

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



Отчет
Картирование открытых
данных в Таджикистане

Февраль – Март 2022 года

Благодарность

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

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.

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

Содержание

Введение	7
Краткий обзор состояния дата журналистики в Таджикистане.....	11
Картирование открытых данных в Таджикистане	15
Основные проблемы для дата журналистов	23
Рекомендации для заинтересованных сторон.....	25
Приложение 1: Инструкция для журналистов по использованию базы данных ...	27
Приложение 2: Контрольный список для журналистов, работающих с данными .	30
Приложение 3: Письмо о запросе данных (Оригинал)	33
Приложение 4: Некоторые инструменты визуализации данных для журналистов	34
Приложение 5: Рекомендации по конвертации PDF	35

Резюме

Общие положения: Новые источники данных, связанные с социально-экономическим развитием, должны улучшить журналистику. Однако из-за отсутствия централизованного хранилища открытых данных в Таджикистане 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 Таджикистане и формированию всеобъемлющей базы данных доступной на «Airtable» ([ссылка](#)) и Google Таблицы ([ссылка](#)). Посредством картирования открытых данных, текущее исследование облегчает доступ к информации, связанной с развитием для журналистов, работающих над социально-экономическими вопросами.

Метод: Опираясь на метод двухэтапного кабинетного исследования организация «Зеркало» составила базу данных, содержащую большую часть источников открытых данных в Таджикистане. Период сбора данных для исследования протекал с 14 февраля по 10 марта 2022 года. В целом исследовательская группа проверила 1032 гиперссылки на 68 веб-сайтах, содержащих соответствующие пакеты данных. В результате организация «Зеркало» выявила 294 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 Банк Республики Таджикистан (48), Агентство по статистике при Президенте Таджикистана (46), Всемирный Банк (32), администрация города Душанбе (21), Университет Центральной Азии (15), Международный валютный фонд (14), Налоговая служба (12), а также Ассоциация микрофинансовых организаций Таджикистана (11).
- Другой вывод связан с форматом наборов данных: PDF (107), веб-страницы (86), Excel (64) и т.д. Данный результат позволяет предположить, что многие пакеты данных преобразованы в PDF файлы, которые не так хорошо подходят для анализа данных, как другие файлы.

- Отображение данных показывает, что существует больше данных в отношении основных тем, таких как Банки и финансы (52), Доходы (18), Торговля (18), Налоги (16) и Региональное развитие (12).
- С точки зрения издателя, данные поступают из государственных учреждений (171), международных организаций (52), местных НПО (29) и научных учреждений (16).
- Что касается языка, пакеты данных доступны на русском (159), английском (64) и таджикском языках (45).
- Инструкция, как ориентироваться в базе данных на «Airtable» (нажмите [здесь](#)) приведена в Приложении 1.

Заключение: В целом, наш отчет показывает, что существует потенциал для появления сильного сообщества дата журналистики в Таджикистане, предлагая несколько областей, которые можно улучшить.

Рекомендации

Для информационных организаций:

- Журналистам следует пользоваться возможностью исследования среды открытых данных в Таджикистане путем обращения к базе данных ([ссылка](#)).
- В целях содействия развитию дата журналистики, информационным организациям следует выделять достаточно ресурсов на репортеров, вовлеченных в анализ данных.
- Исследование предполагает, что журналисты могли бы извлечь пользу из семинаров по визуализации данных.
- Информационным организациям следует устранить технические препятствия, мешающие журналистам, работающим с данными, вводить визуализации данных в свои материалы (Приложение 4).
- Для обеспечения надежности данных журналисты, работающие с данными, могут следовать рекомендациям, приведенным в Приложении 2.
- С этой целью в исследовании приводится шаблон письма с просьбой о предоставлении данных (Приложение 3).

Для правительства:

- Поскольку большинство открытых данных связаны с определенными темами, одна из рекомендаций - побуждать чиновников из других ведомств публиковать данные.
- Правительство должно больше инвестировать в свой потенциал для последовательного сбора и публикации данных.
- Необходимо повышать осведомленность о роли и важности открытых данных для социально-экономического развития, предоставляя данные в легком, безопасном и доступном формате, пригодном для чтения электронными устройствами.
- Государственные чиновники, отвечающие за распространение данных в каждом ведомстве/министерстве, должны своевременно отвечать на журналистские запросы об обновлении данных.

Для финансирующих организаций:

- Организациям гражданского общества и международным донорам, поддерживающим развитие сообщества данных в Таджикистане, необходимо привлечь государственные учреждения, чтобы убедить их в важности и полезности предоставления данных в соответствующем формате.
- В то же время, сотрудников государственных учреждений можно было бы обучить практическим навыкам публикации данных в формате, пригодном для чтения электронными устройствами, а не в виде PDF-документов. Это облегчило бы работу дата журналистов.
- Для решения проблемы слабого журналистского потенциала можно предложить целевые курсы по грамотности в области данных для местных информационных организаций.
- Необходимо реализовать пилотный проект открытых данных для города Душанбе.

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 реализуется такая политика и как она влияет на жизнь людей, связывая политиков с общественностью и обеспечивая их подотчетность.

Этот механизм относительно хорошо работает в развитых демократических странах с сильными и плюралистическими структурами средств массовой информации. Более того, существующие исследования подтверждают, что сильные медиаструктуры оказывают положительное влияние на такие экономические явления, как бедность, неравенство распределения доходов и богатства и коррупция.¹ Напротив, развивающиеся страны с формирующимися рынками СМИ в большей степени страдают от недостатка надежной и объективной информации о деятельности правительства.

Журналисты, работающие с данными, играют особую роль в таких информационных экосистемах. В мире, где данных становится все больше, журналисты, работающие с крупными базами данных, часто могут лучше выполнять свою задачу по привлечению чиновников к ответственности и информированию общественности о реализации политики, предоставляя основанные на данных, заметки о новейшей политике, связанной с управлением, в конкретном контексте. Конечно, при неправильной подготовке или несоблюдении профессиональных стандартов и этики, журналисты и средства массовой информации могут способствовать созданию необъективной информационной среды, повышая уровень дезинформации и манипулирования общественным мнением. В этом случае данные часто могут использоваться для того, чтобы склонить общественность в пользу той политики, которую предпочитают власти. Напротив, при правильном использовании, данные могут предоставлять важные доказательства и бороться с попытками дезинформации и манипулирования.

В отличие от своих коллег, освещающих социально-экономические вопросы традиционным способом, журналисты, работающие с данными, часто упускаются из виду организациями развития, поддерживающими консолидацию гражданского общества. Такое пренебрежение, вероятно, связано с относительной редкостью дата журналистики как отдельной специализации. В основе данного исследования лежит предположение о том, что укрепление работы журналистов,

¹ 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, 2005 г.

работающих с данными, будет способствовать привлечению к ответственности должностных лиц и расширению прав и возможностей общественности посредством отчетности, основанной на фактах.

Для преобразования дата журналистов в сообщество решающее значение имеют совместная работа и доступные данные. Доступные и надежные открытые данные позволяют журналистам предлагать высококачественные отчеты по социально-экономическим вопросам. Для целей данного исследования мы определяем открытые данные как «данные, которые можно свободно использовать, использовать повторно и распространять кем угодно – при условии соблюдения требования об атрибуции и совместном использовании»². В частности, данное исследование сосредоточено на открытых данных, полученных от государственных учреждений, поскольку они располагают наибольшим объемом данных. Другой аспект касается машиночитаемости открытых данных. В исследовании Всемирного банка от 2017 года упоминается, что «для того, чтобы открытые данные имели наибольшую ценность, машиночитаемость должна стать ключевой целью для поставщиков данных».³ В контексте данного исследования машиночитаемые данные относятся к структурированным данным, которые могут быть экспортированы в виде табличных данных (файлы Excel или CSV) для дальнейшей обработки компьютерами с целью расчета различных статистических данных или изменения структуры пакета данных.

Выделяются четыре проблемы:

- Во-первых, данных не хватает, так как развивающиеся государства обычно не имеют возможности предоставлять 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 данные по ВВП за последние пятьдесят лет или данные Всемирного исследования ценностей. Микроуровень относится к данным агрегатного уровня, таким как материалы переписи (т.е. отдельные лица, домохозяйства), а детализация относится к способу классификации данных.

² См. *Справочник открытых данных*, Фонд открытых знаний, <https://opendatahandbook.org/guide/en/what-is-open-data/>

³ 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, такой дефицит данных означает, что журналисты не имеют возможности консультироваться, практиковаться и лучше использовать данные, чтобы сделать убедительный репортаж.
- В-третьих, журналисты в развивающихся странах редко обладают навыками и знаниями для использования новейших инструментов работы с данными, поскольку приобретение таких технических навыков может занимать много времени. В конце концов, чтобы научиться правильно анализировать данные, необходимо обучение, которое может быть дорогостоящим.
- В-четвертых, сама дата журналистика представляет собой трудоемкий процесс. Часто информационные организации не могут позволить себе репортажи о данных, поскольку у редакторов нет средств для соответствующей оплаты такой работы. Например, обычная оплата труда в редакциях составляет в среднем 150-200 сомони (\$11-\$15) за статью. Соответственно, даже у подготовленных журналистов, занимающихся данными, может отсутствовать мотивация работать над статьей о данных и применять полученные знания.

В других контекстах репортажи, основанные на данных, которые близки к журналистским расследованиям, могут представлять личную опасность для журналистов и их семей, что отталкивает их от этого пути. Всемирный опрос дата журналистов показал, что доступ к качественным данным (56%), нехватка времени (49%), отсутствие достаточных знаний в области анализа данных (44%) и обеспечение надежности данных (39%) входят в число основных сложностей, с которыми сталкиваются дата журналисты.⁴

Успешная дата журналистика создает репортажи, представляющие общественный интерес. В этом смысле работа дата журналистов находится на пересечении нескольких различных областей: поиск данных и создание баз данных, разработка репортажей, основанных на данных, использование передовых инструментов визуализации данных и применение новых инновационных технологий для обработки данных.⁵ 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.

⁴ См. Андреа Абеллан, Исследование состояния журналистики данных, 2021 год, <https://datajournalism.com/read/blog/data-journalism-survey-2021>

⁵ 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)

Настоящий документ содержит подробную информацию о результатах исследования. Структура итогового отчета выглядит следующим образом. В Разделе 2 будет рассмотрено состояние дата журналистики в Таджикистане, выделяя существующие правовые и институциональные рамки. Раздел 3 описывает методологию исследования и основные результаты. Раздел 4 указывает на потенциальные проблемы для дата журналистов, предлагая возможные решения. Основываясь на наших выводах, последний раздел включает в себя основные рекомендации.

Краткий обзор состояния дата журналистики в Таджикистане

В Таджикистане растет медиа-сообщество, в основном сосредоточенное в столице страны – Душанбе. По данным исследования предпочтений средств массовой информации, проведенного в 2019 году, в стране было зарегистрировано 376 газет (112 государственных и 264 частных), 245 журналов (114 государственных и 131 частных), 10 информационных агентств (1 государственное и 9 частных), 34 телеканала (8 - национальных и 26 региональных) и 33 радиостанции (4 общенационального охвата).⁶ Из всех зарегистрированных газет, государственные газеты являются самыми распространенными. Например, тираж единственной ежедневной газеты Таджикистана «Джумхурият» составляет 240 000 экземпляров. «Омузгор» (издание Министерства образования), «Минбари Халк» (партийная газета) и «Садои Мардум» (парламентский еженедельник) также широко распространены. Наиболее популярными частными газетами являются три таблоида, публикующие развлекательные новости, за которыми следуют издания «Таджикистан» и «Азия-плюс», предлагающие более аналитические материалы. Так же, как и в остальных странах мира, цифры тиражей подтверждают, что количество проданных экземпляров снижается с каждым годом.⁷ Кроме того, в Таджикистане растет сообщество создателей интернет-контента и онлайн-порталов. Тем не менее, несмотря на кажущееся изобилие источников информации, у изданий нет специальных отделов, использующих данные для создания контента, а если данные и используются, то обычно бегло, без понимания того, как они могут улучшить репортаж. Для поддержки журналистов, использующих данные в Таджикистане, Интерньюс провела несколько тренингов по данным и менторских сессий, а в 2021 году запустила программу по дате журналистике.

Неразвитость дата журналистики как отдельного направления в Таджикистане не означает, что журналисты не используют данные в своих репортажах на различные темы. Несколько примеров, отобранных после обзора нескольких таджикских СМИ, наглядно продемонстрируют данный тезис. Одним из основных онлайн-ресурсов является портал Asia Plus, который публикует разнообразный контент, включая новости экономики и бизнеса.⁸ В Asia Plus есть специальный раздел, посвященный экономике, в котором публикуются более длинные статьи. Например, расследование причин серьезного долга «Барки Точик», национальной

⁶ Медиа предпочтения населения Таджикистана: ТВ, радио, печатные СМИ, сайты, социальные сети и мессенджеры, Медиа-Консалтинг и Зеркало, 2019. http://zerkalo.tj/uploads/files/2020/MediaPreff-2019_Final_RUS.pdf

⁷ Ражаб Мирзо, Как продаются таджикские газеты?, Newreporter.org, 26 августа 2019 года. <https://newreporter.org/2019/08/26/kak-prodayutsya-tadzhikskie-gazety/>

⁸ Например, один медиа-проект состоял из интервью с успешными предпринимателями: <https://asiaplustj.info/projects/Probizness> .

энергетической компании Таджикистана, опирается на множество сравнительных данных о потреблении энергии для объяснения повторяющихся проблем с электроснабжением.⁹ В некоторых отчетах данные используются для того, чтобы рассказать увлекательные истории о влиянии глобального потепления на таяние ледников в Таджикистане, расхождениях в статистике торговли в Центральной Азии, задержках заработной платы, кредитовании, денежных переводах и бюджетных расходах. В другом примере, репортаж на сайте радио «Озоди» освящает экономический рост, не предлагая каких-либо визуальных презентаций.¹⁰ В этом случае, очевидно, сайт медиа компании не позволяет интегрировать определенные фреймы визуализации. Несмотря на то, что некоторые журналисты прошли обучение по работе с данными, краткий обзор медиаконтента показывает, что большинство журналистов используют данные традиционным способом, не применяя современные инструменты для работы с данными.

В других случаях журналисты используют инфографику и визуальные данные от международных экономических организаций, нежели создают собственные. К примеру, в расширенном отчете о бедности в Таджикистане, опубликованном недавно созданным новостным порталом Your.tj, включили инфографику, разработанную Всемирным банком.¹¹ Подобная практика свидетельствует о том, что не все редакции могут позволить себе нанять человека, который будет создавать инфографику. Хотя и возможно, что некоторые ключевые публикации были пропущены, данные примеры указывают на то, что дата журналистика все еще остается развивающейся сферой в Таджикистане. Следовательно, существует много возможностей для улучшения в плане доступности данных и грамотности.

Важно отметить, что существует несколько проектов, направленных на внедрение политики открытых данных, и дата журналистики. В 2015 году при поддержке Института «Открытое общество», Фонда открытых данных и Всемирного банка был опубликован отчет о готовности к использованию открытых данных. Среди основных рекомендаций были создание экспертной группы по открытым данным, которая разрабатывала бы национальную стратегию открытых данных, а также создание тесного сотрудничества между государственными учреждениями, гражданским обществом и сообществом разработчиков в связи с отменой практики отказа в доступе к данным со ссылкой на закон о государственной тайне. В документе также рекомендуется разработка пилотных проектов и программы по наращиванию потенциала

⁹ Пайрав Чоршанбиев, Три причины "бедности" Барки Точик, Asia-Plus, 1 марта 2022 г., <https://asiaplustj.info/ru/news/tajikistan/economic/20220301/tri-prichini-bednosti-barki-tochik>

¹⁰ Абдулло Ашуров, «7,3% рушди иктисоди Тоҷикистон. Аммо чаро муҳочирати қорӣ идома дорад?», Радио Озоди, 13 апреля 2019 г., <https://www.ozodi.org/a/29876593.html>

¹¹ Экономика Таджикистана растет, но бедность сохраняется, Your.tj, <https://your.tj/jekonomika-tadzhikistana-rastet-no-naselenie-vse-ravno-bedneet/>

для государственных служащих.¹² Аналогичным образом, проект «Хома», поддерживаемый Министерством иностранных дел и международного развития и Институтом открытого общества с 2016 по 2020 год, обучал журналистов финансовой грамотности, включая компонент дата журналистики. Главным результатом этого проекта стал специализированный веб-сайт, посвященный экономической журналистике и журналистике данных.¹³ Cabar.asia провела семинары по дата журналистике в рамках Школы молодых аналитиков с некоторыми инструментами онлайн обучения по дата журналистике доступными каждому.¹⁴

Что касается доступа к данным в Таджикистане, важно отметить, что примерно половина всех государственных учреждений имеют веб-форму для связи с общественностью для запроса неопубликованных данных. Это положительный результат, поскольку он свидетельствует об определенной степени открытости в отношении информационных запросов. Однако государственные учреждения различаются по времени предоставления ответа. Исследование Национальной ассоциации независимых СМИ показало, что сайты государственных учреждений не всегда содержат весь объем данных, часто они не располагают надлежащим форматом, не обновляются вовремя, и часто присутствует риск удаления данных. Аналогичное исследование показало, что многие государственные агентства неохотно отвечают на запросы с опозданием на месяцы, отвечая только на несколько вопросов из списка. Более того, иногда письменные запросы полностью теряются или остаются без ответа месяцами.¹⁵

Наше ключевое предположение данного исследования заключается в том, что благодаря установлению и облегчению доступа к данным из различных мест, текущее исследование по отображению данных способствует эффективному управлению. Это достигается путем консолидации доступа к источникам данных для оказания помощи местным медиа-компаниям с низким потенциалом анализа данных. Журналисты будут создавать более качественный материал, когда они смогут легко находить и анализировать необходимые данные. Они будут знать, как использовать данные в своих репортажах, когда ознакомятся с инструкциями, сопровождающими базу данных (см. ниже). Тем самым они помогут гражданам отслеживать эффективность государственной политики в Таджикистане. В то же время, поддерживая доступ журналистов к данным об экономическом развитии, инициатива по

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

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

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

¹⁵ См. Национальная ассоциация независимых СМИ, Отчет о доступе к информационной практике в Таджикистане, 20 июня 2019 года. <http://a2i.nansmit.tj/2019/06/20/> Точное время получения ответа см.: <http://a2i.nansmit.tj/%d0%bf%d0%be-%d0%b0%d0%b4%d1%80%d0%b5%d1%81%d0%b0%d1%82%d0%b0%d0%bc/>

отображению данных может также стимулировать представителей гражданского общества требовать большей прозрачности в отношении создания и распространения данных.

Картирование открытых данных в Таджикистане

Команда по «Картированию открытых данных в Таджикистане» создает базу данных, содержащую доступные открытые источники данных социально-экономической тематики. Исследование, проведенное в феврале-марте 2022 года международной и многокультурной командой, состоящей из специалистов из Кыргызстана, Таджикистана и Канады, было посвящено открытым данным в области экономического развития, публикуемым государством.

Субъекты частного сектора традиционно являются бенефициарами инициатив в области открытых данных, поскольку они могут использовать такие данные для выявления новых возможностей для бизнеса.¹⁶ Часто компании выступают в качестве посредников. Однако они могут создавать и собственные данные. Например, Twitter предоставляет платный доступ к своему архиву данных, а Uber и Airbnb опубликовали некоторые из своих данных в качестве открытых данных. Большинство частных компаний не решаются предоставлять свои данные открыто по причинам, связанным с бизнесом. К сожалению, такие данные частного сектора не выявлены в Таджикистане, а некоторые компании вместо этого продают свои данные. При проведении отображения приоритетными являлись открытые данные, которые могут быть использованы журналистами без каких-либо дополнительных затрат с их стороны. Поскольку такие закрытые базы данных, принадлежащие частным предприятиям, не могут быть использованы журналистами, они не подходят в качестве надежного источника информации для журналистов.

Для создания базы данных организация «Зеркало» использовала двухэтапный подход кабинетного исследования. Во-первых, мы провели кабинетное исследование, выявив все источники, содержащие соответствующие социально-экономические данные по Таджикистану. Мы начали с оценки всех данных, опубликованных государственными органами в Интернете. Операторы Z-Analytics просмотрели 1032 веб-страницы на 91 сайте государственных учреждений, международных организаций, местных и региональных НПО. В результате отображения было выявлено 294 пакета данных из 54 источников по состоянию на 18 марта 2022 года. Как только набор данных был найден, операторы следовали стандартному протоколу, индексируя и

¹⁶ См. пример Открытых данных 150, проект реализуемый совместно с Глобальной сетью Открытых данных 500, с указанием списка канадских компаний, полагающихся на открытые данные для развития своих предприятий: <https://canada.opendata500.com/list-page.html>

записывая его характеристики. Последняя партия наборов данных была добавлена с помощью функции поиска наборов данных Google и в отобранных хранилищах данных.¹⁷

В случаях, когда правительственные данные были устаревшими, операторы переходили ко второму этапу. По согласованию с Интерньюс были сделаны запросы данных за период 2019-2022 годов, чтобы обновить уже собранные наборы данных. В большинстве случаев такие данные запрашивались у учреждений, которые менее активны в плане публикации данных. Запросы не направлялись в те государственные учреждения, которые не публиковали данные. Шаблон для таких запросов данных указан в Приложении 3. В некоторых случаях ассистенты исследователей связывались с чиновниками по телефону или посещали отделы хранения данных в государственных учреждениях. На этом этапе они фиксировали, получили ли они ответ и соблюдают ли органы власти действующее законодательство о доступе к информации.

Основным результатом проекта является база данных с бесструктурными файлами, состоящая из простых строк и столбцов, как показано на Изображении 1 ([ссылка](#)). Структура базы данных идентична электронной таблице, но добавляется уровень сложности за счет включения ряда выпадающих списков. Доступная на английском и русском языках, она содержит информацию о наборах данных, структурированную по 15 основным категориям, каждая из которых соответствует столбцу. Чтобы сделать ее более удобной для пользователей, новая база данных была размещена на платформе организации данных «Airtable» ([ссылка](#)). Набор данных можно легко просматривать, сортировать и обновлять. Он доступен для бесплатной загрузки, распространения и использования.

¹⁷ Список хранилищ открытых данных можно найти здесь: <https://mira.mcmaster.ca/research/open-access-data-repositories>
Другие потенциальные источники включают <http://dataverse.harvard.edu>, data.europe.eu, <https://cyberleninka.ru> и аналогичные архивы.

Изображение 1. База данных с именами столбцов в формате Excel.

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV (I)	XV (II)	XVI (III)	XVII	XVIII	XIX	XX	XXI	XXII
1	Dataset Title	Access Date	Data period	Source (Full name)	Publisher	Type	Publication Area	Link	Utilization Level	Format	Access	Language	Issuance Cycle	Next Release	Main Action	Main Thesis	Main Thesis	Comment (email or pl)			
2	Statistics on ed	02/14/2022	2017-2018	Ministry of Educ	State	Other	https://maorif.tj	National	PDF	Free	Tajik	Annually	Not known	Education	Institutions	Students	(992 37) 221-561 P. 5-9	12-17, 35			
3	Indicators by nu	02/17/2022	The first 3 mont	Ministry of Cult	State	Budget	https://fahar.tj	National	PDF	Free	Tajik	Not known	Not known	Culture	Achievements	Offers	734023, Sharrif P. 16, 29, 39-40				
4	Socio-economic	02/18/2022	2014-2020	Indu: Statistics Agenc	State	Social protection	https://stat.wv	National	PDF	Free	Tajik	Quarterly	Not known	Food security, P. Prices and tariff	Industrial scena	Tajikistan, Dush. Collection	'Toc				
5	Environmental	02/18/2022	1991-2018	Statistics Agenc	State	Agriculture	https://stat.wv	National	PDF	Free	Tajik	Not known	Not known	Environmental (Rational use of investments, pr	Tajikistan, Dush. Environmental	Development o	Indicators and ir health financing	Tajikistan, Dush. Collection	'Hie		
6	Main indicators	02/18/2022	2014-2018	Statistics Agenc	State	Other	https://stat.wv	National	PDF	Free	Tajik	Not known	Not known	Development o	Indicators and ir health financing	Tajikistan, Dush. Collection	'Hie				
7	List of laws on ir	21.02.2022	2005-2012	World Trade org	International	Or Legislation	http://wto.tj/er	National	Web Page	Free	English	Not known	Not known	World Trade Org	Laws, Trade	Agreements	Ministry of Econ	Separate webs			
8	Indicators for in	21.02.2022	2011-2020	World intellect, international	Or	Ownership	https://www.wi	National	Web Page	Free	Russian	Not known	Not known	Economic indica	Technology	Patents	Contact center \	Sale of statistic			
9	Monthly inflati	21.02.2022	September 2016	National Bank o	State	Prices	https://nbt.tj/r	National	PDF	Free	Russian	Monthly	Not known	Banks and Finan	Inflation	Consumer price	Tel.: + (992) 44 6	Link to the pag			
10	Inflation Review	21.02.2022	2016-2021	National Bank o	State	Prices	https://nbt.tj/r	National	PDF	Free	Russian	Monthly	Not known	Banks and Finan	Inflation	Consumer price	Tel.: + (992) 44 6	#VALUE!			
11	Price indices	21.02.2022	1909-2019	National Bank o	State	Prices	https://nbt.tj/r	National	Excel	Free	Russian	Annually	Not known	Banks and Finan	Inflation	Dynamics	Tel.: + (992) 44 6	#VALUE!			
12	Monetary policy	21.02.2022	2022	National Bank o	State	Monetary Policy	https://nbt.tj/q	National	PDF	Free	Russian	Annually	Perhaps 2022-20	Banks and Finan	Forecasts	Inflation	Tel.: + (992) 44 6	#VALUE!			
13	Credit investme	21.02.2022	2000-2021	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	2022	Real sector	Loans	Statistics	Tel.: + (992) 44 6	#VALUE!			
14	Residues of savi	21.02.2022	2002-2021	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	2022	Real sector	Deposits	Statistics	Tel.: + (992) 44 6	#VALUE!			
15	National Bank N	21.02.2022	2001-2007	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	Not known	Monetary sectoi	Assets	Reserve money	Tel.: + (992) 44 6	#VALUE!			
16	National Bank N	21.02.2022	2008-2021	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	Not known	Monetary sectoi	Assets	Reserve money	Tel.: + (992) 44 6	#VALUE!			
17	Monetary Overv	02.22.2022	2001-2007	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	Not known	Monetary sectoi	Assets	Loans: Deposits	Tel.: + (992) 44 6	#VALUE!			
18	Monetary Overv	02.22.2022	2008-2020	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	Not known	Monetary sectoi	Assets	Loans: Deposits	Tel.: + (992) 44 6	#VALUE!			
19	Monetary Overv	02.22.2022	2001-2007	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	Not known	Assets, Liabiliti	Loans, Deposits	Holders, Capl. S	Tel.: + (992) 44 6	#VALUE!			
20	Monetary Overv	02.22.2022	2008-2020	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	Not known	Assets, Liabiliti	Loans, Deposits	Holders, Cash, S	Tel.: + (992) 44 6	#VALUE!			
21	Monetary Overv	02.22.2022	2008-2020	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Tajik	Annually	Not known	Wide money sat	Money	Deposits, Secur	Tel.: + (992) 44 6	#VALUE!			
22	Monetary Overv	02.22.2022	2008-2020	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	Not known	Wide money sat	Money	Deposits, Secur	Tel.: + (992) 44 6	#VALUE!			
23	Weighted avera	02.22.2022	2002-2021	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	Not known	Interest	Betting	Loan rates	Tel.: + (992) 44 6	#VALUE!			
24	Weighted avera	02.22.2022	2002-2021	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	Not known	Interest	Betting	Deposit rates	Tel.: + (992) 44 6	#VALUE!			
25	Information abc	02.22.2022	1995-2021	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	Not known	Interest	Betting	Medium bankin	Tel.: + (992) 44 6	#VALUE!			
26	Refinancing rat	02.22.2022	02/05/2021 to th	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	Not known	Refinancing rat	% per annum	Validity	Tel.: + (992) 44 6	#VALUE!			
27	Results of aucti	02.22.2022	2005-2020	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Excel	Free	Russian	Annually	Not known	Securities of the	State treasury b	Trade, Yield	Tel.: + (992) 44 6	#VALUE!			
28	Calendar of the	02.22.2022	February 2022	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Web Page	Free	Russian	Monthly	March 2022	Monetary sectoi	Securities	Calendar releas	Tel.: + (992) 44 6	#VALUE!			
29	Calendar of the	02.22.2022	2022	National Bank o	State	Banks and Finan	https://nbt.tj/fi	National	Web Page	Free	Russian	Annually	2023	Monetary sectoi	Securities	Calendar releas	Tel.: + (992) 44 6	#VALUE!			
30	Rate	02.22.2022	02.22.2022 (two)	National Bank o	State	Banks and Finan	https://nbt.tj/r	International	Web Page	Free	Russian	Daily	Not known	Foreign Econom	Currency quotat	#VALUE!	Tel.: + (992) 44 6	#VALUE!			
31	Information on	02.22.2022	On 08.28.2015	National Bank o	State	Banks and Finan	https://nbt.tj/r	National	Web Page	Free	Russian	Not known	Not known	Foreign Econom	Currency quotat	#VALUE!	Tel.: + (992) 44 6	#VALUE!			
32	Information on	02.22.2022	On 08.28.2015	National Bank o	State	Banks and Finan	https://nbt.tj/r	National	Web Page	Free	Russian	Not known	Not known	Foreign Econom	Currency quotat	#VALUE!	Tel.: + (992) 44 6	#VALUE!			

Журналисты могут получить доступ к базе данных двумя способами. Они могут получить доступ к ней в виде электронной таблицы Google ([ссылка](#)). Как видно на Изображении 1, электронная таблица включает два раздела: открытые данные и данные по запросу. В разделе открытых данных приведены все доступные данные, а в разделе данных по запросу содержатся данные, которые можно получить, связавшись с владельцами данных. Как в разделе открытых данных, так и в разделе данных по запросу содержится конкретная контактная информация на случай, если пользователям базы данных понадобится дополнительная информация или обновленные версии наборов данных. Журналисты могут использовать шаблон в Приложении 3 для запроса данных по запросу.

Второй вариант включает удобную для пользователя версию набора данных в «Airtable», практичном инструменте для организации данных, позволяющем легко находить данные. Это рекомендуемый способ получения доступа к базе данных, поскольку он обеспечивает удобную визуализацию данных. При этом журналисты могут группировать данные по любому предпочтительному столбцу, что позволяет им легко искать данные и получать доступ к ним. Инструкцию по доступу к данным в «Airtable» можно найти в Приложении 1.

Ниже приведены основные выводы, сделанные на основе анализа собранных данных. На Изображении 2 показано, что большинство наборов данных отображают национальный уровень, а некоторые - региональный и международный. База данных содержит только несколько местных наборов данных.

Изображение 2. Уровень набора данных.



Source: Zerkalo Center • Created with Datawrapper

/Национальный; Региональный; Международный; Местный/

Касательно типа издателя, мы видим, что большинство данных получено от государственных учреждений и международных организаций, в то время как некоторые данные получены от местных НПО и академических учреждений (Изображение 3).

Изображение 3. Наборы данных по издателю.

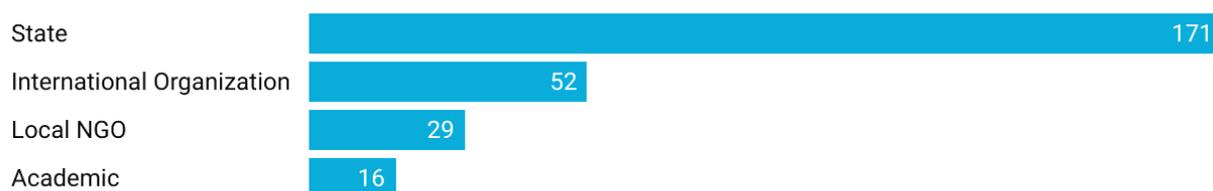


Chart: Ion Marandici • Source: Zerkalo Center • Created with Datawrapper

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

На Изображении 4 представлен формат данных. Он показывает, что в Таджикистане открытые данные обычно публикуются в виде PDF-документов и интегрируются в интернет-страницы. Это, конечно, влияет на то, как журналисты могут получить доступ к данным. Тот факт, что многие наборы данных встроены в PDF-документы, затрудняет доступ СМИ к данным и их извлечение для последующего анализа. В Приложении 5 мы приводим рекомендации по извлечению данных из PDF-документов.

Изображение 4. Формат наборов данных.



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

/PDF; Веб-страница; Excel; Word; По запросу; CSV/

Касательно языка, большинство наборов данных, включенных в базу данных, доступно на русском языке, за которым следуют английский и таджикский. Некоторые наборы данных могут быть доступны на нескольких языках.

Изображение 5. Наборы данных по языкам.

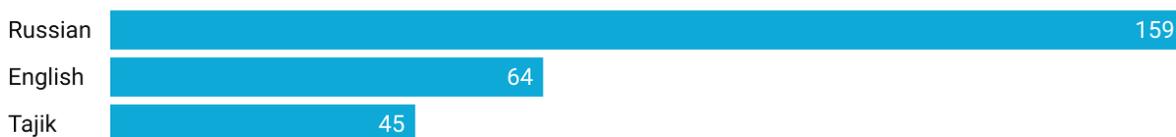


Chart: Ion Marandici • Source: Zerkalo Center • Created with Datawrapper

/Русский; Английский; Таджикский /

Касательно охвата, большинство открытых данных сосредоточено в следующих секторах: банки и финансы, торговля, налоги, доходы и региональное развитие (Изображение 6). Наборы данных по жилищным вопросам, правам собственности и производительности довольно редки. Одна из причин обилия открытых данных в банковском и финансовом секторах может быть связана с интенсивной активностью данных в государственных учреждениях в этой сфере, которые должны информировать общественность об обменных курсах и процентных ставках, обязательных резервах и формулировать прогнозы на основе различных макроэкономических показателей. Другими словами, они могут выдавать большие объемы данных и обнародовать часть из них. Другая причина может быть связана с конкретными требованиями к отчетности. Национальный банк Таджикистана (НБТ) связан определенными обязательствами перед своими кредиторами, такими как Международный валютный фонд, который может потребовать большей прозрачности, а также регулярной публикации данных перед выдачей льготных кредитов. НБТ

также участвует в торгах на международных биржах и поэтому должен соблюдать регулятивные нормы, чтобы получить разрешение на участие.

Изображение 6. Наборы данных по секторам экономики.

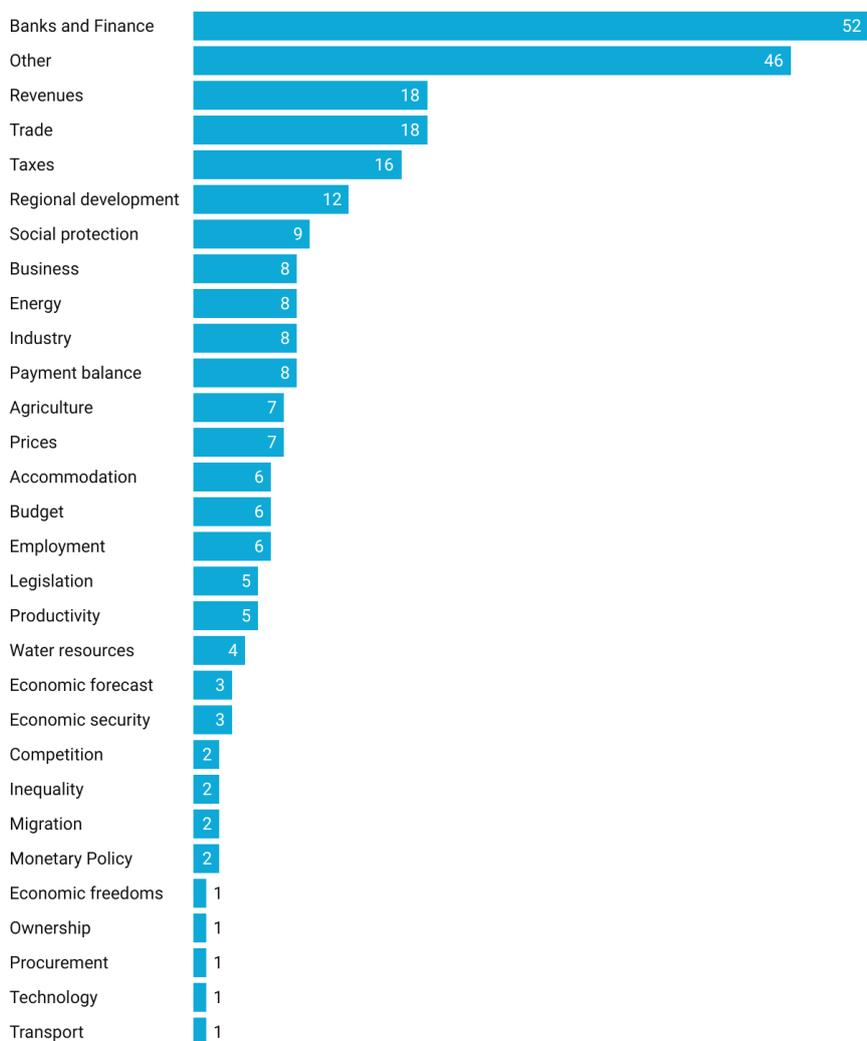


Chart: Ion Marandici • Source: Zerkalo Center • Created with Datawrapper

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

Существует значительный разброс в количестве открытых данных, найденных на правительственных сайтах. В Таблице 1 приведено количество наборов данных по внутренним ресурсам. Она показывает, что три учреждения - Статистический комитет Таджикистана, Национальный банк и Мэрия г. Душанбе – постоянно выдают больше открытых данных, чем

другие учреждения. Три вышеупомянутых учреждения опубликовали более 20 соответствующих наборов данных.

Таблица 1. Таджикские веб-сайты по количеству соответствующих наборов данных
Более 20 наборов данных
Агентство по статистике при Президенте Таджикистана
Национальный банк Таджикистана
Мэрия г. Душанбе
От 2 до 5 наборов данных
Агентство по занятости и трудоустройства (Министерство труда, миграции и занятости)
Аналитический центр «Наво»
Государственный комитет по землеустройству и геодезии
Государственный комитет по инвестициям и управлению государственным имуществом
Министерство финансов
Министерство энергетики и водных ресурсов
Национальный патентно-информационный центр
«Барки Точик», Национальная энергетическая компания
Республиканский союз потребительских обществ – «Таджикматлубот»
Таможенное управление
От 6 до 20 наборов данных
Ассоциация микрофинансовых организаций
Государственное предприятие «Душанбеводоканал»
Министерство экономического развития и торговли
Налоговая служба (Правительство Таджикистана)
Национальная Ассоциация малого и среднего бизнеса
1 набор данных
Генеральная прокуратура
Институт экономики и демографии Академии наук Республики Таджикистан
Комитет по делам женщин и семьи (Правительство Таджикистана)
Министерство культуры
Министерство образования и науки
Министерство транспорта
Торговый портал Таджикистана
Центр стратегических исследований (администрация президента)

Причины публикации открытых данных НБТ были рассмотрены выше, а в случае с мэрией г. Душанбе кажущуюся обширность данных можно объяснить другими факторами. Тщательное изучение наборов данных показало, что городская администрация не публикует высококачественные данные и не выпускает регулярные статистические отчеты. Вместо этого на одной веб-странице часто размещается несколько наборов данных, охватывающих различные темы регионального развития по районам. Кроме того, город обязан публиковать подробные прайс-листы на предоставляемые им услуги, которые также были включены в базу данных. Тем

не менее, у г. Душанбе нет портала открытых данных, как у некоторых крупных городов, но в будущем такая инициатива может оказаться полезной.^{17F}¹⁸

Общедоступные данные часто каталогизируются и продаются в виде брошюр и печатных томов. Агентство по статистике Таджикистана предлагает такие сборники данных.^{18F}¹⁹ Чтобы получить доступ к законодательству Таджикистана, журналисты данных могут сначала обратиться на веб-сайт Национального центра законодательства, где большинство правовых актов публикуется с задержкой в два-три месяца.^{19F}²⁰ Другой полезной базой данных, содержащей основные правовые акты, принятые государственными учреждениями Таджикистана, является «Адлия». Проект Министерства юстиции, USAID и Всемирного банка, в настоящее время «Адлия» предоставляет платный доступ ко всем видам законодательства быстрее, чем Национальный центра законодательства.^{20F}²¹

Открытые данные часто удаляются с сайтов, исчезая из общественного доступа. Иногда это является результатом технических ограничений различных дизайнов сайтов. В других случаях данные удаляются из Интернета, потому что журналистские расследования раскрывают какие-то неудобные аспекты деятельности государственного чиновника. Например, по данным таджикских журналистов, информация о ранее заключенных контрактах была удалена с портала открытых закупок и государственных тендеров, который давно не обновлялся.^{21F}²² Другие порталы с отсутствующими данными включают pozirof.tj, investcom.tj (Иностранные инвестиции) и moh.tj (Министерство здравоохранения). Часто веб-сайты обновляются, и предыдущие данные оказываются удаленными. Генеральный план г. Душанбе был удален с веб-сайта dushanbe.tj, принадлежащего главе исполнительной власти города, а данные о владельцах частных предприятий и других организаций исчезли с веб-сайта фискального реестра andoz.tj.

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 исчезают из сети.

¹⁸ См. для примера портал открытых данных города Торонто, запущенный в 2009 г.: <https://open.toronto.ca/>

¹⁹ Журналисты, заинтересованные в покупке таких данных, могут проверить каталог доступных публикаций здесь: https://stat.wv.tj/library/tj/2021_catalog.pdf

²⁰ Веб-сайт Национального центра законодательства <http://ncz.tj/legislation>

²¹ Тарифные планы варьируются от 100 до 400 долларов США. См. <http://www.adlia.tj/>

²² См. <https://zakupki.gov.tj/ref/zakupki/>. На новой версии веб-сайта указаны последние контракты на закупки: <https://eprocurement.gov.tj/tj/searchanno>

Для обновления некоторых наборов данных 18 марта организация «Зеркало» запросила дополнительные данные у нескольких государственных учреждений: Министерства образования, Агентства по патентам, Министерства энергетики, Комитета по инвестициям, Комитета по делам женщин, Налогового комитета, Исполнительного комитета г. Душанбе, Агентства по статистике (полный список см. [здесь](#)). Из десяти учреждений, с которыми были установлены контакты, ответили девять. Два набора данных в формате pdf были предоставлены Агентством по землеустройству, геодезии и картографии. Комитет по делам женщин перенаправил запрос в Национальное статистическое агентство, в то время как другие государственные органы не предоставили запрошенные данные, указав вместо этого, что все имеющиеся наборы данных опубликованы на официальных веб-сайтах. Представитель Государственного комитета по инвестициям пообещал предоставить данные, если они еще не опубликованы в Интернете, но не выполнил своего обещания. Эти результаты показывают, что получение последних данных от государственных учреждений в Таджикистане является сложной задачей для дата журналистов, поскольку чиновники неохотно идут на такие шаги, чтобы избежать возможных карьерных последствий.

База данных может обновляться постоянными пользователями, которые могут добавлять новые наборы данных. Мы призываем журналистов и средства массовой информации добавлять новые данные через следующую форму ([ссылка](#)). Таким образом, база данных с открытым доступом может стать самодостаточным проектом в том смысле, что она будет поддерживаться сообществом дата журналистов.

Основные проблемы для дата журналистов

Дата журналистика как отдельная форма журналистики развивается в Таджикистане медленно по нескольким причинам. Это явление объясняется несколькими факторами.

Во-первых, некоторым журналистам трудно использовать анализ данных в своих репортажах. Эта тенденция усугубляется трудностями, возникающими в процессе получения данных от государственных учреждений, ограниченностью ресурсов и относительной непопулярностью таких репортажей среди читателей. Кроме того, имеет значение и то, как журналистика как дисциплина преподается в университетах.

Вторая причина может быть связана с тем, как работают СМИ в стране. В частности, свобода СМИ в Таджикистане ухудшается. Эта тенденция, безусловно, имеет последствия для журналистов, расследующих чувствительные темы и запрашивающих данные у государственных

служащих. Доступ к информации по определенным темам может быть ограничен. Согласно Всемирному индексу свободы печати за 2021 год, составленному организацией «Репортеры без границ», Таджикистан занимает 162 место из 180 стран по уровню свободы СМИ.²³ Международные средства массовой информации также сообщают о том, как таджикские журналисты и создатели онлайн-контента подвергаются запугиванию в связи со своей работой, причем пандемия еще больше усугубляет такие тенденции, поскольку власти вводят новые ограничительные меры.²⁴

В-третьих, общая нехватка открытых данных является еще одной проблемой, с которой сталкиваются таджикские журналисты. В некоторой степени эта проблема носит круговой характер, поскольку маловероятно, что кто-то станет журналистом данных в условиях нехватки данных. Отсутствие данных журналистов также означает, что открытые данные запрашиваются не так часто, как это было бы возможно. Таким образом, таджикские чиновники, возможно, не привыкли к запросам данных. Но даже если бы спрос на открытые данные был велик, наше исследование показало, что данные, скорее всего, будут доступны в файлах PDF, которые потребуют дополнительной обработки для извлечения необходимой информации для более глубокого анализа. Для решения этой проблемы в Приложении 5 приведены некоторые практические рекомендации по конвертации таких PDF-файлов в машиночитаемые форматы.

В-четвертых, данные могут быть защищены. Некоторая информация, представляющая общественный интерес является закрытой согласно специальному закону Таджикистана как государственная тайна, и, следовательно, ее невозможно получить.²⁵ В других случаях закон не определяет определенные типы данных как секретные, но государственные учреждения отказываются их предоставлять и могут предложить их только журналистам, участвующим в пресс-конференциях.

В-пятых, когда государственные учреждения обновляют свои веб-сайты, часто данные могут быть удалены из общественного доступа. Предотвратить удаление можно только путем архивирования таких веб-сайтов.

²³Репортеры без границ, *Всемирный индекс свободы печати за 2021 год*, <https://rsf.org/en/tajikistan>

²⁴Международное партнерство по правам человека, *Цена за молчание против цены за высказывание: Свобода слова в СМИ в Таджикистане*, июль 2020 г. <https://www.iphronline.org/wp-content/uploads/2020/07/ENG-Media-Report-TJ.pdf>

²⁵Закон о государственных тайнах Таджикистана, 2014 (с поправками в 2015 и 2020 гг.). http://ncz.tj/system/files/Legislation/1095_ru.pdf

Рекомендации для заинтересованных сторон

Рекомендации для средств массовой информации

- Исследование выявило многочисленные наборы данных, связанных с социально-экономическим развитием Таджикистана. Для включения в репортажи рекомендуется, чтобы журналисты ознакомились с источниками данных, включенными в базу данных, и самостоятельно искали другие источники данных.
- Учитывая недостаток репортажей, основанных на больших данных, разумно предположить, что местным журналистам может не хватать компетенции для обработки данных. С этой целью можно организовать тренинги и семинары по анализу данных для повышения уровня информационной грамотности таджикского сообщества СМИ.
- Для стимулирования развития дата журналистики средства массовой информации могут выделять ресурсы для аналитиков данных.
- Журналисты данных могут быть ограничены техническими возможностями веб-сайтов своих изданий. Средства массовой информации должны устранить технические препятствия, мешающие журналистам данных встраивать различные виды визуализации в свои репортажи.
- 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 таджикскими журналистами, заинтересованными в подготовке репортажей, основанных на больших данных.
- Рекомендуется, чтобы журналисты скачивали наборы данных и делали скриншоты исходной веб-страницы, так как во многих случаях данные были удалены органами власти. Журналисты также могут восстановить сохраненные страницы из интернет-архива ([ссылка](#)).

Рекомендации для правительства

- Одна из рекомендаций касается секторов с нехваткой данных. Анализ нашей базы данных показывает, что существует множество открытых данных по банкам и финансам, но не так много по жилищным вопросам, производительности и занятости. Это означает, что органы власти, контролирующие определенную политику, с меньшей вероятностью будут выдавать открытые данные. Одним из решений этой проблемы могут стать шаги по улучшению институционального потенциала различных ведомств, чтобы можно было генерировать больше данных.
- Похожая рекомендация касается секторов, по которым данные существуют, но не публикуются. Необходимо провести дополнительные исследования, чтобы определить, почему определенные данные, представляющие общественный интерес, не доступны в Интернете.
- Еще одна идея для рассмотрения заключается в обучении сотрудников государственных учреждений обработке и публикации данных в форматах, пригодных для анализа. Для этого правительство могло бы вступить в партнерство с международными организациями, которые могли бы работать над изменением того, как государственные учреждения и научно-

педагогические кадры в Таджикистане публикуют данные, поощряя их переходить от PDF и изображений к машиночитаемым форматам.

- Опираясь на опыт других стран, правительство могло бы облегчить доступ к открытым данным посредством создания единого портала открытых данных.

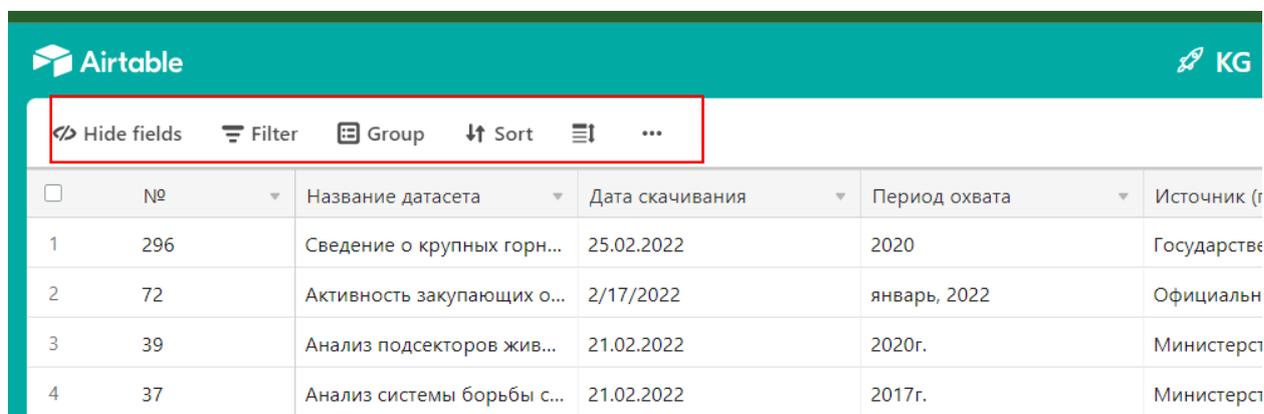
Рекомендации для донорских организаций

- Донорские организации могли бы предложить финансирование курсов по повышению грамотности в области данных, которые повысили бы компетентность таджикских журналистов в области данных. Такое обучение может помочь создать резерв журналистов, заинтересованных в более глубокой подготовке в области анализа данных.
- Донорские организации могли бы поощрять государственные учреждения к созданию большего количества машиночитаемых открытых данных.
- Организации, предоставляющие помощь в целях развития, могли бы запустить инициативу по открытым данным в городе Душанбе в виде пилотного проекта, при этом городские предприятия и организации гражданского общества получили бы значительные выгоды от такой инициативы.

Приложение 1: Инструкция для журналистов по использованию базы данных

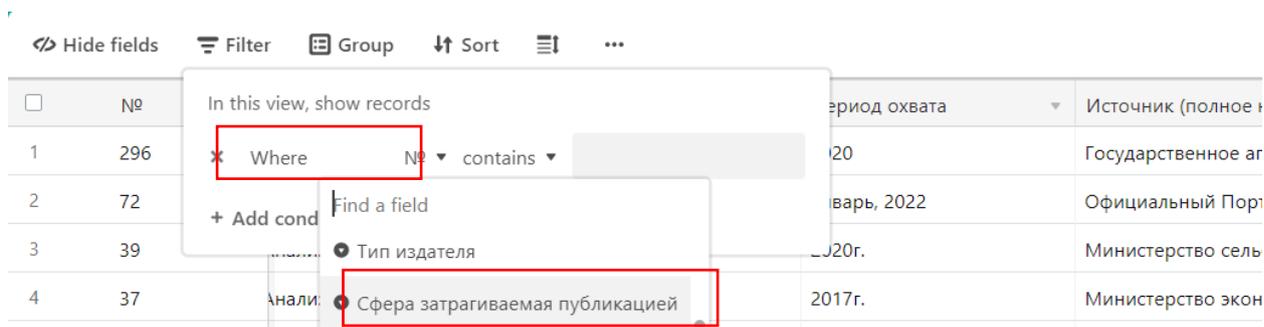
Airtable обладает простым в использовании и понятным интерфейсом. Получив доступ к базе данных на Airtable ([ссылка](#)), выполните несколько простых шагов, чтобы получить нужные данные. Основные инструменты для использования базы данных находятся в левом верхнем углу и включают следующие функции:

1. Фильтровать
2. Группировать
3. Сортировать
4. Высота строки

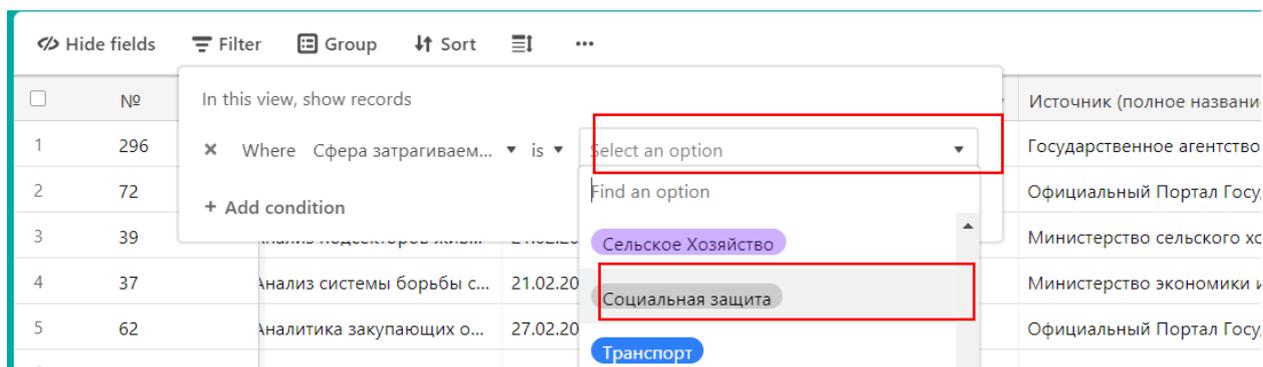


Скрыть поля – эта функция позволяет скрыть выбранные столбцы. Например, когда у Вас слишком много столбцов и они не помещаются на экране, функция «Скрыть поля» позволяет убрать с экрана ненужные столбцы.

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».



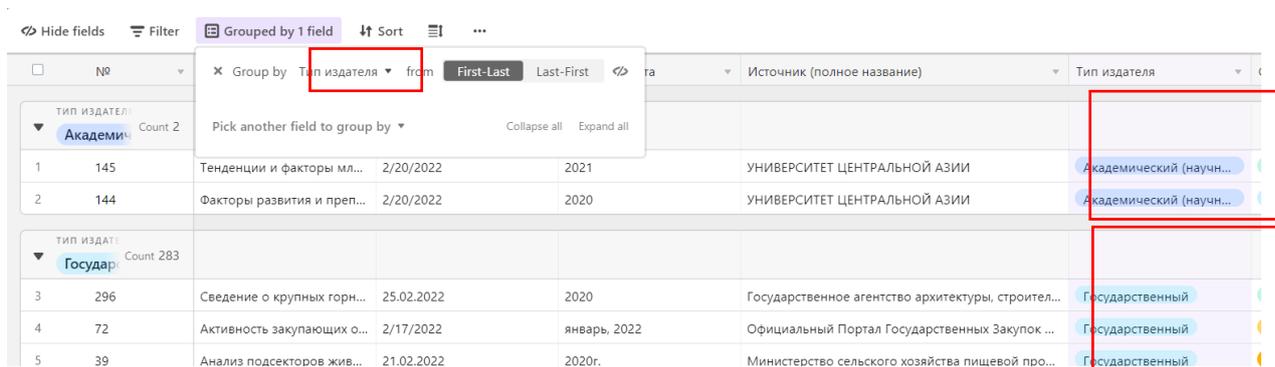
Далее выберите опцию «Социальная защита» в меню.



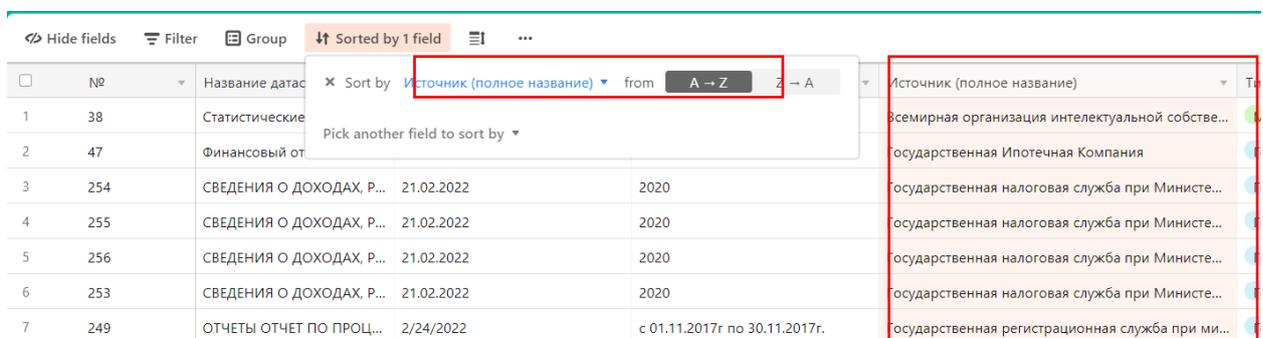
Теперь в таблице будут отображаться только наборы данных, относящиеся к области социальной защиты.

Чтобы удалить примененный фильтр, нажмите **×** 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».



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», который содержит ключевые слова. 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.

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 пользователи данных выбирают только ту информацию, которая подтверждает их утверждения, оставляя без внимания данные, которые могут опровергнуть их версию.

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. 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, группы меньшинств и т.д.?

Остерегайтесь ошибочных логических выводов. Это 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: Письмо о запросе данных (Оригинал)

ЛОГОТИПИ РЕДАКСИЯ

Сод _____

Ба кй:

Ба кучо:

ДАРХОСТ

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

1.

Сана:

Ному насаб, вазифа, ташкилот

Телефон барои тамос

Приложение 4: Некоторые инструменты визуализации данных для журналистов

1. [Datawrapper.de](https://www.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 донести свою историю и исполнить роль хранителей общественных интересов. Для ознакомления с Datawrapper, пожалуйста, нажмите здесь. Если вас интересуют более сложные темы, мы предлагаем воспользоваться ресурсами Академии Datawrapper - <https://www.datawrapper.de/>

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://flourishr.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>

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

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](#). Он позволяет

конвертировать отсканированные изображения в текст (до 8 Мб). Если Word и Img2TXT не работают, Вы можете приобрести АBBYY 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 АBBYY FineReader.

Вариант 3: Введите данные в машиночитаемые файлы вручную.