

# Как проверять данные и источники информации

2020-й стал не только годом пандемии, но и годом открытых данных.

Многие организации, связанные со здравоохранением, ежедневно в режиме реального времени публиковали новейшую информацию о распространении вируса по всему миру, делая беспрецедентный объем данных доступным для публики. Для журналистов трудность заключалась в том, чтобы точно проанализировать эту информацию и эффективно донести ее до своей аудитории.

Очень важно, чтобы журналисты сначала сами разобрались в данных, с которыми работают. И хотя в сегодняшнем непрерывном новостном потоке часто приходится спешить с публикациями, неточная информация может принести больше вреда, чем пользы. Во время такого кризиса, как COVID-19, данные помогают повысить осведомленность о важных проблемах. Но искаженная информация может привести к еще большему риску.

Всегда подходите к анализу данных со здоровым скептицизмом. Журналисты должны исследовать, откуда получены данные, которые они используют. Важно понять, кто собрал и опубликовал эти данные, кто финансировал исследование.

Журналисты также должны исправлять нелогичные или отсутствующие данные и убирать неправильно маркированные цифры. Такие ошибки могут возникать в процессе ввода данных - независимо от того, вводили их вручную или автоматически.

Например, как [сообщил местным СМИ](#) бывший министр здравоохранения Иордании Саад Джабер, раньше Министерство здравоохранения этой страны вручную вводило те результаты тестов на COVID-19, которые не загружались в правительственную базу данных автоматически. Когда количество ежедневных тестов увеличилось, результаты были потеряны - и появились ошибки, связанные с именами и результатами тестов.

Также имейте в виду, что даже в таких надежных программах, как Microsoft Excel, в базы данных могут закрасться сделанные людьми ошибки. Например, [так произошло в Великобритании](#) в прошлом году: 16 000 записей о пациентах с COVID-19 были случайно удалены из официальной базы данных, что привело к распространению неточной информации, что, в свою очередь, затруднило важные для борьбы с вирусом усилия, включая отслеживание контактов.

Чтобы избежать публикации неточных данных, используйте информацию из заслуживающих доверия источников и проверяйте данные. Вот о чем нужно при этом помнить.

## Прозрачность

Ищите ресурсы, открыто объясняющие, как они собрали и задокументировали данные, включая информацию о технологиях и алгоритмах, которые использовались во время всего процесса. Чем более прозрачны процессы сбора данных, тем более вероятно, что эти данные точны.

Убедитесь, что понимаете, как источники, на которые вы опираетесь, собирали свои данные. Это поможет вам лучше анализировать и верифицировать информацию, прежде чем включать ее в свои материалы.

## Методология

Публикуя набор данных, обязательно прикрепляйте соответствующие файлы с [метаданными](#) - это поможет объяснить, как данные были собраны. Такой файл может также включать информацию о размере выборки, пределах погрешности и недостающих значениях и список терминов и аббревиатур.

К примеру, в [Италии журналисты обнаружили несоответствия](#) в предоставленных правительством официальных данных о COVID-19, подвергли эту информацию сомнению. Такие несоответствия могли быть вызваны рядом факторов, включая то, что правительство несколько раз в течение 2020 года меняло политику тестирования, а также различия в методах отслеживания случаев заражения вирусом в разных регионах. В результате данные были представлены непоследовательно. Если бы к данным был прикреплен файл с метаданными, эти ошибки легко можно было бы идентифицировать.

## Контекст

При анализе данных контекст критически важен. Например, подумайте, как представить информацию об общем числе случаев и уровне заболеваемости. Когда правительство представляет региональные данные о количестве инфицированных COVID-19, больше всего случаев будет зарегистрировано в крупных городах. Но это не обязательно значит, что там самый высокий уровень заражения: такие высокие показатели могут быть просто следствием того, что этот район густонаселен.

Лучше сравнивать данные по количеству случаев заболеваемости на 100 человек. Представленная таким образом информация будет точнее отражать степень распространения вируса.

## Как понять данные

Не начинайте работать с базой данных, если вам непонятно, какую информацию они представляют. Задайте себе следующие вопросы:

- На что указывают данные?
- Понимаю ли я все термины и определения, связанные с данными?
- Какая информация, которая могла бы предоставить дополнительный контекст, не включена в данные?
- В каких единицах все измеряется?
- Можно ли для подтверждения сопоставить эти данные с данными из других источников?

Для проверки данных нужно проводить исследования и анализ. К счастью, чтобы это сделать, журналистам не нужно быть экспертами в анализе данных. Журналистские ценности, наработанные в журналистике навыки, профессиональный инстинкт – все это эффективно помогает проверять данные. Проверка вручную может быть более эффективной, чем автоматическая проверка с помощью алгоритмов. Хотя технологии не всегда помогают определить достоверность данных, они могут предоставить журналистам полезные инструменты и справочную информацию.

Всегда задавайте вопросы, относитесь к информации со скепсисом. Где возможно, проверяйте данные и сверяйте их с аналогичной информацией. На приведенной ниже иллюстрации показаны шаги, которые я предпринимаю при работе с информацией в базе данных. Возможно, это поможет вам разработать собственную стратегию проверки.

<https://ijnet.org/ru/story/%D0%BA%D0%B0%D0%BA-%D0%BF%D1%80%D0%BE%D0%B2%D0%B5%D1%80%D1%8F%D1%82%D1%8C-%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5-%D0%B8-%D0%B8%D1%81%D1%82%D0%BE%D1%87%D0%BD%D0%B8%D0%BA%D0%B8-%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%B8>