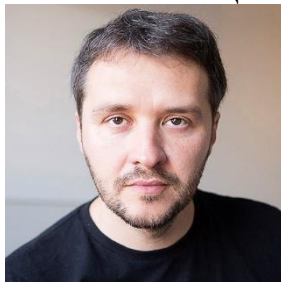


УДК 004.4

## КРИЗИС РЕПРЕЗЕНТАЦИИ ДАННЫХ



**А.М. Шинкаренко**

*Директор ООО «Датаверитас», Аналитик,  
Архитектор систем Business Intelligence*

*LinkedIn: <https://www.linkedin.com/in/aleksei-shinkarenko/>*

*E-mail: [aleksei@esnes.by](mailto:aleksei@esnes.by)*

### **А.М. Шинкаренко**

*Окончил Белорусский национальный технический университет. Работает в области аналитики данных. Преподает в «IT-Academy» основы визуального анализа данных. Проводит научные исследования визуальной топологии структур данных и их дискретных состояний.*

**Аннотация.** Мы переживаем случай кризиса универсального типа, в структуру которого интегрированы практически все измерения человеческого существования: биологическое измерение кризиса переплетается и оказывает деформирующее воздействие на экономику, социальные, политические, научные и другие планы существования и развития. Текст фокусируется на исследовании симптомов проявления кризиса в среде аналитики данных. Это прежде всего «инфляция доверия» данным, связанная с изменением их природы и поведения, с их парадоксально одновременным существованием в двух противоположных версиях. Их управляемая репрезентативность, как ключевая базовая способность выражать с математической точностью реальные события, оказывается в условиях кризиса низко эффективна при использовании в процедурах убеждения или процедурах принятия решений. В данной работе произведен апостериорный анализ серии аналитических ситуаций, в которых маркируются косвенные признаки текущего кризиса репрезентации данных. Чтобы обозначить горизонт проблемы репрезентации данных и их отождествления с реальным миром, рассмотрены примеры аналитических коммуникаций, эксплицированы паттерны выражения признаков кризиса в текущей практике аналитической репрезентации данных. Цель работы, сфокусировать внимание на специфике кризиса репрезентации данных в контексте универсального кризиса и создать коррелятивные отношения развития между кризисом и квалификационными требованиями для аналитиков нового поколения.

**Ключевые слова:** системная аналитика, аналитика данных, системы поддержки принятия решений, визуализация данных, проблема репрезентации, бизнес анализ, большие данные.

**Введение.** Данные в условиях кризиса стали вести себя не так, как в условиях стабильности. Поэтому невозможно внятно сказать, «что на самом деле происходит», и увидеть в данных, «что будет дальше». Графики заболеваемости и смертности, демонстрируя разные версии события, вызывают мало доверия [1]. Предсказания на основе исторических данных нарушаются выбросами новых данных в эпицентрах распространения вируса. Оберегающие наше будущее логические модели нарушены эпидемией.

В то же время что-то скрывается за самой озабоченностью аналитиков восстановлением моделей, планов и прогнозов, их тревогой, связанной с утратой данными их управляемой *репрезентативности как способности выразить с математической точностью реальные события*. Аналитика данных оказалась подвержена какому-то событию, выходящему за рамки того, что мы обозначаем «пандемией». И проблема в том, что неясно, каким научным инструментарием нужно воспользоваться, чтобы его назвать и тем самым локализовать для дальнейшего изучения. Аналитикам в этих условиях с огромным усилием

приходится отождествлять данные с картиной мира, и все равно им не хватает какого-то недостающего элемента знания о происходящем.

Здравому смыслу традиционно удавалось удерживать данные от проявления своего дивергентного, несогласованного, невозможного (*impossibilité* [2]) состояния. У него было как минимум два достаточных основания для этого — механизмов рынка, доведенных до высокой степени автоматизма процессов, и оскудения палитры ожидаемых исследователями результатов, которые нужно было подтвердить. Можно отметить, что у здравого смысла всегда был временной запас, чтобы успеть создавать интерпретацию, распределяя поступающие данные по соответствующим «картинам мира». Находящиеся в движении данные искусственно затормаживались и перерабатывались алгоритмами. Усилия направлялись на распределение данных между интерпретациями мира, вынуждая таким образом мир размножаться в его интерпретациях и закрепляя эти версии мира в границах невозможного существования [3]. Таким образом данные о «неправильной» картине мира получалось изолировать. Но потеря преимущества в скорости интерпретации и выброс непонятных, новых данных привел к ситуации, когда разобщенные когда-то в серии данные наводнили один и тот же мир. Оказались в одной таблице, парадоксальной матрице-сцене, где одновременно выполняют роль и растущего, и обваливающегося рынка, и зараженного, и не зараженного вирусом мира.

**Открыть данные и понять, что все хорошо.** Для создания архитектур систем поддержки принятия решений (Business Intelligence) приходится создавать описания формальных требований к процессам непрерывного анализа данных. Распространённый способ для сбора требований — это аналитическая коммуникация в форме интервью [4]. Тщательная работа с транскрипциями, записанными в процессе сбора не формализованных требований помогает заметить признаки кризиса репрезентации. В некотором смысле аналитик догадывается, что кризисная ситуация уже сложилась, и его задача — спровоцировать ее появление, помочь выразить ее признаки своими вопросами [5]. Приведем несколько типовых примеров аналитических коммуникаций, в которых сохранены речевые конструкции, но изменены коммерческие и персональные признаки интервьюируемых. Обратим внимание на то, как интервьюируемые сочетают слова для выражения проблемы, связанной с данными и аналитической задачей.

Одна из самых распространенных ситуаций репрезентативных действий — интервьюируемый отождествляет понятия и цифры.

**Пример 1.** *«Нам нужен инструмент ежедневного контроля за достижениями показателей продаж. Мы мониторим эти штуки каждый месяц... Если открыть данные, по сути, это сделки по оборудованию, просто в рамках одного месяца. Соответственно, каждую сделку продавцы вносят вот в этот эксель, и, соответственно, потом трекают там... то есть статус ее меняется или не меняется, то есть они все эти данные вносят в эксель там. Продавцы, они вносят эти сделки в ERP-систему и плюс вносят в отдельные файлы, а уже менеджеры вышестоящие, то есть там, на директорском уровне, им интересны как бы именно вот уже цифры, им там как бы конкретные сделки не интересны. Им интересны цифры. И нужна креативность в смысле в том плане, как можно обыграть эти же таблицы, только другими средствами, и показать, какие возможности появляются в связи с этим. Например, временная шкала и, допустим, отслеживание каких-то может быть трендов — тут как бы увеличивается и уменьшается, да? Или сравнительный анализ, как я относительно прошлого месяца, допустим, выгляжу, и понять... Потому что сейчас вот как бы этих инструментов вот нет в целом, просто цифры какие-то. Вот, хотелось бы знать, по отношению к предыдущему я вот уменьшаюсь или там увеличиваюсь?»* (из интервью с руководителем отдела продаж)

В приведенном примере данные превращаются «по сути» в «сделки», внесением в ячейки таблицы как «цифры». Для интервьюируемого они взаимозаменяемы, вступают в

отношение «соответствия». Подобное отождествление позволяет сделать любопытный перенос, где лицу, принимающему решение, необходимы уже не сделки, а их абстрактные числовые представления, и на передний план аналитической задачи риторически выводится задача преподнесения данных в понятиях «уменьшаюсь-увеличиваюсь». Уже обыгранные цифры нужно показать, формируя точку зрения в границах оппозиции различий, формируя представления о ситуации в продажах через «креативность в смысле» в «обыгранных таблицах».

**Пример 2.** *«Мне хочется открыть дашборд и понять: все хорошо, и пойти спокойно дальше. А если что-то не так, то я хочу понять, в каком месте бизнеса, или кто-что повлияло на негативное значение... Мы по всей выручке провисли, либо где-то в чем-то конкретном. Я хочу, чтобы мне что-то просигнализировало, должен я дальше копать или не должен, например, увижу цифру 25% и она еще и красная, это значит, что мне нужно будет думать, куда лезть разбираться».* (из интервью с руководителем)

Отождествление цифры с эмоциональными характеристиками — «хорошо» и «спокойно» — часто встречается при формировании аналитической задачи. Интервьюируемым важно, чтобы цифра обладала потенциалом к сопереживанию, и часто создают установку на отображение данных как позитивной картины: «хочу открыть дашборд и понять: все хорошо».

**Пример 3.** *«... В итоге он согласился, да. Мы просто смогли обосновать в данных... Ну то есть почему мы это сделали. Ну, просто я понимаю, что, например, ну для меня лично, если б я была на его месте, для меня это был бы не идеальный вариант. Ааа... потому что, опять же, это очень относительная вещь... Этот график показывает позиции категорий отгрузок относительно друг друга за конкретный месяц... То есть, в моем понимании, нужно было бы в идеале определить какие-то четкие понимания вот этих секторов рынка... Ааа и... вот эту модель анализа ко всем секторам рынка, чтобы можно было вот как-то спроецировать вот эти отгрузки».* (из интервью с аналитиком)

Так же часто, как требование подать данными сигнал к действию, встречается запрос на аналитическое убеждение. В аналитической практике приспособление данных к позитивному сценарию развития зачастую связывается с игнорированием негативных причин, следствий или последствий. Аналитическое усилие направляется на дробление задачи, чтобы возможные проявления негативного сценария не могли проявить себя, тотальная сборка данных была неосуществима ни на уровне отделов, ни на уровне руководства компаний. Решаемая задача направлена на поддержку равновесного положения вещей и событий, как в Примере 4. Установка на поддержку точки зрения звучит практически всегда при согласовании требований. И успешной аналитикой становится та, которая смогла максимально отдалить данные, в которых может содержаться риск проявления несогласованности бизнес-моделей и описываемой ими реальности.

**Пример 4.** *«Мы ограничены во времени и в бюджете. Нужно постараться сделать максимально просто. Тем более, в нашей конкретной ситуации, у нас эти данные уже собраны, и надо работать вот с тем, что есть. Нам интересно правильно преподнести историю накопленную, чтобы сразу же получить результат. Для своих уже там подвижек, для своих выводов, куда дальше двигаться. История нам больше важна, как бы важнее то, как ты это рассказываешь и какие ты доводы приводишь, чем по факту на самом деле, как оно есть».* (из интервью с руководителем отдела)

В некотором смысле аналитик вынуждено подыгрывает подобным сценариям, выясняя максимально точно, какую точку зрения или предубеждение нужно будет поддерживать репрезентацией данных. Возможно, за этим скрыто стремление защитить себя от множеств вариантов, и по своему не усмотрению создать аналитическую конструкцию на основе данных, которая может инициировать кризис.

**Пример 5.** *«Ты каждый раз приходишь на предприятие, рисуешь их бизнес процессы, начинаешь раскладывать данные между собой и видишь, что ну как бы фигня какая-то происходит. Дальше по сути успешность тех проектов, с которыми я работал, зависела только от величины моего административного рычага. У меня был проект, на котором административный рычаг большой, и я за 3 месяца делал аналитику и все заходило»* (из интервью с бизнес аналитиком)

**Пример 6.** *«Чтобы это рассказать мне слов не хватало, и мне нравилось, чтобы это не рассказать, а это показать. Попробовал другими средствами, ну когда не думал про визуализацию данных, сделал большую таблицу, в которой мне казалось, что мне понятно и все видно, показывал менеджерам и они долго смотрели и, короче, не могли понять, о чем я им говорю»* (из интервью с аналитиком данных)

В установке на необходимость интерпретационно-репрезентативной поддержки при выводе данных заинтересованы не только аналитики (Пример 5 и Пример 6), но и руководители, как можно увидеть в Примере 7.

**Пример 7.** *«На своем примере знаю, насколько я или кто-то другой может исказить ситуацию, не потому что мозг в этот момент плохо обрабатывает информацию, а есть фильтры эмоций, культуры, знаний, предубеждений, заблуждений, накопленных ошибок, опыта. И через эту призму фильтров проходит поток данных, даже не информации в виде образов, а данных. И моя задача — узнавать эти фильтры и останавливать их влияние при принятии решений. На больших скоростях это не всегда получается. Особенно трудно исключить эмоциональный фон. Я прекрасно понимаю, что в потоке больших объемов информации, которые через меня за день проходят, вероятность серьезных ошибок велика, но как эти искажающие фильтры работают, я пока не совсем понимаю. Там, где мозг на подсознательном уровне сопротивляется данным, где он не хочет очень глубоко изучать, потому что там созидательного ничего нет мне нужна помощь».* (из интервью с руководителем компании)

Не является ли аналитика данных инструментом установления границ разумного, искусством «выражения»? Выражения интересов, произрастающих из логико-математического схематизма мышления. Расположением точек зрения на события и вещи, как будто совершается попытка утверждения опытного знания отрицанием иллюзии, выражением факта, якобы изъятая опытным знанием из ткани событий. И уже эта конструкция из подтвержденного опытом знания предъявляет «истолкованное» и как результат действия, и как представление настоящести и завершенности опыта. Позитивное, опытное знание словно требует фиксации в изображении всех меняющихся отношений между мыслящим и мыслимой вещью для утверждения его достоверности [6].

Аналитик, вступая в сопротивление обманчивой иллюзии, функционирует как куратор, организатор постоянных пересборок фрагментов ситуативных данных; комбинирует конфигурации данных вокруг предмета, нежели отыскивает его правильное значение. Искусство репрезентации данных — способ удержания аналитиком динамических форм на грани их остановки, разворачивания утонченной языковой игры с сознанием зрителя.

**Эдвард Тафти и преднамеренное подтверждение.** Книги Доктора Эдварда Тафти, великого художника статистики, о представлении данных — прекрасное свидетельство того, как экономика стала тождественна искусству. Осенью 2016 года Тафти выступил с предостережением о кризисе сравнений [7]. Он высказывает свою озабоченность тем, что в отчетах преобладают *преднамеренные подтверждения* (confirmation bias). Пропустим социальный мотив в его опасениях и посмотрим, как фигура кризиса выделяется на фоне текстов самого Тафти. Возможно ли сегодня исследование без преднамеренных подтверждений, использование данных вне предустановленных смыслов? Тафти утверждает новое правило, новую границу «до и после», — «...Теперь, видимо, настал момент, когда мы должны провести жесткое разграничение между подтверждающими заблуждения отчетами и

исследовательско-детективной работой» [8]. Канонически за фундаментальной задачей анализа утверждена установка причин и следствий между свидетельствами, умозаключениями и выводами. В эпоху данных это реализуют с помощью данных и методов сравнения.

Новое правило, вероятно, нужно для того, чтобы сравнивать точнее. Неким жестом усиливается основание для сравнения данных. Вводится дополнение к правилам через дробление существующих и их пересборку. Но вопрос «С чем сравнивать?» существует и реализуется как метод благодаря бесконечности его продления. В этом его опасность. [прим] Возможно, система становится самождественной, когда желания субъектов больше не вступают в противоречие с преднамеренными подтверждениями. Разотождествить эти понятия – нетривиальная задача и поэтому продолжается добавление новых правил деления. Система становится все мельче, еще сложнее, что только увеличивает путаницу и создает эффект камуфляжа ложных выводов о «правдивых данных».

**Посмотри и будь осторожен.** Сценарий кризиса аналитики данных регулярно проигрывается в художественных фильмах, фантастический сценарий которых может указывать на отдаленное во времени событие, которое мы удерживаем на дистанции максимально натуралистичным, «фотографическим» описанием. В фильме «Предел риска» [9] о кризисе 2008 года, кризис начинается с момента передачи данных — момента критического соединения данных в особенную аналитическую фигуру. Для обхода защитных механизмов, предотвращающих от тотальности данных, способных вырваться из моделей интерпретации, не хватало только небольшого фрагмента — достижения данными критической формы и точки невозврата, «когда музыка уже не звучит».

Аналитика из отдела рисков, работавшего с этими данными, сокращают. Покидая офис, он передает флешку с незавершённой работой своему преемнику и говорит: «Будь осторожен». Молодой аналитик изучает материал и дополняет данные недостающим фрагментом, доводя их до состояния запретной полноты, после чего кризисные события моментально разворачиваются.

В некотором смысле, собрание уникальной коллекции данных осовременило древнюю страсть людей к коллекционированию как способу овладения миром [10]. Принципы репрезентативности цифр, синтаксиса математики, технологий вычислений позволили произвести замену «проекции вещей» их «числовыми проекциями». Установление этой сцепки через данные позволяет достигнуть предельно тесной и общей связи вещей с себе подобными. Эту операцию и проделал аналитик в «Пределе риска».

Каким-то образом вещи оказываются максимально выраженными через данные. На этой степени близости происходит «электрический разряд» — производство катастрофического события. Будучи оцифрован, предмет или явление становится идеальным с точки зрения избавления от всех своих изначальных функций. Идеальным экземпляром для того, чтобы им обладать [11], встроить его в новую структуру, совершенную, упорядоченную структуру баз данных и таблиц, новых форм коллекций. Но сама структура не идеальна, а функциональна. В кризис структура и связи ослабли. Неожиданно данные перемешиваются и отказываются поддерживать функцию универсальной линейки, выражающей вещи и события. Каким образом данные собираются и распадаются сейчас? В каких структурах мы можем их уловить и использовать для сборок своих репрезентативных моделей реальности?

*Заключение.* Нужно принять во внимание, что данные будут приведены в соответствие своему привычному состоянию тождественности здравому смыслу. Но чему можно научиться, пока не произошло то, что описывают фразой «кризис пройдет»? Может звучать странно, но аналитическая практика мало озабочена согласованностью данных, она скорее ориентирована на их репрезентативные свойства, способность поддерживать автоматизм и симплификацию. Одно из действий кризиса в том, чтобы устранить накапливающееся несоответствие на разных уровнях репрезентативной сборки и помочь тем самым аналитикам.

Провести в некотором смысле сброс агрегированных состояний данных, чтобы измельчиться, пересобраться, мутировать и выйти на новый цикл сборки и цикл роста.

К признакам кризиса данных важно внимательно присмотреться хотя бы для того, чтобы выделить их «деятельность» из множеств методов смешения реальности с репрезентацией. Поставить под сомнение принятие на «веру» того, что данные воплощают такие абстрактные сущности, как деньги, бизнес, рынок, клиент, товар. Данные вовлечены и молчаливо работают над «тотализацией» [12] и пересборкой картины мира. Все, что есть по отдельности, безотносительно друг к другу, бессвязно, раздроблено, обволакивается ими, заполняя пустоту между структурами и связями, уже подчиненными самой логике данных. Великая цифровая «грибница» больших данных. С математической точностью понятий, склеенных из абстракций информационных сообщений, поддерживается цифровая версия динамической и последовательной модели мира. Идеально выверенное сцепление данных с идеями. Нет, вероятно, способа полностью отделить данные от их репрезентативной функции долговременно и настолько надежно, чтобы обслуживаемые ими иллюзии были навсегда заблокированы. Так же, как нет способа обеспечить сопротивление аналитической мысли идеологическому измерению, имплантированному вещам. Перманентная мутация образов и пересборка понятий и смыслов требует от современных аналитиков новой квалификации. Постоянного переизобретения мер предосторожности против того, как данными овеществляют понятия. Необходимо обновление и самого анализа данных, непрерывное исследование и диагностика функциональности данных в создании политических и идеологических понятий — той роли, которую они стали сегодня играть в наших воображаемых решениях наших реальных проблем.

#### **Список литературы**

- [1]. [Электронный ресурс]. - Электронные данные. - Режим доступа <https://www.colta.ru/articles/society/24006-mihail-yampolskiy-o-tom-kak-pandemiya-razrushaet-illyuziyu-znaniya> — Михаил Ямпольский «Эпидемия: незнание и правда»
- [2]. Лейбниц, Готфрид. Сочинения в 4-х томах. Том 3. М.: Мысль, 1984. — 734 с. — (Философское наследие. Том 92, фрагмент «Абсолютно первые истины»). — С. 123
- [3]. Делёз, Жиль. Складка. Лейбниц и барокко / посл. В.А. Подороги. Пер. Б.М. Скуратова — М.: изд. «Логос», 1997. «Глава 5. «Несовозможность», индивидуальность, свобода.» — С. 103
- [4]. BABOK. A Guide to the Business Analysis Body of Knowledge. International Institute of Business Analysis, Toronto, Ontario, Canada. Version 3.0. published 2015. — p. 290
- [5]. Silver, Nate. The signal and the noise : why most predictions fail but some don't. Penguin Press, 2012. — p. 14
- [6]. Подорога, Валерий Александрович. Метафизика ландшафта: коммуникативные стратегии в философской культуре XIX-XX вв. Наука, 1993 («Парадоксия стиля: выражение и опыт») — С. 286
- [7]. Tufte, Edward - The Future of Data Analysis. Slide «The crisis in data analysis: most published studies are false.» [https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg\\_id=000AIz](https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=000AIz)
- [8]. Microsoft Machine Learning & Data Science Summit 2016. Keynote Session: Dr. Edward Tufte - The Future of Data Analysis Sep 28, 2016. (00:04:27 --> 00:04:32) <https://channel9.msdn.com/Events/Machine-Learning-and-Data-Sciences-Conference/Data-Science-Summit-2016/MSDSS11>
- [9]. Предел риска (Margin Call). Режиссер: Джей Си Чендор Год: 2011. <https://www.kinopoisk.ru/film/503671/>
- [10]. Беньямин, Вальтер. О коллекционерах и коллекционировании. М.: ЦЭМ, V-A-C press, 2018. С. 21, 75
- [11]. Маркс, Карл. Экономическо-философские рукописи 1844 года. Маркс К., Энгельс Ф. Сочинения. М., 1974. Т.42, — С. 120
- [12]. Джеймисон, Фредерик. Постмодернизм, или Культурная логика позднего капитализма / пер. с англ. Д. Кралечкина; под науч ред. А. Олейникова.—М.: Изд-во Института Гайдара, 2019. — С. 642

## **CRISIS OF DATA REPRESENTATION**

**A. Shinkarenko**  
*Dataveritas CEO,  
Analyst, Business Intelligence  
systems Architect*

*Linkedin: <https://www.linkedin.com/in/aleksei-shinkarenko/>*

*E-mail: [aleksei@esnes.by](mailto:aleksei@esnes.by)*

### **A. Shinkarenko**

*Belarusian National Technical University graduate. Works in the field of data analytics. Teaches the basics of data analysis in “IT Academy”, Minsk. Conducts research on the visual topology of data structures and their discrete states.*

**Abstract.** We are experiencing a case of a universal type of crisis, in which almost all dimensions of human existence are integrated: the biological dimension of the crisis is intertwined and has a deforming effect on the economy, social, political, scientific and other planes of existence and development. The text focuses on the study of symptoms of crisis in the data analytics environment. This is primarily an “inflation of confidence” in data, which is associated with changes in their nature and behavior, with their paradoxically simultaneous existence in two opposite versions. Their managed representativeness, as a key basic ability to express real events with mathematical accuracy, is low effective in crisis situations when used in persuasion or decision-making procedures. In this paper, a posteriori analysis of a series of analytical situations is performed, in which indirect signs of the current crisis of data representation are marked. To define the horizon of the problem of data representation and identification with the real world, examples of analytical communications are considered and patterns of expression of crisis signs by the analyst are revealed. The purpose of this work is to draw attention to the crisis of data representation as something that requires new qualifications from modern analysts.

**Keywords:** System analytics, data analytics, decision support systems, data visualisation, representation problem, business analysis, big data.