



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



# Report Mapping Open Data in Kyrgyzstan

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

<b>Executive Summary</b>	<b>4</b>
<b>1. Introduction</b>	<b>6</b>
<b>2. The State of Data Journalism in Kyrgyzstan</b>	<b>8</b>
<b>3. Mapping Open Data in Kyrgyzstan</b>	<b>9</b>
<b>4. Challenges for Data Journalists in Kyrgyzstan</b>	<b>14</b>
<b>5. Recommendations</b>	<b>16</b>
Annex 1: Database Instructions for Journalists	18
Annex 2: Checklist on How to Work with Open Data	21
Annex 3: A Sample Data Request Letter	23
Annex 4: Top-3 Data Visualization Tools	24
Annex 5: Recommendations on PDF Conversion	25

## Executive Summary

**Context:** New data sources related to socio-economic development should improve journalistic reporting. Still, despite an abundance of open data in Kyrgyzstan, journalists may face delays in getting quick access to important datasets.

**Goal:** To facilitate the access to development-related information for journalists working with data on socio-economic issues, Zerkalo conducted a mapping study, identifying all the open data sources in Kyrgyzstan and putting together a comprehensive database accessible on Airtable ([link](#)) and Google Spreadsheets ([link](#)).

**Methods:** Relying on a two-stage desk research approach, the Zerkalo team constructed a database containing open data sources in Kyrgyzstan. The main data collection phase for the mapping study was conducted from February 14 to March 10, 2022. Overall, the research team examined 1986 hyperlinks to 131 websites, holding relevant datasets. As a result, Zerkalo identified 389 datasets on socio-economic themes, which were included in the database.

### Results:

- The mapping revealed that most datasets are held by the following organizations: the National Statistical Agency of the Republic of Kyrgyzstan (140), the Ministry of Economy and Commerce (21), Central Bank (21), the Ministry of Finance (17), The Antimonopoly Agency (11), the World Bank (10) etc.
- Another finding concerns the format of the datasets: Web (237), Excel (48), PDF (61) and Word (10).
- The mapping shows that there are more open datasets on topics such as Agriculture (75), Employment (51), Economic Forecasting (38), Banks and Finance (23), Prices (20), Budget (19), Business (12), Regional Development (11) etc.
- In terms of publisher, datasets originate from state agencies (290) and international organizations (54), international NGOs (8), local NGOs (2) and academic institutions (2).
- As regards language, the datasets are in Russian (302), English (52) and Kyrgyz (2).
- For easier visualization, the database can be accessed on Airtable ([click here](#)). Instructions on how to navigate the database on Airtable are provided in Annex 1.

**Conclusion:** Taken together, our report finds that the amount of open data available to the community of data journalists in Kyrgyzstan is growing. More and more datasets are being published online mostly due to the numerous projects involving partnerships between state institutions and international providers of development aid. Building on the findings of the mapping exercise, the report formulates below a couple of recommendations for various stakeholders.

## **Recommendations**

### To Media Organizations:

- Journalists would make the best use of the database by exploring its content on Airtable ([link](#))
- Media organizations should offer targeted training for potential data journalists focusing on how to start research on a data-related topic, how to conduct basic data analysis and exploration, and how to write data-driven stories.
- A key skill for data journalists includes visualization. Annex 4 contains a list of recommended data visualization tools selected by Zerkalo as well as links to tutorials on how to use them.
- When working with data, journalists should follow certain rules of thumb to ensure data quality and reliability (Annex 2).

### To Government:

- The government should continue implementing the open data programs, offering data free of charge, online, in formats amenable to advance processing.
- As most open data cover certain themes and originate from only a few institutions, one recommendation would be to prompt officials from other agencies to release sector-specific data.
- Government should invest more in its capacity to consistently collect and publish data. It should increase public awareness on the role and importance of open data for socio-economic development providing such data in an easily and safely accessible machine-readable format.

### To Donor Organizations:

- Donor organizations should integrate data publishing into development projects as well as fund projects aimed at consolidating the capacity of civil society organizations to produce data.
- Donor organizations could offer training opportunities for the officials in charge of collecting and publishing data so that open data are provided in formats suitable for data analysis.

## 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 will be implemented, connecting policymakers to the public. This mechanism works 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.<sup>1</sup> By contrast, developing countries with fledgling media markets are more likely to suffer from a lack of reliable and unbiased information about various public policies.<sup>2</sup>

Data journalists play a special role in such information ecosystems. To hold public officials accountable and monitor the implementation of public policies, journalists need to inform the public. In a world in which data are increasingly abundant, journalists working with large datasets can fulfill this mission by publishing data-driven stories about the latest socio-economic policies in a particular context. Unlike their peers reporting on socio-economic issues in a traditional manner, journalists working with data have often been overlooked by development organizations. Such neglect is probably linked to the relative recency of data journalism as a distinct specialization. The underlying assumption of this study is that the dominant approach needs to focus on empowering journalists to hold state officials accountable by providing evidence-based information.

For data journalists to emerge as a community, accessible data are crucial. Reliable and open data enable journalists to offer high quality reports. To avoid further confusion, 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.”<sup>3</sup> Another aspect concerns machine-readability. 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” in developing countries.<sup>4</sup> While in developed countries there is a wealth of data sources that journalists can retrieve to write their stories, data availability and accessibility are major issues in developing countries.

Four problems stand out:

- First, open government 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

---

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

<sup>2</sup> Christopher Starke, Teresa K. Naab, & Helmut Scherer. “Free to Expose Corruption: The Impact of Media Freedom, Internet Access and Governmental Online Service Delivery on Corruption.” *International Journal of Communication* [Online], 10 (2016): 21.

<sup>3</sup> See *The Open Data Handbook*, Open Knowledge Foundation, <https://opendatahandbook.org/guide/en/what-is-open-data/>

<sup>4</sup> Audrey Ariss, Machine-readable Open Data: How it’s Applicable to Developing Countries, Data Blog, March 2, 2017. <https://blogs.worldbank.org/opendata/machine-readable-open-data-how-it-s-applicable-developing-countries>

points collected over longer periods of time. Examples include GDP figures over the last decades or the World Value Surveys. Micro-level refers 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 with 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. 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. While the needs of data journalists in each country may be different, 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 journalists in this field.<sup>5</sup>

Successful data journalism produces public interest stories. In this sense, the work of data journalists is situated at the intersection of several distinct areas. They find data and construct databases, develop stories centering on data, use advanced data visualization tools, and apply innovative technologies to process data.<sup>6</sup> 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 presents the results of the study. The structure of the report is as follows. Section 2 discusses the state of data journalism in Kyrgyzstan, highlighting the existing legal and institutional framework. Section 3 describes the research methodology and key findings. Section 4 points out potential challenges for data journalists. Building on our results, the last section includes several recommendations to media organizations, government and donors.

---

<sup>5</sup> See Andrea Abellan, The State of Data Journalism Survey 2021, <https://datajournalism.com/read/blog/data-journalism-survey-2021>

<sup>6</sup> 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)

## The State of Economic and Data Journalism in Kyrgyzstan

Over the last decade, Kyrgyzstan made significant progress with regards to open data availability. The Kyrgyz state has benefitted from the presence of various universities and international organizations, which have introduced data education courses in the curriculum, generating much interest toward the topic. For instance, the American University of Central Asia and some state universities offer courses in data journalism. Data-related workshops are regularly held for journalists in Bishkek.<sup>7</sup> This led to the publication of numerous data-driven stories, which demonstrate that there is a growing community of data enthusiasts in Kyrgyzstan.<sup>8</sup> Moreover, the School of Data Kyrgyzstan, a local NGO promoting data use and access, is part of a global network committed to advancing data literacy among civil society organizations. Another notable initiative from the country is the effort to compile a free library of datasets for each of the five Central Asian republics.<sup>9</sup>

Our assessment revealed that the Kyrgyz state in partnership with international organizations has developed the capacity to organize and present data online effectively. For instance, the websites of the National Statistical Committee of the Kyrgyz Republic ([link](#)), the Ministry of Economy and Commerce ([link](#)), and the National Bank ([link](#)) are well-organized and user-friendly. Such improvements are the result of a series of reforms which began in the 2010s with the attempt to publish information about public procurement contracts, making the flow of budgetary financial flows more transparent.<sup>10</sup> Other initiatives such as the National Strategy for Introducing Electronic Governance were launched in 2014 with substantial financial support from the UNDP.<sup>11</sup> In 2017, Kyrgyzstan made further progress by joining the Open Government Partnership. Currently, the government is continuing its efforts to implement digital governance projects such as the “Digital Kyrgyzstan 2019-2023” concept, the system of inter-agency electronic interaction “Tunduk”, which will enable the sharing of more open data. Another project concerns the modernization of the statistical data collection and distribution system.<sup>12</sup> This should lead to more open data being available for data

---

<sup>7</sup> See Hackathon on Data Journalism “Hack Poverty”, February 12, 2018. <http://project533785.tilda.ws/>. For other Open Data projects in Kyrgyzstan see <http://opendata.kg/page19106310.html>

<sup>8</sup> See for instance the following report on the deteriorating elementary schools in Kyrgyzstan: Valentina Galich, Diana Svetlichnaya, Taalajbek Talantbekullu, Ekaterina Afanasieva, Sezim Digerbaeva, Zhuzumkan Askhatbekova “The Crack in the Wall: Why Children in Kyrgyzstan Study in Extreme Environments,” Radio Azzatyk. <https://rus.azattyk.org/a/schools/31439009.html>; Also see the data-rich Tazabek online resource at <https://www.tazabek.kg/>; For a list of data-driven publications see here: <http://opendata.kg/page19725884.html>; Kloop.kg has a section dedicated to data-driven stories: <https://kloop.kg/blog/category/data-zhurnalistika/>

<sup>9</sup> See the School of Data organization in Kyrgyzstan and their database <http://opendata.kg/page19981037.html>

<sup>10</sup> See Jyldyz Kasymova & Marco Aurelio Marques Ferreira & Suzanne J. Piotrowski, 2016. "Do Open Data Initiatives Promote and Sustain Transparency? A Comparative Analysis of Open Budget Portals in Developing Countries," Public Administration and Information Technology, in: Jing Zhang & Luis Felipe Luna-Reyes & Theresa A. Pardo & Djoko S. Sayogo (ed.), *Information, Models, and Sustainability*, pp. 137-155, Springer.

<sup>11</sup> See Open Public Data and its Impact on the Development of Kyrgyzstan, UNDP website, November 2014. <https://www.kg.undp.org/content/kyrgyzstan/en/home/presscenter/pressreleases/2014/11/17/open-public-data-and-its-impact-on-development-of-kyrgyzstan.html>

<sup>12</sup> “Zapushchen Proekt po Modernizatsij Nalogovogo Administrirovanya i Statisticheskoy Sistemy,” National Statistical Committee Website, March 9, 2021.



journalists over the next three years. Moreover, the World Bank is implementing Pillar 2 of the Effective Governance for Economic Development (EGED) programme funded by the UK Government with the aim to support the design and delivery of evidence-based economic policies. This will include assisting governments in their efforts to generate better data, use data for policy design and implementation, improved coordination within government on policy implementation, and better engagement with citizens. The World Bank will focus on improving the regional development policy by mapping micro-level data at village (i.e. ayyl aymak) and town level, made available as interactive maps on the platform of the National Statistics Committee.

Yet, even for Kyrgyz journalists the available open data is not easy to find online. Access to public information in Kyrgyzstan is regulated by the Access to Information Law (Law Nr. 213 from 2006 last amended in 2017), which provides the legal framework and the rules to be followed when requesting information from authorities.<sup>13</sup> Law Nr. 213 also applies to data, which Kyrgyz journalists may request from various agencies at national, regional, and local level. The law imposes time constraints on public officials, establishing that information of public interest should be provided within two weeks from the day it was requested. Most journalists work under tight deadlines and have to operate under pressure. It is thus challenging for them to request and quickly find the data they need. They need to go through multiple tables and indicators on disparate websites until they identify the necessary information. In this regard, the pooled database constructed by the Z-Analytics team provides straightforward access to economic data from multiple online sites as well as contact information to request additional data. Journalists, academics and interested citizens writing about economic development in Kyrgyzstan can validate the information about key economic indicators by comparing it to similar data from other sources.

## Mapping Open Data in Kyrgyzstan

The “Open Data Mapping in Kyrgyzstan” project put together a database containing key open data sources. Carried out over the February-March 2022 period by an international and multicultural team from Kyrgyzstan, Tajikistan, and Canada, the project focused on open data in the area of economic development.

To construct the database, Z-Analytics adopted a two-stage desk research approach. First, we identified all the sources containing relevant economic data about Kyrgyzstan. We began with an evaluation of the published data by state agencies. Z-Analytics operators indexed and browsed 49 governmental websites. Then, Z-Analytics surveyed data from international organizations, local and regional NGOs, identifying 389 datasets as of March 18, 2022. Once a data source was located, the operators followed a standard protocol,

---

<http://www.stat.kg/ru/news/zapushen-proekt-po-modernizacii-nalogovogo-administrirovaniya-i-statisticheskoy-sistemy/>

<sup>13</sup> Ministry of Justice, Law Nr. 213, 2006. <http://minjust.gov.kg/ru/content/672>

recording its characteristics. The last batch of datasets was added using a keyword search on major search engines (Google, Yandex) and a targeted search within select data repositories.

In cases where governmental data were not open, the operators proceeded to the second stage, requesting access. The template for such data requests is included in Annex 3. In some instances, the research assistants contacted officials by phone. At this stage, they documented 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 with data organized into rows and columns as seen in Figure 1. The structure of the database is identical to a Google or Excel spreadsheet with an added layer of complexity consisting of a series of drop-down lists, which distinguish it from existing databases. 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. It can be easily searched, sorted, downloaded, and updated.

Figure 1. Database columns.

Users can access the database in two ways. They can open 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 institutional holders. Both the open and on-demand data sections include contact information in case database users need more details and want to contact the relevant data holders.

The second way to access the database is via Airtable, a useful organizing tool allowing for the easy retrieval of data. The Airtable version ([link](#)) is more user-friendly. 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 study it. Instructions on how to access the Airtable data can be found in Annex 1.

Below are some of the findings regarding the open data landscape in Kyrgyzstan. Figure 2 shows that the identified open data were concentrated in the following fields: agriculture, labor, banks and finance, economic forecasting, and prices.

*Figure 2. Datasets by sector.*

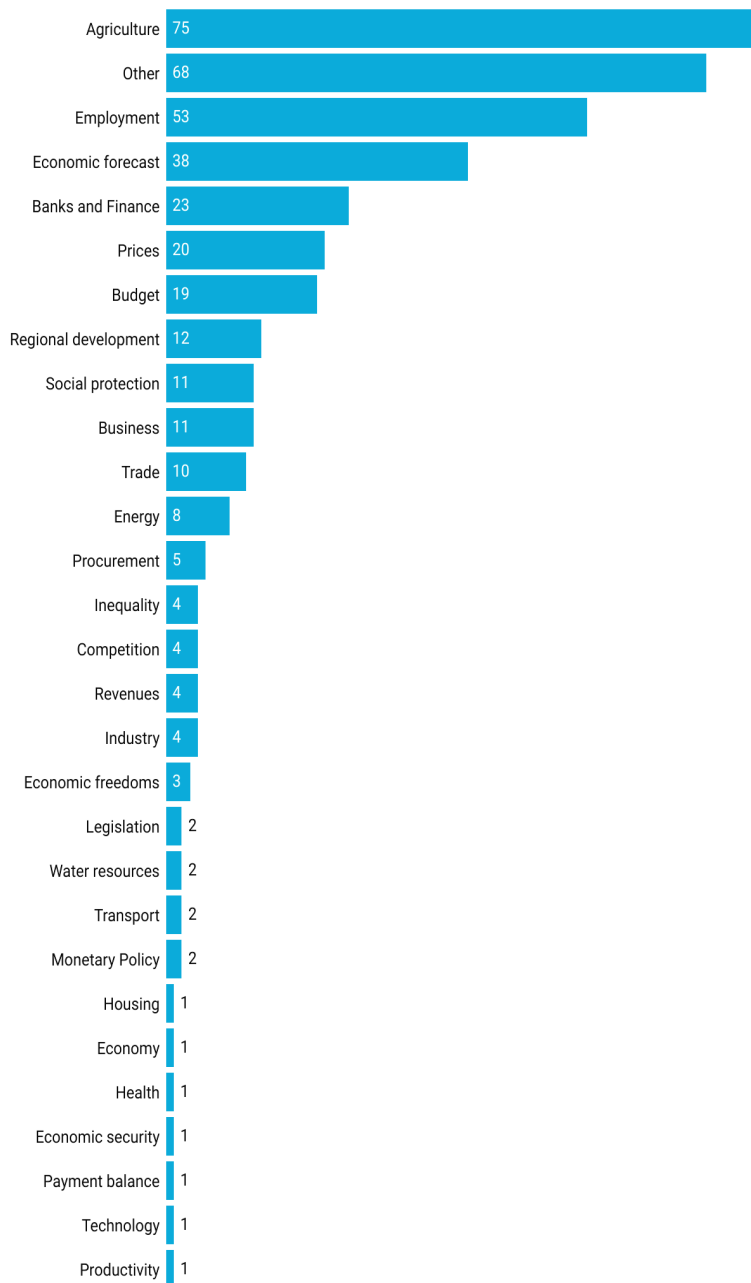


Chart: Ion Marandici • Source: Zerkalo Center • Created with Datawrapper

Most of the datasets identified are embedded in a web page with some available as PDF files (63) and only 51 are downloadable in an Excel format (Figure 3). Since PDF documents are not amenable to advanced data analysis, in Annex 5, we provide recommendations to data journalists on how to extract data from such files.

Figure 3. Dataset Format.



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

Another finding concerns the origins of the datasets. As seen in Figure 4, most of the datasets are from state authorities with only a few from international NGOs, international organizations, academic and local NGOs. This suggests that data journalists may face difficulties in corroborating official data with data from alternative sources.

Figure 4. The type of dataset publisher.

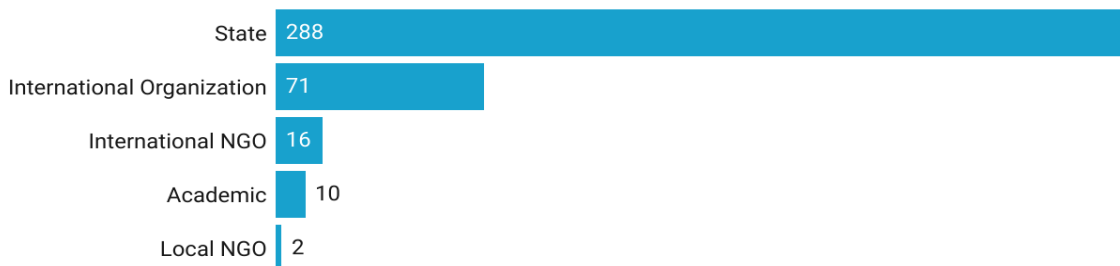
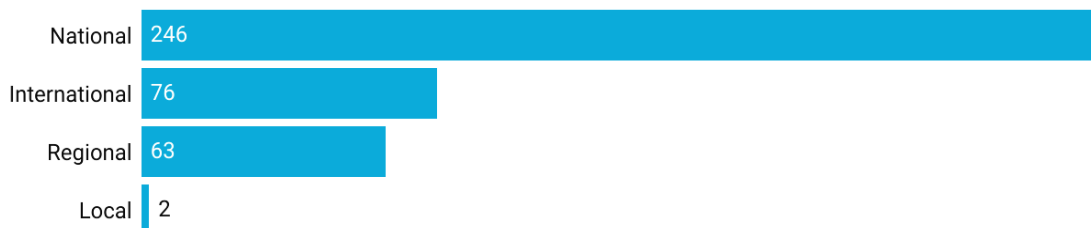


Figure 5 indicates that most of the datasets are at the national level (246) with some at regional (63) and international levels (76). The database contains only two local datasets.

Figure 5. Dataset level.



The database includes a special column, specifying the next expected release. Figure 6 provides some insights on the updating frequency of the cataloged datasets. Most of them are updated annually, while some of them result from one-time projects, lacking information about the next release.

Figure 6. Frequency of Data Releases.

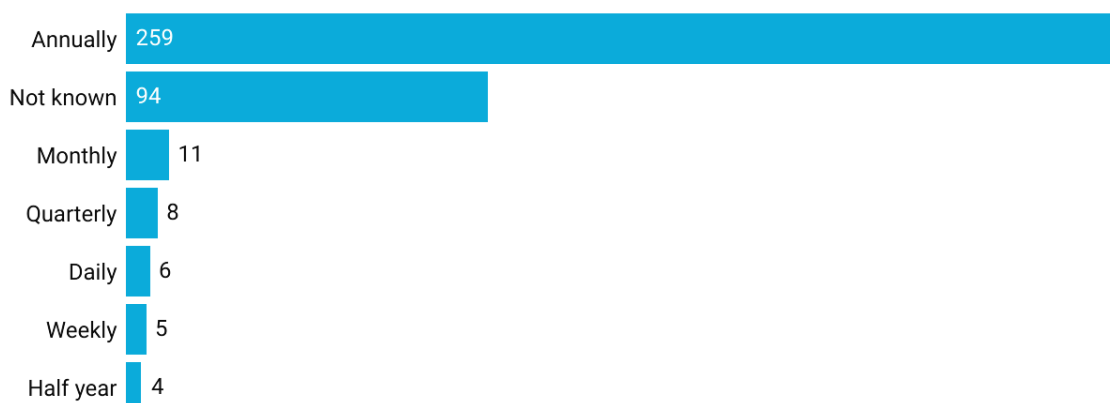


Chart: Ion Marandici • Source: Zerkalo Center • Created with Datawrapper

As part of the data mapping exercise, researchers from Zerkalo have contacted state officials requesting supplementary information in those cases in which the datasets were outdated. Reactions to such data requests varied. While the Ministry of Finance published on their website the requested data, some institutions removed data from their websites or redirected data requests to other officials. For instance, in its official response, the deputy-Minister of Justice stated that the institution is not mandated to hold and release data about judicial matters, redirecting the request to the Supreme Court of Kyrgyzstan. In the case of the Ministry of Emergency Situations, the data were provided as a PDF file. Another issue concerning data access has to do with the preparedness of the officials in charge of releasing the data. In several cases, the public employees simply did not know how to respond to the request. Zerkalo researchers contacted them by phone to describe the scope of the data request and spell out how it should be implemented. On a positive note, it should be mentioned that none of the institutions contacted for additional data attempted to sell data or request informal payments. Also, in most cases, the institutions responded to data requests in a timely manner. Moreover, numerous key data are freely available for download as brochures and booklets from the website of the National Statistical Committee of Kyrgyzstan.<sup>14</sup>

Open data are often removed from websites, disappearing from the web. At times, this is the result of the technical limitations of various website designs. In other cases, data are deleted from the public domain, because journalistic investigations revealed some inconvenient aspects about a public official. 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 will enable 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 was removed from the web.

<sup>14</sup> For a list of publications released by the Statistical Committee of Kyrgyzstan see the following link: <http://www.stat.kg/ru/publications/>

## Challenges for Data Journalists in Kyrgyzstan

Despite an abundance of data sources about various aspects of the Kyrgyz socio-economic development, there are some areas which present major challenges.

One area that could be improved concerns data releases. Most of the datasets in our database are updated annually, which means that the data may be outdated by the time journalists can use them. In such cases, journalists would need to request updated information from authorities using the template provided in Annex 3.

Another issue has to do with the official reactions to journalistic work. Despite the growing community of data journalists in Kyrgyzstan, the key problem is not so much related to data literacy, but rather pertains to ways in which authorities react to data-driven investigations. A pertinent example is the case of Bolot Temirov, who was arrested in 2022 after he published an investigation report about corruption.<sup>15</sup> Similarly, two media organizations reported hacking attacks against their staffers, while the 2020 law combating fake news introduced new limitations on the freedom of expression in Kyrgyzstan.<sup>16</sup> Similarly, Article 19, an international human rights organization monitoring the freedom of expression, noted multiple incidents in which access to information was not respected and journalists were harassed.<sup>17</sup> A report using data from 2015-2021 revealed that only a few of the individuals attacking journalists were ever prosecuted.<sup>18</sup>

Then there is the challenge of having access only to official data. Journalists often need to verify official statements by relying on alternative sources of data. Such data are, however, often simply absent. This is similar to the well-documented problem described as *press-release journalism*, when media organizations tend to rely on official press releases rather than conduct their own research on a particular topic.<sup>19</sup>

A related issue is the absence of legal consequences for those officials, who refuse to provide the data. Often officials refuse to release the data under the pretext that they constitute a state secret or simply offer no explanation for their decision.<sup>20</sup>

---

<sup>15</sup> Davron Hotam, Kyrgyz Media Protest Arrest of Investigative Journalist, VOA, January 25, 2022. <https://www.voanews.com/a/kyrgyz-media-protest-arrest-of-investigative-journalist/6412175.html>

<sup>16</sup> Kyrgyzstan Parliament Approves “False Information” Bill, Committee to Protect Journalists, August 10, 2021. <https://cpj.org/2021/08/kyrgyzstan-parliament-approves-false-information-bill> ; Government’s Pressure against Media Mounts in Kyrgyzstan, OCCRP, February 21, 2021. <https://www.occrp.org/en/daily/16000-government-s-pressure-against-media-mounts-in-kyrgyzstan>

<sup>17</sup> Kyrgyzstan: Attacks on journalists and access to information following disputed election, Article 19, October 8, 2020.

<https://www.article19.org/resources/kyrgyzstan-attacks-on-journalists-and-access-to-information-following-disputed-election/>

<sup>18</sup> Saadat Tologonova, Zhurnalistov v Kyrgyzstane postoyanno izbivayut, presleduyut i sudyat - Data-Analiz, Kloop.kg, February 19, 2022. <https://kloop.kg/blog/2022/02/19/zhurnalistov-v-kyrgyzstane-postoyanno-izbivayut-presleduyut-i-sudyat-data-analiz-kloopa/>

<sup>19</sup> Elira Turdubaeva, Media landscape in Kyrgyzstan: Caught between elite capture and control of political and business interests, The Foreign Policy Center, March 1, 2021. <https://fpc.org.uk/media-landscape-in-kyrgyzstan-caught-between-elite-capture-and-control-of-political-and-business-interests/>

<sup>20</sup> “Kyrgyz Journalists Restricted Access to Information,” Analytical Center for Central Asia, August 9, 2019. <https://acca.media/en/kyrgyz-journalists-restricted-access-to-information/>

There is then the challenge of getting access to machine-readable tabular data as the most preferred format to work with data. Some datasets are provided as pdf documents and it takes time to convert such data into Excel or csv tables.

## Recommendations to Stakeholders

### Recommendations to Media Organizations

- The “Mapping Open Data” project has put together an extensive collection of datasets. It is recommended that journalists make the best use of it by exploring it on Airtable ([link](#)).
- The project would bring additional benefits if media organizations could offer instructional workshops for journalists in three additional areas:
  - how to start the research on a data-rich topic
  - how to conduct basic data analysis and data exploration
  - how to write data-driven 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 Kyrgyz journalists interested in delivering data-driven stories.
- Annex 2 includes practical recommendations for journalists on how to avoid common pitfalls when working with data.
- It is recommended that journalists download the datasets and take screenshots of the source webpage as in many cases data were deleted by state officials. Journalists could also retrieve saved pages from the Internet archive ([link](#)).
- Journalists should check the website of the National Statistical Committee as it would release over the next year micro-level data at village (i.e. ayyl aymak) and town level.

### Recommendations to Government

- The government should further implement the principles of open government data by providing complete datasets in a timely and accessible manner. This includes publishing data in formats that allow more advanced processing.
- The government should offer non-proprietary data, eliminating any fees or other restrictions such as identity checks.
- The government should provide the data online in a stable format on a permanent website. The open data should be safe to open and accompanied by proper documentation (such as codebooks).
- The government should evaluate the impact on open data distribution of the ongoing changes related to the Sanarip Aymak reforms implemented by the Statistical Committee of Kyrgyzstan. Specifically, it would be useful to know whether these changes will have different effects across different types of datasets contingent on their collection method.



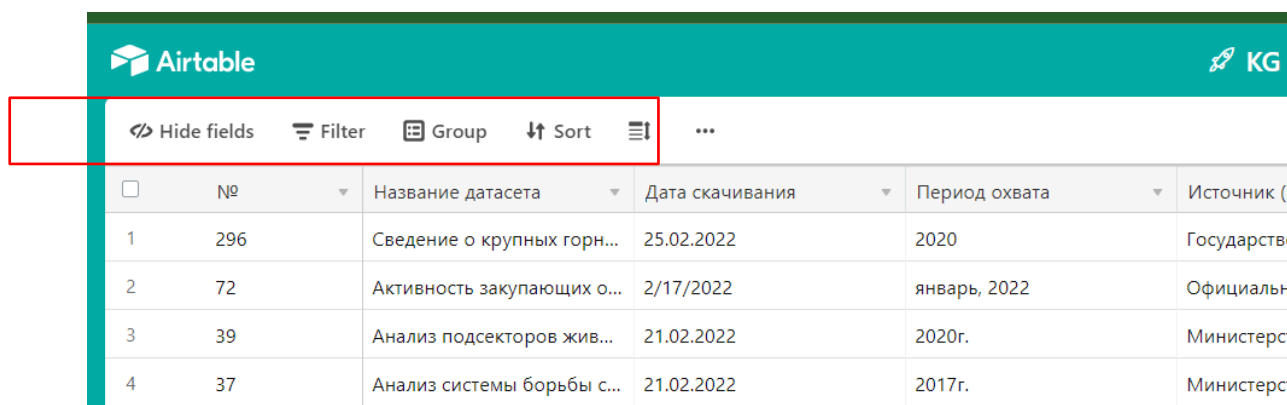
### **Recommendations to Donor Organizations**

- Donor organizations should unlock the potential of open data and participatory governance by funding projects that increase the amount of data placed in the public domain.
- Donor organizations should continue contributing to the improvement of data literacy among the Kyrgyz journalists by providing opportunities for exchanges among data journalists from various countries, funding data workshops and rewarding excellent data journalistic work.
- Donor organizations could offer training opportunities for the officials in charge of collecting and publishing data at various levels. The goal is to reduce the costs of data publishing and to eliminate fears that data publishing may backfire.

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

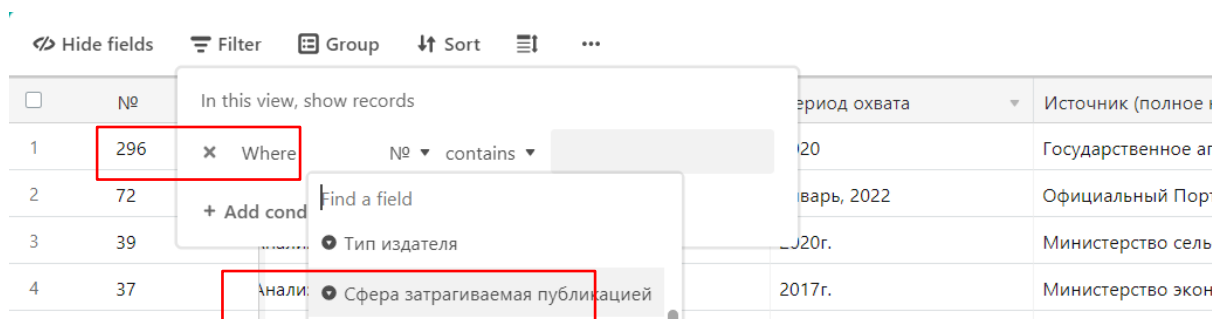
Airtable’s interface is easy to use and straightforward. Once you access the database on Airtable ([here](#)) follow some simple steps to get the data you need. The key tools for using the database can be found in the upper-left corner and include the following functions:

1. Hide Fields
2. Filter
3. Group
4. Sort
5. Row height

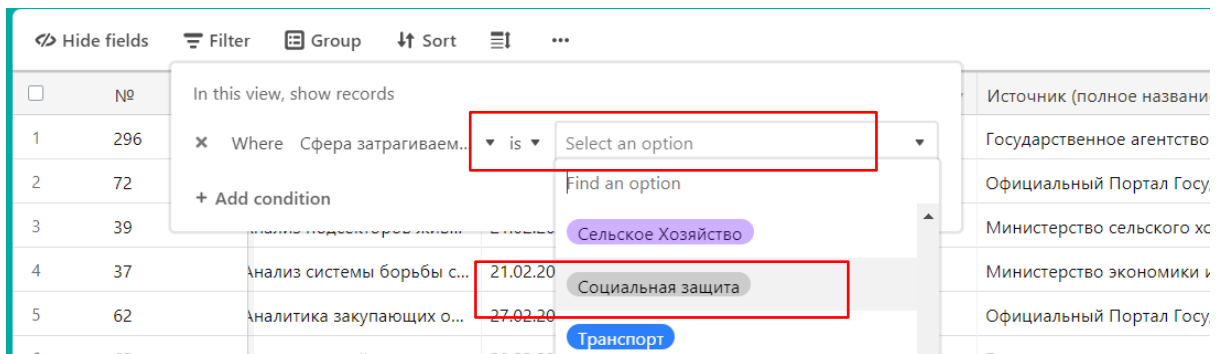


**Hide fields** – this function allows you to hide selected columns. For instance, when you have too many columns and they do not fit on your screen, Hide Fields allows you to remove from view the unnecessary columns.

**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.”



Next, in the “select an option” menu select the desired sphere, for instance “Social protection.”



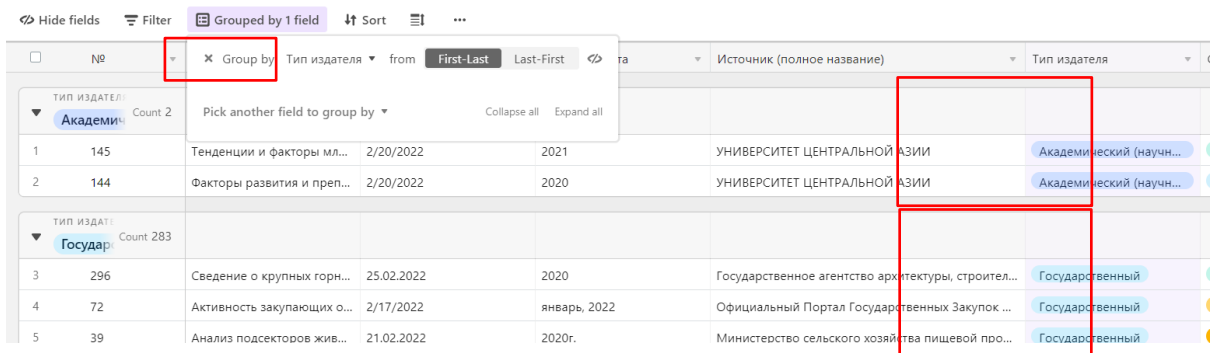
Now, the table will display solely the datasets pertaining to the area of social protection.

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

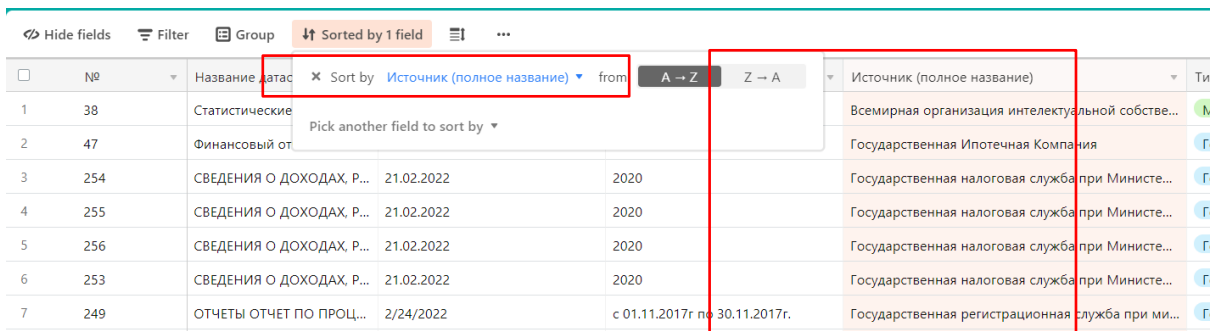
**Group** – the Group function enables the user to arrange the data based on the content of the column.

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

Similarly, the datasets can be grouped by any column of interest.



**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.

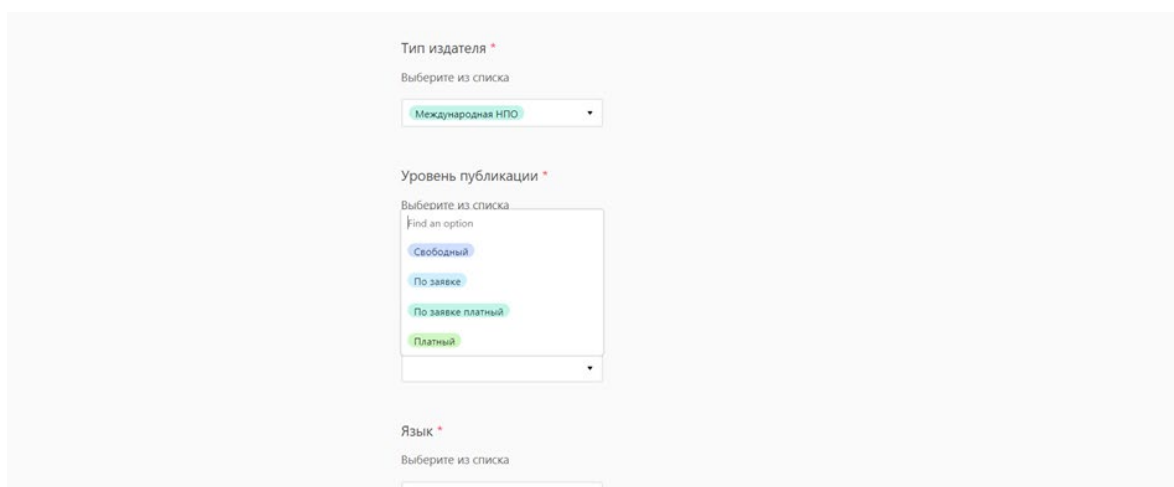


**Row height** – this option can be used to change the height of each row. One may select the minimum height of a row, so that data are displayed in a compact manner. Setting the maximum of row height increases readability as it provides more space between the text in each row.

**Tags column** – this option allows users to group datasets by the “Tags” column. 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.

### Data Entry Form

Users can contribute to the database by completing a data entry form. The data entry form consists of the same field classifications as the database. It includes three types of fields: text, drop-down lists, and multiple choice. The fields marked with an asterisk (\*) are mandatory.



The image shows a data entry form with three mandatory fields, each marked with an asterisk (\*). The first field is 'Тип издателя' (Publisher Type), with a dropdown menu showing 'Международная НПО' (International NGO). The second field is 'Уровень публикации' (Publication Level), with a search box and a list of options: 'Свободный' (Free), 'По заявке' (On request), 'По заявке платный' (Paid on request), and 'Платный' (Paid). The third field is 'Язык' (Language), with a dropdown menu.

Once all the mandatory fields are filled out, the user should click on the “Submit” button to have the manually entered data automatically transferred to the main database.

## **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 in order 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 are presented in a way that supports a certain political or economic agenda with inconvenient data being left out. In some cases, data presentations are designed to go viral.

**How were the questions worded?** In surveys, the phrasing 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. It is worth asking whether the pollster misled 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, LGBT etc.?

**Beware of the ecological inference fallacy!** This error may occur 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: Template of a Data Request Letter

Исх. №:

«\_\_\_» \_\_\_\_\_ 20\_\_ г.

Полное наименование государственного органа

Или если обращение конкретно на чье-либо имя, то должность и ФИО

*Например:*

*Министерство финансов Кыргызской Республики*

*или*

*Министру финансов Кыргызской Республики*

*Бакетаеву А.К.*

### ЗАПРОС

(Наименование запрашиваемой организации) на основании Закона КР «О доступе к информации, находящейся в ведении государственных органов и органов МСУ» (28.12.2006) для создания аналитического материала, а также учитывая, что запрашиваемые данные представляют общественный интерес, просит предоставить в машиночитаемом формате текстовом/электронном формате следующую официальную информацию. Просим вас предоставить данные в формате Excel или csv с ... по .. год со следующими столбцами:

1. Бюджет КР на 2022 год и прогноз на 2023-2024 гг.

Напоминаем, что согласно ст.11 закона «О доступе к информации, находящейся в ведении государственных органов и органов МСУ», ответ на письменный запрос о предоставлении информации должен носить исчерпывающий характер, исключая необходимость повторного обращения заинтересованного лица по тому же предмету запроса.

Все данные могут быть представлены в электронном виде. Просим ответ отправить на почту:

\_\_\_\_\_

Если у Вас есть какие-либо вопросы по поводу запроса, пожалуйста, свяжитесь с нами по указанному ниже адресу:

Полное наименование организации, адрес, контактный номер телефона.

С уважением,

Должность

Ф.И.О.

## Annex 4: Top-3 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 in Tajikistan 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 it 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>

In cases in which PDFs consist of scanned images, extracting the data might be more difficult for 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](#), which allows the conversion of scanned images into text (up to 8 Mb). If Word and [Img2TXT](#) do not work, you may purchase ABBYY FineReader, 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 data in machine-readable files.