another source.

How representative is the sample? This question does not refer to the sample representing the population. Instead it asks the journalist and media organizations to pay attention to marginalized communities and disadvantaged groups. Have certain individuals been excluded from the sample such as people with disabilities, children, women, people in need, minority groups etc.?

Beware of the ecological inference fallacy. This happens when inferences about individuals in a group are based on findings about the group as a whole.

When interpreting data it is important to keep in mind that **averages are often useless** as they do not tell the whole story. For instance, in countries with high economic inequality, a high average GDP per capita does not necessarily mean that all of the population is prosperous.

Annex 3: Data Request Letter (Original)

ЛОГОТИПИ РЕДАКСИЯ

Сод _____

Ба кй:

Ба кучо:

ДАРХОСТ

_____ (номи ВАО) дар асоси моддаи 18 Қонуни Ҷумҳурии Тоҷикистон “Дар бораи мурочиати шахсони воқеӣ ва ҳуқуқӣ” бо мақсади омода намудани маводи таҳлилий аз Шумо хоҳиш дорем маълумотҳои зеринро дар шакли Excel ва ё csv аз то сол дар ҷадвали зерин пешниҳод намоед:

1.

Сана:

Ному насаб, вазифа, ташкилот

Телефон барои тамос

Annex 4: Some Data Visualization Tools for Journalists

1. [Datawrapper.de](https://datawrapper.de)

Description: Datawrapper allows users without any coding experience to create visually appealing charts, maps, and tables with a few clicks. It can help data journalists convey their story and fulfill their role as guardians of the public interest. For an introduction to Datawrapper, please click [here](#). If you are interested in more advanced topics, we suggest you use the Datawrapper Academy resources ([link](#)).

2. [RawGraphs.io](https://rawgraphs.io)

Description: RAW Graphs is an open-source data visualization framework built with the goal of making the visual representation of complex data easy for everyone. It does not require any coding experience. RAW Graphs provides the missing link between spreadsheet applications (e.g. Microsoft Excel) and graphics editors (e.g. Adobe Illustrator, Inkscape, Sketch). For quick tutorials check [this link](#).

3. [Flourish](https://flourishhq.com)

Description: Launched in 2018, Flourish builds on R and does not require coding skills. The company made the tool free for journalists. For examples on how to use Flourish click [here](#).

Annex 5: Recommendations on PDF Conversion

Zerkalo has the following recommendations for data journalists with regards to Portable Document Format (PDF) datasets. If the PDF was created by exporting from a Windows app, then the solution is to:

Option 1: Download the PDFs Open the document with Adobe Acrobat. From the Adobe Acrobat DC menu, the user may choose the option Save As and then pick Word or Excel as the desired format of the final document. This solution is the recommended one as the quality of the conversion is high.

Option 2: If you don't have Acrobat reader installed follow these steps. Download the PDF dataset on your Desktop. Right-click on the document icon and open it with Microsoft Word. Word will then convert and open the PDF as a Word document.

Option 3: Use reliable and safe online conversion websites. In such cases, the user may upload the PDF to a platform and download the converted document.

Suggested platforms:

Free Adobe Converter: <https://www.adobe.com/ca/acrobat/online/pdf-to-word.html>

Smallpdf.com: <https://smallpdf.com/pdf-to-word>

IlovePDF: <https://www.ilovepdf.com/>

In cases in which PDFs consist of scanned images, extracting the data might be more difficult for data journalists. Such PDFs can be identified easily as users cannot select the text within the document. They may choose one of the following solutions.

Option 1: Use optical character recognition software tools to convert the PDF to another format. In such cases there might be some errors in the converted file. Hence the data journalist would need to verify the accuracy of the conversion. Word has an in-built OCR, which may convert the scanned PDF into editable text. Another option is the free online Optical Character Recognition converter - [Img2TXT](#). It allows the conversion of scanned images into text (up to 8 Mb). If Word and [Img2TXT](#) do not work, you may purchase ABBYY Finereader, which is one of the most recommended OCR tools.

Option 2: Take screenshots of the scanned data and use one of the optical recognition software tools available online such as ABBYY FineReader.

Option 3: Manually introduce the data in machine-readable files.