

Эффективное управление для экономического развития  
Компонент 2: Гражданское общество за подотчетность



# Отчет

## Картирование открытых данных в Кыргызстане

Февраль – Март 2022 года

## **Благодарность**

Данный отчет подготовлен по заказу Интерньюс и составлен Zerkalo-Analytics в рамках проекта «Эффективное управление для экономического развития», Компонент 2: Гражданское общество за подотчетность». Данный проект реализуется при финансовой поддержке Правительства Великобритании в сотрудничестве с ACTED.

В создании данного отчета участвовало много людей. Автором отчетов по исследованию является Йон Марандич (д.ф.н). База данных собрана командой в лице: Кахрамона Бакозода (директор), Берика Кожомкулова (координатор), Йона Марандич, Чолпон Бейшекеевой, Сайфулло Садыкова, Рухшоны Джалоловой, Искандара Ражабова, Фарруха Садыкова и других сотрудников из организации «Зеркало». Автор отчета хотел бы поблагодарить следующих рецензентов из Интерньюс за предоставление полезных комментариев в отношении предварительных отчетов данного исследования: Саида Маниева, Изабель Шлэфер, Айзада Токтогулова, Асель Омуралиева, Рэйчел Кэйвуд, а также Зарина Эргашева.

Данный материал подготовлен при финансовой поддержке Правительства Великобритании. Мнения и взгляды, содержащиеся в материале, не обязательно отражают официальную точку зрения Правительства Великобритании.

## Содержание

<b>Резюме</b>	<b>4</b>
<b>1. Введение</b>	<b>6</b>
<b>2. Состояние экономической и дата журналистики в Кыргызстане</b>	<b>9</b>
<b>3. Картирование открытых данных в Кыргызстане</b>	<b>11</b>
<b>4. Основные проблемы для дата журналистов в Кыргызстане</b>	<b>16</b>
<b>5. Рекомендации</b>	<b>17</b>
Приложение 1: Инструкция для журналистов по использованию базы данных	20
Приложение 2: Контрольный список для журналистов, работающих с данными	23
Приложение 3: Шаблон письма о запросе данных	26
Приложение 4: Три лучших инструмента визуализации данных для журналистов	27
Приложение 5: Рекомендации по конвертации PDF	28

## Резюме

**Общие положения:** Новые источники данных, связанные с социально-экономическим развитием, должны улучшить журналистику. Однако из-за отсутствия централизованного хранилища открытых данных в Кыргызстане журналисты могут сталкиваться с задержкой при получении доступа к важным базам данных.

**Цель:** Для того чтобы решить данную проблему, организация «Зеркало» провела исследование по выявлению источников открытых данных в Кыргызстане и формированию всеобъемлющей базы данных доступной на «Airtable» ([ссылка](#)) и Google Таблицы ([ссылка](#)).

**Методы:** Опираясь на метод двухэтапного кабинетного исследования организация «Зеркало» составила базу данных, содержащую большую часть источников открытых данных в Кыргызстане. Период сбора данных для исследования составил с 14 февраля по 10 марта 2022 года. В целом исследовательская группа проверила 1986 гиперссылок на 131 веб-сайтах, содержащих соответствующие пакеты данных. В результате организация «Зеркало» выявила 389 наборов данных на социально-экономические темы и включила их в базу данных.

### Результаты:

- В ходе отображения данных выявлено, что большинство баз данных находятся в ведомстве следующих организаций: Национальный статистический комитет Кыргызской Республики (140), Министерство экономики и коммерции Кыргызской Республики (21), Центральный банк (21), Министерство финансов (17), Антимонопольное агентство (11), Всемирный банк (10) и т.д.
- Другой вывод связан с форматом данных: Веб (237), Excel (48), PDF (61) и Word (10).
- Отображение данных показывает, что существует больше данных в отношении основных тем, таких как Сельское хозяйство (75), Занятость (51), Прогноз экономического развития (38), Банки и финансы (23), Цены (20), Бюджет (19), Бизнес (12), Региональное развитие (11) и т.д.
- С точки зрения издателя, данные поступают из государственных учреждений (290), международных организаций (54), международных НПО (8), местных НПО (2) и научных учреждений (2).

- Что касается языка, данные доступны на русском (302), английском (52) и кыргызском (2) языках.
- Для более удобной визуализации, доступ к базе данных можно получить на «Airtable» (нажмите [здесь](#)). Инструкция, как ориентироваться в базе данных на «Airtable» приведена в Приложении 1.

**Заключение:** В целом, наш отчет показывает, что количество открытых данных доступных сообществу дата журналистов в Кыргызстане растет. Все больше и больше наборов данных публикуется в Интернете, в основном благодаря многочисленным проектам, включающим партнерство между государственными учреждениями и международными поставщиками помощи в целях развития. Основываясь на результатах проведенного анализа, в отчете сформулированы несколько рекомендаций для различных заинтересованных сторон.

## **Рекомендации**

### Для СМИ:

- Журналисты могли бы наилучшим образом использовать базу данных, изучая ее содержание на «Airtable» ([ссылка](#))
- Информационным организациям следует предложить целевое обучение для потенциальных журналистов, специализирующихся на данных, с акцентом на то, как начать исследование темы, связанной с данными, как проводить базовый анализ и изучение данных, а также как делать репортажи, основанные на данных.
- Ключевым навыком для журналистов, пишущих о данных, является визуализация. В Приложении 4 указан список рекомендуемых инструментов визуализации данных, отобранных организацией «Зеркало», а также ссылки на учебные пособия по их использованию.
- При работе с данными журналистам следует придерживаться определенных правил, чтобы обеспечить качество и надежность данных (Приложение 2).

### Для правительства:

- Правительству необходимо продолжать реализацию программ открытых данных, предлагая данные бесплатно, онлайн, в форматах, удобных для предварительной обработки.

- Поскольку большинство открытых данных охватывают определенные темы и исходят лишь от нескольких учреждений, одна из рекомендаций заключается в том, чтобы побудить чиновников из других ведомств публиковать данные по конкретным секторам.
- Правительству следует больше инвестировать в свой потенциал для последовательного сбора и публикации данных. Необходимо повысить осведомленность общественности о роли и важности открытых данных для социально-экономического развития, предоставляя такие данные в легком и безопасном доступном машиночитаемом формате.

#### Для донорских организаций:

- Финансирующим организациям следует ввести публикацию данных в проекты развития, а также финансировать проекты, направленные на укрепление потенциала организаций гражданского общества по производству данных.
- Финансирующие организации могли бы предложить возможности обучения для должностных лиц, отвечающих за сбор и публикацию данных, чтобы открытые данные предоставлялись в форматах, пригодных для анализа.

## 1. Введение

Жители всего мира получают информацию из средств массовой информации (радио, телевидение, газеты), от создателей онлайн-контента и из социальных сетей. Журналисты выполняют ключевую функцию в современном обществе, поскольку они информируют граждан о новых политических инициативах и часто следят за тем, как эта политика будет реализовываться, связывая политиков с общественностью. Этот механизм хорошо работает в развитых демократических странах с сильными и плюралистическими структурами СМИ. Более того, существующие исследования подтверждают, что сильные медиа-структуры оказывают положительное влияние на такие экономические явления, как бедность, неравенство имущественного положения и коррупция.<sup>1</sup> Напротив, развивающиеся страны с только формирующимися рынками СМИ чаще страдают от недостатка надежной и объективной информации о различных направлениях государственной политики.<sup>2</sup>

Дата журналисты играют особую роль в таких информационных экосистемах. Чтобы привлечь к ответственности государственных чиновников и контролировать реализацию

<sup>1</sup> Сангхамитра Бандьопадхьяй, Экономическое развитие, основанное на знаниях: Средства массовой информации и невесомая экономика. DARP (74), Международные центры экономики и смежных дисциплин Suntory и Toyota, Великобритания: Лондон.

<sup>2</sup> Кристофер Старк, Тереза К. Нааб и Хельмут Шерер. «Свобода разоблачать коррупцию: Влияние свободы СМИ, доступа к Интернету и предоставления государственных онлайн-услуг на коррупцию». Международный журнал коммуникации [Онлайн], 10 (2016): 21.

государственной политики, журналисты должны информировать общественность. В мире, где данных становится все больше, журналисты, работающие с большими массивами данных, могут выполнять эту миссию, публикуя основанные на данных репортажи о новейшей социально-экономической политике в конкретном контексте. В отличие от своих коллег, освещающих социально-экономические вопросы в традиционной манере, журналистов, работающих с данными, часто игнорируют организации, занимающиеся вопросами развития. Такое пренебрежение, вероятно, связано с относительной редкостью дата журналистики как отдельной специализации. В основе данного исследования лежит предположение о том, что основной подход должен быть направлен на расширение возможностей журналистов по привлечению государственных чиновников к ответственности путем предоставления информации, основанной на фактах.

Для становления сообщества журналистов, работающих с данными, решающее значение имеет доступность данных. Надежные и открытые данные позволяют журналистам предлагать высококачественные отчеты. Чтобы избежать дальнейшей путаницы, мы определяем открытые данные как «данные, которые можно свободно использовать, применять повторно и распространять – при соблюдении лишь, в крайнем случае, требования об атрибуции и совместном использовании».<sup>3</sup> Другой аспект касается машиночитаемости. В исследовании Всемирного банка за 2017 год упоминается, что «для того, чтобы открытые данные имели наибольшую ценность, машиночитаемость должна стать ключевой целью для поставщиков данных» в развивающихся странах.<sup>4</sup> В то время как в развитых странах существует множество источников данных, которые журналисты могут использовать для написания своих материалов, в развивающихся странах наличие и доступность данных являются основными проблемами.

Выделяются четыре проблемы:

- Во-первых, открытых правительственных данных мало, так как развивающиеся государства обычно не имеют возможности производить высококачественные данные. Под высококачественными данными мы понимаем разрозненные данные, временные ряды, данные микроуровня и детализированные данные с информацией о геолокации. Временной ряд – это набор данных, включающий точки данных, собранных за длительные периоды времени. В качестве примера можно привести показатели ВВП за последние десятилетия или Исследования жизненных ценностей населения. Микроуровень относится к данным на уровне единиц измерения, таким как информация переписи населения (т.е. отдельные лица, домохозяйства), а детализация относится к способу классификации данных.

---

<sup>3</sup> См. *Справочник открытых данных*, Фонд открытых знаний, <https://opendatahandbook.org/guide/en/what-is-open-data/>

<sup>4</sup> Одри Аррисс, *Машиночитаемые открытые данные: Как их применять для развивающихся стран*, запись в блоге, 2 марта 2017 г. <https://blogs.worldbank.org/opendata/machine-readable-open-data-how-it-s-applicable-developing-countries>

- Во-вторых, такая нехватка данных означает, что журналисты имеют возможность консультироваться, практиковаться и лучше использовать данные, чтобы рассказать убедительную историю.
- В-третьих, журналисты в развивающихся странах редко обладают навыками и умеют использовать новейшие инструменты для работы с данными, поскольку приобретение таких технических знаний может занять много времени. В конце концов, чтобы научиться правильно анализировать данные, требуется обучение, которое может быть дорогостоящим.
- В-четвертых, сама журналистика данных является трудоемким процессом. Часто информационные организации не могут позволить себе репортажи, основанные на данных, поскольку у редакторов нет ресурсов для соответствующего вознаграждения данной работы. Соответственно, даже у подготовленных журналистов, специализирующихся на данных, может отсутствовать мотивация для работы над репортажем о данных и применять полученные знания.

В других случаях репортажи, основанные на данных, которые приближаются к журналистским расследованиям, могут представлять личную опасность для журналистов и их семей, что отталкивает их от этого пути. Хотя потребности журналистов, работающих с данными, в каждой стране могут быть разными. Всемирный опрос журналистов, работающих с данными, показал, что доступ к качественным данным (56%), нехватка времени (49%), отсутствие достаточных знаний в области анализа данных (44%) и обеспечение надежности данных (39%) являются одними из главных трудностей, с которыми сталкиваются журналисты в этой области.<sup>5</sup>

Успешная дата журналистика публикует репортажи, представляющие общественный интерес. В этом смысле работа дата журналистов находится на пересечении нескольких различных областей. Они находят данные и создают базы данных, делают репортажи, основывающиеся на данных, используют передовые инструменты визуализации данных и применяют инновационные технологии для обработки данных.<sup>6</sup> В этом смысле исследование «Картирование открытых данных» вносит вклад в первую область, создавая обширную базу данных, которая облегчает доступ к надежным открытым данным из различных источников.

Настоящий документ содержит подробную информацию о результатах исследования. Структура итогового отчета выглядит следующим образом. В Разделе 2 будет рассмотрено состояние дата журналистики в Кыргызстане, выделяя существующие правовые и институциональные рамки.

<sup>5</sup> См. Андреа Абеллан, Исследование состояния журналистики данных, 2021 год., <https://datajournalism.com/read/blog/data-journalism-survey-2021>

<sup>6</sup> Анджел Арресе (2022) «В начале были данные»: Экономическая журналистика как/и журналистика данных, *Исследования журналистики*, ИЦО: [10.1080/1461670X.2022.2032803](https://doi.org/10.1080/1461670X.2022.2032803)

Раздел 3 описывает методологию исследования и основные результаты. Раздел 4 указывает на потенциальные проблемы для дата журналистов. Основываясь на наших результатах, последний раздел включает несколько рекомендаций для информационных организаций, правительства и доноров.

## 2. Состояние экономической и дата журналистики в Кыргызстане

За последнее десятилетие Кыргызстан добился значительного прогресса в отношении доступности открытых данных. Кыргызское государство выиграло от присутствия различных университетов и международных организаций, которые ввели в учебные программы курсы по изучению данных, вызвав большой интерес к этой теме. Например, Американский университет Центральной Азии и некоторые государственные университеты предлагают курсы по дата журналистике. Для журналистов в Бишкеке регулярно проводятся семинары в области данных.<sup>7</sup> Это привело к публикации многочисленных репортажей, основанных на данных, которые демонстрируют, что в Кыргызстане растет сообщество любителей сбора данных.<sup>8</sup> Более того, Школа данных Кыргызстана, местная НПО, продвигающая использование данных и доступ к ним, является частью глобальной сети, занимающейся продвижением грамотности в области данных среди организаций гражданского общества. Другой заметной инициативой страны является попытка собрать бесплатную библиотеку наборов данных для каждой из пяти республик Центральной Азии.<sup>9</sup>

Наша оценка показала, что кыргызское государство в партнерстве с международными организациями развило потенциал для эффективной организации и представления данных в Интернете. Например, веб-сайты Национального статистического комитета Кыргызской Республики (ссылка), Министерства экономики и торговли (ссылка), а также Национального банка (ссылка) хорошо организованы и удобны для пользователей. Такие улучшения являются результатом ряда реформ, которые начались в 2010-х годах с попытки опубликовать информацию о контрактах на государственные закупки, что сделало оборот бюджетных финансовых потоков более прозрачным.<sup>10</sup>

---

<sup>7</sup> См. Хакатон по журналистике данных «Взломать бедность», 12 февраля 2018 г. <http://project533785.tilda.ws/>. О других проектах открытых данных в Кыргызстане см. <http://opendata.kg/page19106310.html>

<sup>8</sup> См. например, следующий отчет об ухудшении состояния начальных школ в Кыргызстане: Валентина Галич, Диана Светличная, Таалайбек Талантбек улуу, Екатерина Афанасьева, Сезим Дигербаева, Жузумкан Асхатбекова «Трещина в стене: Почему дети в Кыргызстане учатся в экстремальных условиях», Радио Азаттык. <https://rus.azattyk.org/a/schools/31439009.html> ; Также смотрите богатый данными онлайн-ресурс Tazabek <https://www.tazabek.kg/> ; Список публикаций, основанных на данных, см. здесь: <http://opendata.kg/page19725884.html>; На сайте Kloop.kg есть раздел, посвященный репортажам, основанным на данных: <https://kloop.kg/blog/category/data-zhurnalistika/>

<sup>9</sup> См. Школа организации данных в Кыргызстане и база данных <http://opendata.kg/page19981037.html>

<sup>10</sup> См. Жылдыз Касымова и Марко Аурелио Маркес Феррейра и Сюзанна Дж. Пиотровски, 2016. «Способствуют ли инициативы в области открытых данных повышению и поддержанию прозрачности? Сравнительный анализ порталов открытого бюджета в

Другие инициативы, такие как Национальная стратегия по внедрению электронного управления, были запущены в 2014 году при значительной финансовой поддержке ПРООН.<sup>11</sup> В 2017 году, Кыргызстан добился дальнейшего прогресса, присоединившись к партнерству «Открытое правительство». В настоящее время правительство продолжает работу по реализации проектов цифрового управления, таких как концепция «Цифровой Кыргызстан 2019-2023», система межведомственного электронного взаимодействия «Тундук», которая позволит обмениваться большим количеством открытых данных. Еще один проект касается модернизации системы сбора и распределения статистических данных.<sup>12</sup> Это должно привести к тому, что в течение следующих трех лет журналистам будет доступно больше открытых данных. Кроме того, Всемирный банк реализует первый компонент программы «Эффективное управление для экономического развития» (EGED), финансируемой Правительством Великобритании, с целью поддержки разработки и реализации экономической политики, основанной на фактических данных. Это будет включать в себя оказание помощи правительствам в их усилиях по созданию более качественных данных, использованию данных для разработки и реализации политики, улучшению координации внутри правительства по реализации политики и лучшему взаимодействию с гражданами. Всемирный банк сосредоточится на совершенствовании политики регионального развития путем отображения данных микроуровня в рамках сел (т.е. айыл аймак) и городов, доступных в виде интерактивных карт на платформе Национального статистического комитета.

Тем не менее, даже для кыргызских журналистов доступные открытые данные нелегко найти в Интернете. Доступ к публичной информации в Кыргызстане регулируется Законом о доступе к информации (Закон № 213 от 2006 года с последними изменениями от 2017 года), который обеспечивает правовую основу и правила, которым необходимо следовать при запросе информации от органов власти.<sup>13</sup> Закон № 213 также распространяется на данные, которые кыргызские журналисты могут запрашивать у различных ведомств на национальном, региональном и местном уровнях. Закон накладывает временные ограничения на государственных служащих, устанавливая, что информация, представляющая общественный интерес, должна быть предоставлена в течение двух недель со дня запроса. Большинство журналистов работают в сжатые сроки и вынуждены действовать под давлением. Поэтому им сложно запрашивать и быстро находить нужные данные. Им приходится просматривать множество таблиц и показателей на разных сайтах, пока они не найдут нужную

---

развивающихся странах», Государственное управление и информационные технологии, в: Jing Zhang & Luis Felipe Luna-Reyes & Theresa A. Pardo & Djoko S. Sayogo (ed.), Информация, модели и устойчивость, стр. 137-155, Springer.

<sup>11</sup> См. Открытые общественные данные и их влияние на развитие Кыргызстана, веб-сайт ПРООН, ноябрь 2014

г. <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> «Запущен проект по модернизации Налогового управления и статистической системы», веб-сайт Национального статистического комитета, 9 марта 2021 года.

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

<sup>13</sup> Министерство юстиции, Закон № 213, 2006 г. <http://minjust.gov.kg/ru/content/672>

информацию. В связи с этим объединенная база данных, созданная командой Z-Analytics, предоставляет прямой доступ к экономическим данным с многочисленных сайтов, а также контактную информацию для запроса дополнительных данных. Журналисты, ученые и заинтересованные граждане, пишущие об экономическом развитии Кыргызстана, могут проверить информацию об основных экономических показателях, сравнив ее с аналогичными данными из других источников.

### **3. Картирование открытых данных в Кыргызстане**

В рамках проекта «Картирование открытых данных в Кыргызстане» была создана база данных, содержащая основные источники открытых данных. Выполненный в феврале-марте 2022 года международной и многокультурной командой из Кыргызстана, Таджикистана и Канады, проект направлен на сбор данных в области экономического развития.

Для создания базы данных Z-Analytics использовала двухэтапный подход кабинетного исследования. Сначала мы определили все источники, содержащие релевантные экономические данные о Кыргызстане. Мы начали с оценки данных, опубликованных государственными органами. Операторы Z-Analytics проиндексировали и просмотрели 49 правительственных сайтов. Затем Z-Analytics изучила данные международных организаций, местных и региональных НПО, выявив 389 наборов данных по состоянию на 18 марта 2022 года. После того как источник данных был найден, операторы следовали стандартному протоколу, записывая его характеристики. Последняя партия наборов данных была добавлена с помощью поиска по ключевым словам в основных поисковых системах (Google, Yandex) и целевого поиска в избранных хранилищах данных.

В тех случаях, когда государственные данные не были открыты, операторы переходили ко второму этапу, запрашивая доступ. Шаблон для таких запросов данных приведен в Приложении 3. В некоторых случаях ассистенты исследователей связывались с чиновниками по телефону. На этом этапе они отмечали, получили ли они ответ и соблюдали ли власти действующее законодательство о доступе к информации.

Основным результатом проекта является создание базы данных с бесструктурными файлами с данными, организованными в строки и столбцы, как показано на Изображении 1. Структура базы данных идентична электронной таблице Google или Excel с дополнительным уровнем сложности, состоящим из ряда выпадающих списков, что отличает ее от существующих баз данных. Доступная на английском и русском языках, она содержит информацию о наборах данных, структурированную по 15 основным категориям, каждая из которых соответствует одному столбцу. Чтобы сделать ее более удобной для пользователей, новая база данных была размещена на платформе организации данных «Airtable». Ее можно легко просматривать, сортировать, загружать и обновлять.



Изображение 2. Наборы данных по секторам.

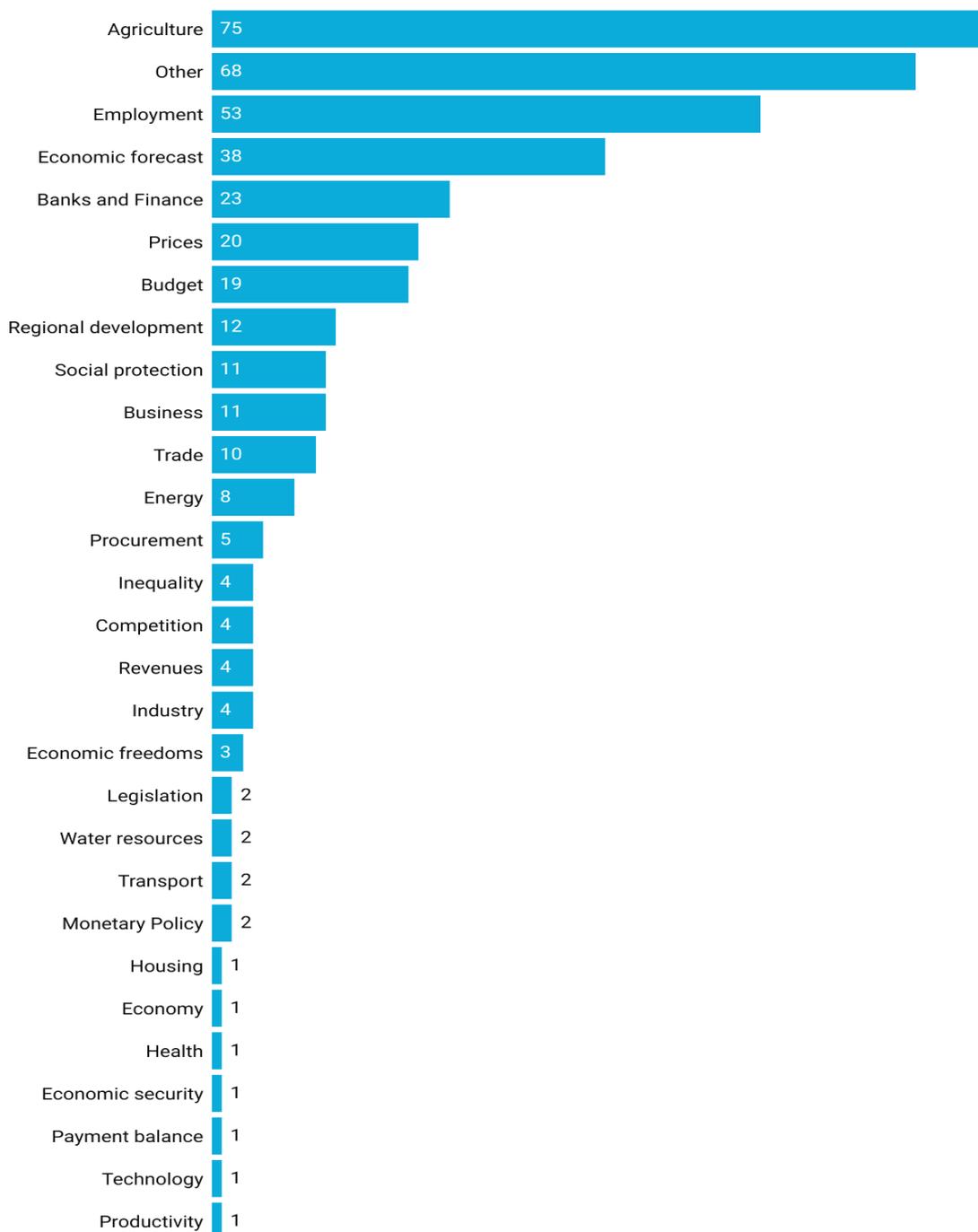


Chart: Ion Marandici • Source: Zerkalo Center • Created with Datawrapper

*/Сельское хозяйство; Прочее; Занятость; Экономический прогноз; Банки и финансы; Цены; Бюджет; Региональное развитие; Социальная защита; Бизнес; Торговля; Энергетика; Закупки; Неравенство; Конкуренция; Доходы; Промышленность; Экономические свободы; Законодательство; Водные ресурсы; Транспорт; Денежно-кредитная политика; Жилищные вопросы; Экономика; Здоровье; Экономическая безопасность; Платежный баланс; Технологии; Производительность/*

Большинство выявленных наборов данных встроены в веб-страницу, некоторые доступны в виде PDF-файлов (63), и только 51 можно загрузить в формате Excel (Изображение 3). Поскольку

документы PDF не поддаются расширенному анализу данных, в Приложении 5 мы даем рекомендации журналистам по извлечению данных из таких файлов.

*Изображение 3. Формат наборов данных.*

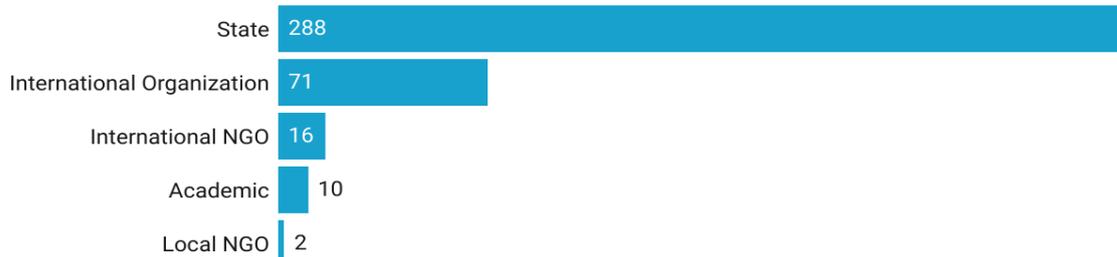


Chart: Ion Marandici • Source: Zerkalo Center • Created with Datawrapper

*/ Веб-страница; PDF; Excel; Word/*

Еще один вывод касается происхождения наборов данных. Как видно из Изображения 4, большинство наборов данных получено от государственных органов, и лишь несколько - от международных НПО, международных организаций, академических и местных НПО. Это говорит о том, что журналисты данных могут столкнуться с трудностями при подтверждении официальных данных при помощи данных из альтернативных источников.

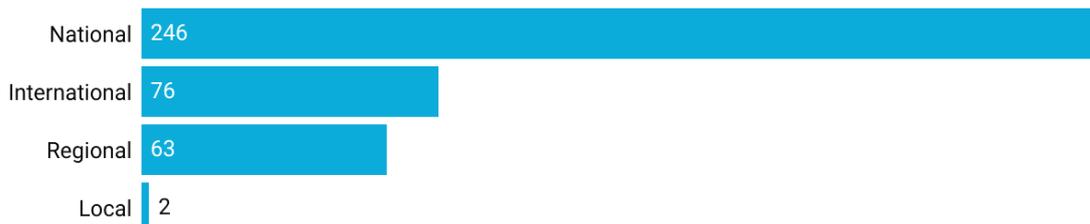
*Изображение 4. Тип издателя набора данных.*



*/Государство; Международная организация; Международная НПО; Академическое учреждение; Местная НПО/*

На Изображении 5 показано, что большинство наборов данных отображают национальный уровень (246), а некоторые – региональный (63) и международный (76) уровни. База данных содержит только два местных набора данных.

*Изображение 5. Уровень набора данных.*



*/Национальный; Международный; Региональный; Местный/*

В базе данных есть специальный столбец, в котором указывается следующая ожидаемая публикация. Изображение 6 демонстрирует некоторое представление о частоте обновления каталогизированных наборов данных. Большинство из них обновляются ежегодно, в то время как некоторые из них являются результатом разовых проектов и не имеют информации о следующей публикации.

Изображение 6. Частота публикации данных.

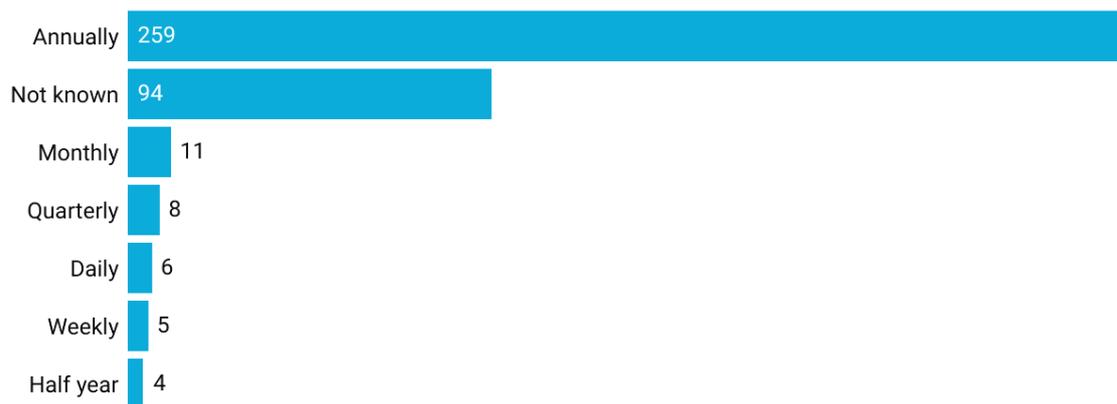


Chart: Ion Marandici • Source: Zerkalo Center • Created with Datawrapper

*/Ежегодно; Неизвестно; Ежемесячно; Ежеквартально; Ежедневно; Еженедельно; Раз в полгода/*

В рамках работы по отображению данных, исследователи из организации «Зеркало» связались с государственными чиновниками, запросив дополнительную информацию в тех случаях, когда наборы данных были устаревшими. Реакция на такие запросы данных была разной. В то время как Министерство финансов опубликовало на своем веб-сайте запрошенные данные, некоторые учреждения удалили данные со своих веб-сайтов или перенаправили запросы на получение данных другим чиновникам. Например, в своем официальном ответе заместитель министра юстиции заявил, что учреждение не уполномочено хранить и предоставлять данные о судебных делах, перенаправив запрос в Верховный суд Кыргызстана. В случае с Министерством по чрезвычайным ситуациям, данные были предоставлены в виде PDF-файла. Другой вопрос, касающийся доступа к данным, связан с подготовленностью чиновников, ответственных за выдачу данных. В нескольких случаях государственные служащие просто не знали, как ответить на запрос. Исследователи организации «Зеркало» связывались с ними по телефону, чтобы описать объем запроса данных и объяснить, как он должен быть выполнен. Из положительного стоит отметить, что ни одно из учреждений, к которым обращались за дополнительными данными, не пыталось продать данные или потребовать неофициальную оплату. Кроме того, в большинстве случаев учреждения своевременно отвечали на

запросы данных. Более того, многие ключевые данные доступны для свободного скачивания в виде брошюр и буклетов на веб-сайте Национального статистического комитета Кыргызстана.<sup>14</sup>

Открытые данные часто удаляются с сайтов, исчезая из сети. Иногда это является результатом технических ограничений различных дизайнов сайтов. В других случаях данные удаляются из общего доступа, потому что журналистские расследования раскрывают какие-то неудобные аспекты деятельности государственного чиновника. В целях решения данной проблемы, организация «Зеркало» заархивировала все наборы данных. Некоторые данные были сохранены на Google Drive, в других случаях копии были сохранены с помощью цифрового архива платформы всемирной паутины (см. Archive.org). Сохранение этих наборов данных позволяет журналистам получать доступ к копиям наборов данных на своих устройствах, а также к различным общедоступным данным, заархивированным в Интернете, в случаях, когда URL-адреса нарушены или данные удаляются из сети.

#### 4. Основные проблемы для дата журналистов в Кыргызстане

Несмотря на обилие источников данных о различных аспектах социально-экономического развития Кыргызстана, есть некоторые области, которые представляют собой серьезные проблемы.

Одна из областей, которую можно улучшить, касается публикации данных. Большинство наборов данных в нашей базе данных обновляются ежегодно, что означает, что данные могут устареть к тому времени, когда журналисты смогут их использовать. В таких случаях журналистам необходимо запросить обновленную информацию у властей, используя шаблон, представленный в Приложении 3.

Еще одна проблема связана с официальной реакцией на журналистскую работу. Несмотря на растущее сообщество дата журналистов в Кыргызстане, основная проблема связана не столько с грамотностью в области данных, сколько с тем, как власти реагируют на расследования, основанные на данных. Актуальным примером является дело Болота Темирова, который был арестован в 2022 году после публикации отчета о расследовании коррупции.<sup>15</sup> Аналогичным образом, две информационные организации сообщили о хакерских атаках на своих сотрудников, когда закон о борьбе с фейковыми новостями от 2020 года ввел новые ограничения на свободу слова в Кыргызстане.<sup>16</sup> Аналогично, «Артикль 19» – международная правозащитная организация, осуществляющая мониторинг свободы

---

<sup>14</sup> Список публикаций, выпущенных Статистическим комитетом Кыргызстана см. по следующей ссылке: <http://www.stat.kg/ru/publications/>

<sup>15</sup> Даврон Хотам, Кыргызские СМИ протестуют против ареста журналиста, занимающегося расследованиями, VOA, 25 января 2022 г. <https://www.voanews.com/a/kyrgyz-media-protest-arrest-of-investigative-journalist/6412175.html>

<sup>16</sup> Парламент Кыргызстана одобрил законопроект о «ложной информации», Комитет по защите журналистов, 10 августа 2021 г. <https://cpj.org/2021/08/kyrgyzstan-parliament-approves-false-information-bill/>; В Кыргызстане усиливается давление правительства на СМИ, Центр по исследованию коррупции и организованной преступности, 21 февраля 2021 г. <https://www.occrp.org/en/daily/16000-government-s-pressure-against-media-mounts-in-kyrgyzstan>

слова – отметила многочисленные инциденты, когда доступ к информации не соблюдался, а журналисты подвергались преследованиям.<sup>17</sup> Отчет, в котором использовались данные за 2015-2021 годы, показал, что лишь некоторые из лиц, нападавших на журналистов, были привлечены к ответственности.<sup>18</sup>

Кроме того, существует проблема доступа только к официальным данным. Журналистам часто приходится проверять официальные заявления, опираясь на альтернативные источники данных. Однако такие данные часто просто отсутствуют. Это похоже на хорошо задокументированную проблему, называемую *журналистикой пресс-релизов*, когда информационные организации склонны полагаться на официальные пресс-релизы, а не проводить собственные исследования по определенной теме.<sup>19</sup>

Схожей проблемой является отсутствие правовых последствий для тех чиновников, которые отказываются предоставить данные. Часто чиновники отказываются предоставлять данные под предлогом того, что они составляют государственную тайну, или просто не дают никаких объяснений своему решению.<sup>20</sup>

Также существует проблема получения доступа к машиночитаемым табличным данным как наиболее предпочтительному формату для работы с данными. Некоторые наборы данных предоставляются в виде документов в формате pdf, и требуется время для конвертации таких данных в таблицы Excel или csv.

## 5. Рекомендации

### Рекомендации для средств массовой информации

- Проект «Картирование открытых данных» собрал обширную коллекцию наборов данных. Журналистам рекомендуется оптимально использовать ее, используя «Airtable» ([ссылка](#)).
- Проект принесет дополнительную пользу, если информационные организации смогут проводить обучающие семинары для журналистов по трем дополнительным направлениям:
  - как начать исследование по теме с большим количеством данных
  - как проводить базовый анализ данных и исследование данных

---

<sup>17</sup> Кыргызстан: Нападения на журналистов и доступ к информации после спорных выборов, Артиклъ 19, 8 октября 2020 г.

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

<sup>18</sup> Саадат Тологонова, Журналистов в Кыргызстане постоянно избивают, преследуют и судят – Анализ данных, Kloop.kg, 19 февраля 2022 г. <https://kloop.kg/blog/2022/02/19/zhurnalistov-v-kyrgyzstane-postoyanno-izbivayut-presleduyut-i-sudyat-data-analiz-kloopa/>

<sup>19</sup> Элира Турдубаева, Медийное пространство в Кыргызстане: Между захватом элиты и контролем политических и деловых интересов, Центр внешней политики, 1 марта 2021 г. <https://fpc.org.uk/media-landscape-in-kyrgyzstan-caught-between-elite-capture-and-control-of-political-and-business-interests/>

<sup>20</sup> «Кыргызские журналисты ограничены в доступе к информации», Аналитический центр по Центральной Азии, 9 августа 2019 г. <https://acca.media/en/kyrgyz-journalists-restricted-access-to-information/>

- как составлять репортажи, основанные на больших данных
- Важным навыком, связанным с данными, является визуализация данных. Организация «Зеркало» провела обзор существующих инструментов визуализации данных для журналистов и включила в Приложение 4 список из трех инструментов визуализации данных, которые могут быть использованы кыргызскими журналистами, заинтересованными в подготовке репортажей, основанных на больших данных.
- Приложение 2 включает практические рекомендации для журналистов в отношении того, как избежать распространенных трудностей при работе с данными.
- Рекомендуется, чтобы журналисты скачивали наборы данных и делали скриншоты исходной веб-страницы, так как во многих случаях данные были удалены государственными чиновниками. Журналисты также могут восстановить сохраненные страницы из интернет-архива ([ссылка](#)).
- Журналистам рекомендуется проверять сайт Национального статистического комитета, поскольку в следующем году он опубликует микроуровневые данные на уровне села (т.е. айыл аймак) и города.

#### **Рекомендации для правительства:**

- Правительство должно и дальше внедрять принципы открытых правительственных данных, предоставляя полные наборы данных своевременно и доступно. Это включает в себя публикацию данных в форматах, позволяющих более глубокую обработку.
- Правительство должно предлагать данные, не являющиеся собственностью, исключая любые сборы или другие ограничения, такие как проверка личности.
- Правительство должно предоставлять данные онлайн в стабильном формате на постоянном веб-сайте. Открытые данные должны быть безопасными для открытия и сопровождаться надлежащей документацией (например, кодировочными книгами).
- Правительству следует оценить влияние на распространение открытых данных текущих изменений, связанных с реформами Санарип Аймак, проводимыми Статистическим комитетом Кыргызстана. В частности, было бы полезно узнать, будут ли эти изменения иметь различный эффект для разных типов наборов данных в зависимости от метода их сбора.

#### **Рекомендации для донорских организаций:**

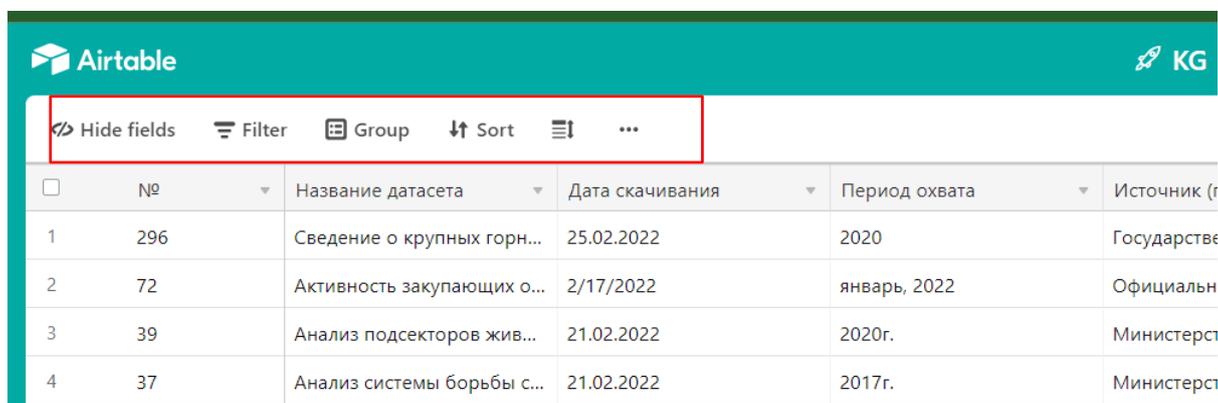
- Донорские организации должны раскрыть потенциал открытых данных и репрезентативного управления посредством финансирования проектов, увеличивающих объем данных, размещаемых в открытом доступе.

- Донорские организации должны продолжать содействовать повышению уровня грамотности в области данных среди кыргызских журналистов, предоставляя возможности для обмена опытом между журналистами данных из разных стран, финансируя семинары по данным и поощряя отличную журналистскую работу в области данных.
- Донорские организации могут предложить возможности обучения для чиновников, отвечающих за сбор и публикацию данных на различных уровнях. Цель является снижение затрат на публикацию данных и устранение опасений, что публикация данных может привести к нежелательному эффекту.

## Приложение 1: Инструкция для журналистов по использованию базы данных

«Airtable» обладает простым в использовании и понятным интерфейсом. Получив доступ к базе данных на «Airtable» ([ссылка](#)), выполните несколько простых шагов, чтобы получить нужные данные. Основные инструменты для использования базы данных находятся в левом верхнем углу и включают следующие функции:

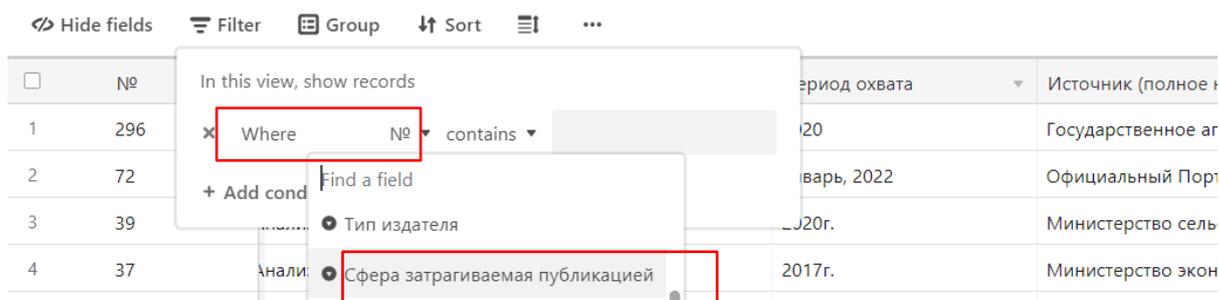
1. Скрыть поля
2. Фильтровать
3. Группировать
4. Сортировать
5. Высота строки



№	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

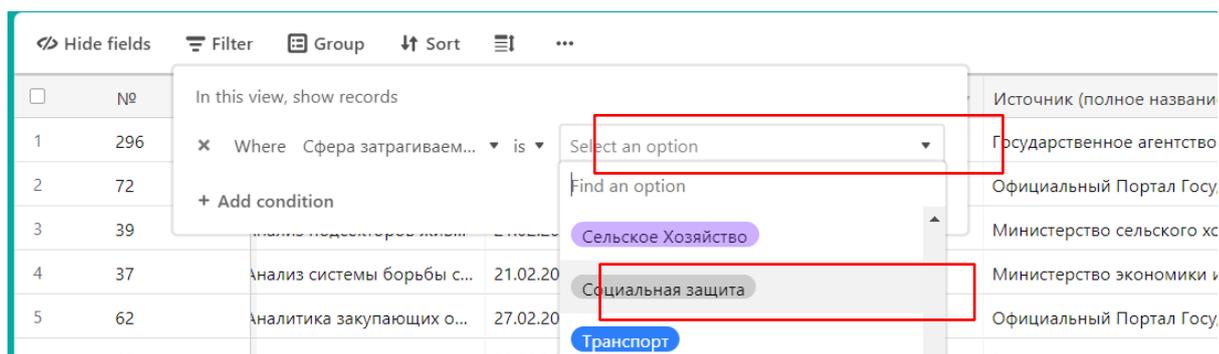
**Скрыть поля** – эта функция позволяет скрыть выбранные столбцы. Например, когда у Вас слишком много столбцов и они не помещаются на экране, функция «Скрыть поля» позволяет убрать с экрана ненужные столбцы.

**Фильтровать** – это ключевая функция для работы с данными. Пользователь может настроить параметры и найти интересующие его данные. Пример: Предположим, Вы хотите увидеть все данные, связанные с социальной защитой. Для этого нажмите на «Добавить условие», а затем в настройке «Где» выберите «Область публикации».



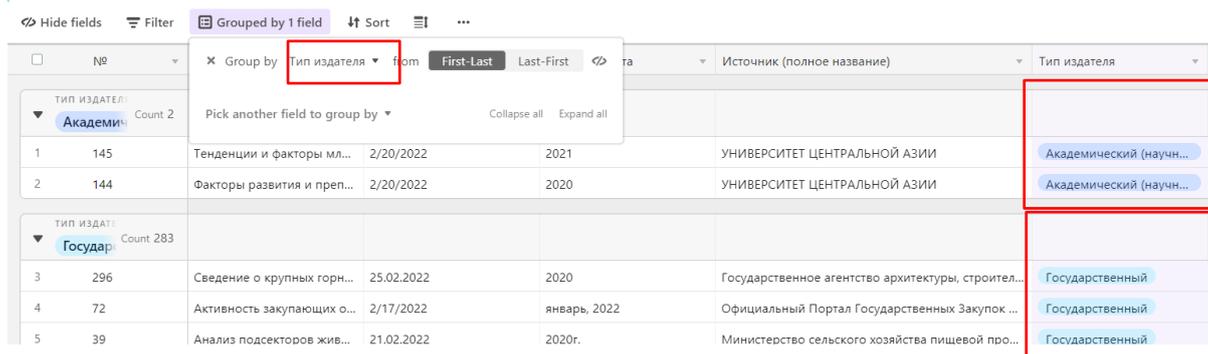
№	Период охвата	Источник (полное н
1	20	Государственное аг
2	январь, 2022	Официальный Порт
3	2020г.	Министерство сель
4	2017г.	Министерство экон

Далее в меню «выбрать опцию» выберите нужную сферу, например «Социальная защита».

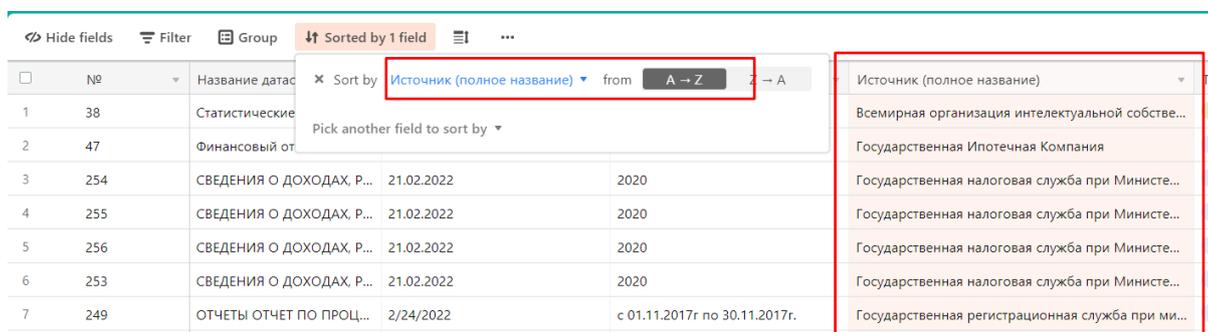


Теперь в таблице будут отображаться только наборы данных, относящиеся к области социальной защиты. Чтобы удалить примененный фильтр, нажмите **x** слева от настройки «Где».

**Группировать** – функция «Группировать» позволяет пользователю упорядочить данные на основании содержимого столбца. На скриншоте ниже данные классифицированы по типу издателя с помощью колонки «Тип издателя». Аналогично, наборы данных могут быть сгруппированы по любому нужному столбцу.



**Сортировать** – эта опция позволяет классифицировать данные на основании содержимого конкретного столбца. В данном случае сортировка осуществляется в алфавитном порядке.

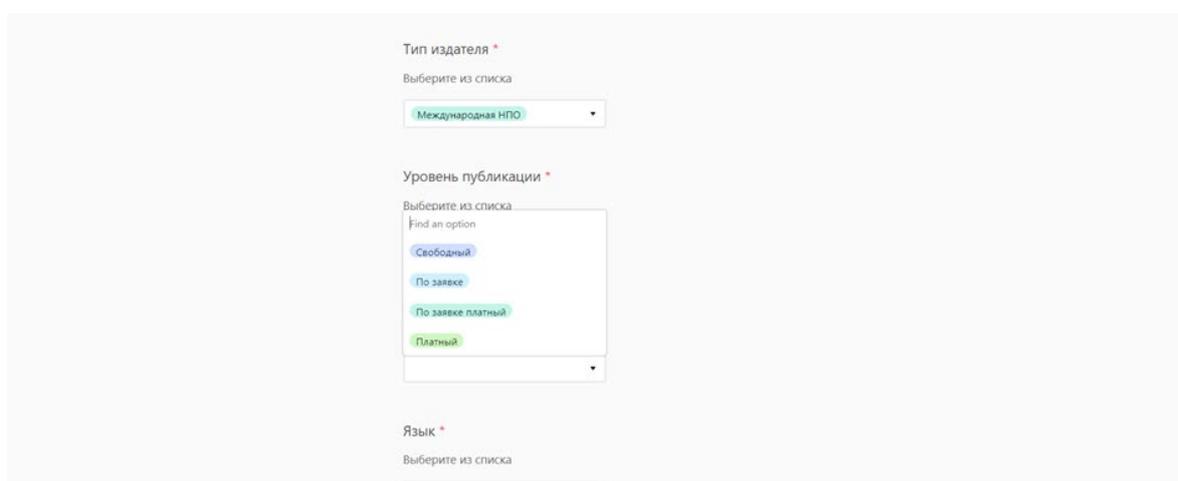


**Высота строки** – эту опцию можно использовать для изменения высоты каждой строки. Можно выбрать минимальную высоту строки, чтобы данные отображались компактно. Настройка максимальной высоты строки повышает удобочитаемость, так как обеспечивает большее пространство между текстом в каждой строке.

**Столбец «Теги»** – эта опция позволяет пользователям группировать наборы данных по столбцу «Теги». Чтобы настроить способ отображения информации в базе данных, пользователи должны использовать функцию группировки, описанную выше, и выбрать «Столбец «Теги» в качестве столбца группировки, чтобы найти данные по схожим темам.

### **Форма ввода данных**

Пользователи могут внести свой вклад в базу данных, заполнив форму ввода данных. Форма ввода данных состоит из тех же классификаций полей, что и база данных. Она включает три типа полей: текст, выпадающие списки и несколько вариантов ответа. Поля, отмеченные звездочкой (\*), являются обязательными для заполнения.



The image shows a portion of a web form with three required fields, each marked with a red asterisk (\*). The first field is 'Тип издателя \*' (Publisher type \*), with a dropdown menu showing 'Международная НПО' (International NGO) selected. The second field is 'Уровень публикации \*' (Publication level \*), with a searchable dropdown menu showing 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 that is partially visible at the bottom of the screenshot.

После заполнения всех обязательных полей пользователь должен нажать кнопку «Отправить», чтобы введенные вручную данные были автоматически перенесены в основную базу данных.

## **Приложение 2: Контрольный список для журналистов, работающих с данными**

Чтобы избежать предвзятости, журналисты должны придерживаться ряда основных правил при работе с данными. Один из способов добиться этого - проявлять критическое мышление и задавать следующие вопросы:

**Кто собрал данные?** Заслуживают ли доверия организации, собирающие данные? Сбор данных - это сложный и дорогостоящий процесс. Выяснив, кто собрал данные, журналисты могут обрести больше уверенности в том, что эти данные являются надежными. Организация, известная своими исследованиями данных, является положительным знаком, в то время как неизвестная организация, представляющая данные, требует дополнительной проверки.

**Как были собраны данные?** Получили ли люди, собирающие данные, специальную подготовку? Сбор данных часто требует навыков, которые можно приобрести с помощью организованного образования и обучения. Это не означает, что любители сбора данных не могут выдавать и публиковать данные, но это повышает вероятность того, что данные, которые Вы собираетесь использовать, являются надежными.

**Использовал ли производитель данных эффективную методологию?** Доступна ли методология в открытом доступе, чтобы журналисты могли изучить ее? Чтобы проверить это, журналист может попросить владельца данных предъявить методологию, использованную для сбора данных. Как правило, методологии исследований публикуются вместе с данными. Аналогичным образом, поставщик данных должен предоставить кодировочную книгу.

**Кто распространяет данные? Когда и почему?** Может случиться так, что производитель данных не является распространителем данных. Например, частная компания может заказать опрос о своей продукции. Социологическая компания, проводящая опрос, предоставляет данные, но частная компания решает опубликовать только те данные, которые рекламируют ее продукцию. В таких случаях журналистам рекомендуется поинтересоваться, кто финансировал опрос и почему его результаты публикуются в определенное время. Существует множество доказательств того, что в некоторых странах опросы общественного мнения публикуются перед выборами, чтобы повлиять на их результаты. Именно поэтому некоторые страны ввели ограничения на публикацию результатов опросов общественного мнения в день выборов.

**Как представляются данные и с какой целью?** Способ представления данных имеет большое значение. Часто данные представляются таким образом, чтобы поддержать определенную политическую или экономическую повестку дня, а неудобные данные упускаются. В некоторых случаях презентации данных создаются для того, чтобы стать вирусными.

**Как были сформулированы вопросы?** В опросах формулировка вопросов оказывает большое влияние на получаемый ответ. Понимание респондентом термина или вопроса может не совпадать со значением, придаваемым ему исследователем. Целесообразно спросить, ввел ли интервьюер респондентов в заблуждение, выбрав определенные слова.

**Насколько надежны данные?** Для опросов: достаточно ли велик размер выборки? Журналисты могут воспользоваться одним из доступных онлайн-калькуляторов выборки. С этим справится калькулятор, предлагаемые «SurveyMonkey» ([ссылка](#)).

**Какова допустимая погрешность в случае использования данных опроса?** Всякий раз, когда отбирается выборка, предел погрешности показывает, насколько поставщик данных уверен в том, что выборка представляет все население. Одно из главных правил заключается в том, что предел погрешности составляет 3 процента.

**Существуют ли альтернативные наборы данных, которые могут подтвердить официальные данные?** По возможности журналисты должны сравнивать данные из одного источника с данными, полученными из другого источника.

**Насколько показательна выборка?** Этот вопрос не относится к выборке, представляющей население. Вместо этого, это просьба к журналисту и средствам массовой информации обратить внимание на маргинализованные слои населения и неблагополучные группы населения. Были ли исключены из выборки определенные лица, например, люди с ограниченными возможностями, дети, женщины, нуждающиеся люди, ЛГБТ и т.д.?

**Остерегайтесь ошибочных логических выводов!** Эта ошибка происходит, когда логические выводы об отдельных людях в группе основываются на выводах о группе в целом.

При интерпретации данных важно помнить, что **средние показатели часто бесполезны**, поскольку они не рассказывают всего. Например, в странах с высоким уровнем экономического неравенства высокий средний показатель ВВП на душу населения не обязательно означает, что все население является преуспевающим.

### Приложение 3: Шаблон письма о запросе данных

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 csvc ... 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.

## Приложение 4: Три лучших инструмента визуализации данных для журналистов

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

**Описание:** Datawrapper позволяет пользователям без опыта кодирования создавать визуально привлекательные графики, карты и таблицы с помощью нескольких щелчков мыши. Он может помочь журналистам данных в Таджикистане исполнить роль хранителей общественных интересов. Для ознакомления с Datawrapper, пожалуйста, нажмите [здесь](#). Если Вас интересуют более сложные темы, мы предлагаем воспользоваться ресурсами Академии Datawrapper ([ссылка](#)).

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

**Описание:** RAW Graphs - это фреймворк визуализации открытых данных, созданный с целью сделать визуальное представление сложных данных простым для всех. Он не требует опыта кодирования. RAW Graphs обеспечивает недостающее звено между приложениями электронных таблиц (например, Microsoft Excel) и графическими редакторами (например, Adobe Illustrator, Inkscape, Sketch). Для получения краткого руководства посмотрите эту [ссылку](#).

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

**Описание:** Flourish запущен в 2018 году, основан на R и не требует навыков кодирования. Компания сделала инструмент бесплатным для журналистов. Примеры как использовать Flourish можно посмотреть [здесь](#).

## Приложение 5: Рекомендации по конвертации PDF

Организация «Зеркало» представляет следующие рекомендации для дата журналистов для работы с наборами данных в формате Portable Document Format (PDF). Если PDF был создан путем экспорта из приложения Windows, то решение следующее:

**Вариант 1:** Загрузите PDF-файл. Откройте документ с помощью Adobe Acrobat. В меню Adobe Acrobat DC пользователь может выбрать опцию «Сохранить как», а затем выбрать Word или Excel в качестве желаемого формата конечного документа. Данное решение является рекомендуемым, так как качество конвертации является высоким.

**Вариант 2:** Если у Вас не установлен Acrobat reader, выполните следующие действия. Загрузите набор данных в формате PDF на рабочий стол. Щелкните правой кнопкой мыши на значке документа и откройте его с помощью Microsoft Word. Word конвертирует и откроет PDF как документ Word.

**Вариант 3:** Используйте надежные и безопасные веб-сайты для онлайн-конвертации. В этом случае пользователь может загрузить PDF-файл на платформу и скачать конвертированный документ.

Предлагаемые платформы:

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

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

В случаях, когда PDF-файлы состоят из отсканированных изображений, извлечение данных может оказаться более сложным для журналистов. Такие PDF-файлы можно легко распознать, поскольку пользователи не могут выделить текст в документе. Они могут выбрать одно из следующих решений.

**Вариант 1:** Используйте ПО оптического распознавания символов для конвертации PDF-файла в другой формат. В таких случаях в конвертированном файле могут быть ошибки. Поэтому журналисту необходимо проверить точность конвертации. Word имеет встроенный OCR, который может конвертировать отсканированный PDF-файл в редактируемый текст. Другой вариант - бесплатный онлайн конвертер оптического распознавания символов - [Img2TXT](#), который позволяет конвертировать отсканированные изображения в текст (до 8 Мб). Если Word и [Img2TXT](#) не работают, Вы можете приобрести ABBYY FineReader, который является одним из наиболее рекомендуемых инструментов OCR.

**Вариант 2:** Сделайте скриншоты отсканированных данных и воспользуйтесь одним из программных инструментов оптического распознавания, доступных онлайн, например ABBYY FineReader.

**Вариант 3:** Введите данные в машиночитаемые файлы вручную.