

24 января 2025

Достоверность данных — необходимое условие доверия к ним

Опрошенные эксперты делятся своими наработками по повышению доверия к данным. Центральное место в их рекомендациях отводится мерам по обеспечению достоверности данных.

Достоверность данных — одно из главных условий доверия к ним со стороны бизнес-пользователей. Но, очевидно, это не единственный фактор, влияющий на доверие. Как достоверность данных поможет обеспечить доверие к ним? Какого рода действия и инициативы в наибольшей степени способствуют повышению достоверности данных, улучшению их качества и росту доверия к ним при принятии решений? На эти и другие вопросы мы попросили ответить экспертов, планирующих участвовать в форуме «Качество данных — 2025».

ОТ ДОСТОВЕРНОСТИ — К ДОВЕРИЮ

Эксперты сходятся во мнении о том, что **доверие к данным напрямую связано с тем, как воспринимают данные их пользователи.**



«Доверие к данным определяется в первую очередь психологией людей, корпоративной культурой компании и отношением к информации. Если решения в компании принимаются без аналитического обоснования, сотрудники не будут использовать ни сами данные, ни инструменты для работы с ними. Повысить доверие к данным поможет совокупность факторов: авторитет и доказанная эффективность работы CDO, появление более сложных задач, требующих качественной аналитики, внедрение инструментов, позволяющих отслеживать происхождение и историю изменения данных, и пр. — считает Арсен Кудзиев, руководитель отдела аналитического департамента компании PARMA TG. — Похожую задачу наши специалисты решили в ходе проекта в Пермском крае, где в нескольких тысячах бюджетных учреждений была внедрена централизованная облачная система управления финансово-хозяйственной деятельностью, основанная на дата-центричном подходе и включающая специально разработанные аналитические инструменты для получения сводных показателей. Благодаря им любая цифра консолидированных отчетов может быть разложена на понятные составляющие, а источники данных отслеживаются вплоть до первичных документов и отдельных бухгалтерских проводок».

Иван Вахмянин, управляющий партнер Visiology, предлагает исходить из того, что **людей нельзя заставить использовать данные и обязать принимать на их основе решения, тем не менее, можно сформировать культуру работы с данными**: «Сотрудники должны видеть, что призыв опираться на данные — это не просто слова. Например, если на совещаниях мнения не принимаются

без их обоснования на основе данных, уровень управленческой культуры сразу вырастает — а с ним и достоверность данных. Другими словами, наиболее важным инструментом в контексте работы с данными всегда является личный пример руководителя».

Сергей Полехин, владелец продукта PIX BI в компании PIX Robotics, рекомендует **отталкиваться от бизнес-задачи**: «Например, можно считать достоверными данные, содержащие корректную информацию обо всех сделках компании за последний год. Однако ответ на вопрос, можем ли мы доверять таким данным, зависит от конкретных задач. Если в 99% сделок их объем равен примерно 1000 рублям, а 1% сделок заключались на миллионы рублей, то можно говорить о достоверности данных в целом, но эти данные будут статистически недостоверными — без дополнительной обработки их нельзя использовать для обучения алгоритмов прогнозирования. Чтобы это стало возможно, необходимо применить к данным инструменты математической статистики и корректно удалить из исходного набора статистические выбросы. Но очищенный набор данных становится непригодным для корректного расчета премий продавцам, поскольку из него выпадают самые крупные сделки. Таким образом, важно рассматривать характеристики достоверности данных и доверия к ним только в контексте конкретных бизнес-задач».

Виктор Мастеров, директор департамента НСИ и интеграции компании SOFROS, предлагает **понимать под доверием к данным доверие к их**

источникам (поставщикам) и распорядителям (операторам) данных: «Сами по себе данные вряд ли целесообразно рассматривать в качестве объекта доверительного отношения. Полагаем, что достоверность, особенно если трактовать ее как воспроизводимую и контролируемо подтверждаемую точность, может быть отнесена к числу тех параметров качества данных, обеспечение которых в наибольшей степени способствует повышению репутации и доверия к работе поставщиков и распорядителей данных. Оптимальный способ достижения этой цели можно рассматривать как реализацию классической маркетинговой триады “возможность — преимущество — выгода”: высокое качество данных и, в частности, их достоверность можно соотнести с получаемым преимуществом, которое может быть использовано для повышения эффективности бизнеса и достижения коммерческого успеха. Демонстрация четкой зависимости между повышением достоверности данных и улучшением показателей бизнеса будет лучшей верительной грамотой для тех, кто предлагает такой продукт целевой аудитории».

Григорий Бокштейн, ведущий эксперт по управлению данными компании TData, приводит **пример того, как достоверность данных влияет на бизнес:** «Большая производственная компания, прежде чем выделить бюджет на создание, вывод на рынок и продвижение нового продукта, анализирует производственные мощности в разрезе географии присутствия, а также продажи в каждом конкретном регионе, потребительское поведение разных целевых аудиторий и многое другое. Если на каком-то этапе принятия

решения данные окажутся недостоверными, неполными, будут неверно проанализированы или интерпретированы, бизнес понесет убытки», — поясняет Бокштейн.

Александр Учаев, менеджер по продукту «1С:MDM» фирмы «1С», рассматривает **доверие к данным как результат тщательно организованного управления их жизненным циклом**, включающего выстраивание процессов управления их качеством, в том числе актуализации данных, проведения консолидированных проверок и постоянного мониторинга качества: «Получение достоверных и качественных результатов и их бесперебойное предоставление пользователям является основой, если угодно, репутацией, позволяющей судить о доверии к данным в этой организации».

По наблюдениям Олега Гиацинтова, технического директора DIS Group, **достоверные данные существенно повышают уровень доверия к данным в случае, если имеются явные доказательства высокого качества данных в соответствии с метриками каждого конкретного пользователя**: «Например, при построении важного бизнес-отчета обычно формируется привязанный к нему отчет о качестве использованных данных, представленный в метриках, релевантных для конкретного пользователя. Подобная работа должна быть точечной, направленной на персональное предоставление информации о качестве данных».

Доверие к данным напрямую зависит от того, насколько они достоверны, полны и точны, уверена Наталья Кудрявцева, функциональный архитектор компании Navicon: «Этого можно достичь, только если сделать процесс управления данными прозрачным и автоматизировать его. Сотрудники должны понимать, как данные собираются, хранятся, кто вносит в них изменения. При этом процессы, связанные с работой с данными, должны оставаться достаточно гибкими».

Одной только достоверности данных недостаточно для того, чтобы пользователи им доверяли, подчеркивает генеральный директор компании «Современные бизнес-аналитические решения» Виталий Миронов: «Необходимо также обеспечить прозрачность процесса работы с данными и добиться вовлечения пользователей. Для этого у них должен быть доступ к информации о происхождении данных, их проверке и обработке (включая описания метаданных, бизнес-правил и расчетов показателей, стандартов форматов, правила работы и др.). Кроме того, необходимы регулярные автоматизированные проверки и возможности для аудита данных. Нужно также обеспечить быструю реакцию на ошибки в данных с уведомлением об их исправлении — особенно это касается ошибок, о которых сообщают сами пользователи. Наконец, нужно предоставить им наглядные и понятные средства визуализации и обеспечить доступность данных, включая возможности для проверки выборок и промежуточных этапов расчетов».

Михаил Роцин, заместитель директора отделения департамента НСИ и интеграционных сервисов IBS, призывает **обратить внимание на последствия потери доверия к данным**: «Восстановить его бывает крайне сложно, это долгий и кропотливый процесс — необходимо на регулярной основе показывать, где возникают некорректные данные, на что они влияют и как найденные ошибки устраняются. Другими словами, нужны максимальная прозрачность и открытость данных. Поэтому лучше внедрять подсистемы контроля качества как можно раньше. Желательно, чтобы процессы проверки данных появлялись одновременно с развертыванием критически важных для бизнеса систем».



Алиса Школьникова,
Руководитель направления
Data Governance
департамента аналитических
решений ГК «КОРУС
Консалтинг»

Очевидно, что, если в данных постоянно встречаются ошибки, сотрудники перестают им доверять. Возврат к доверию происходит небыстро, но он точно возможен благодаря регулярно проводимым процедурам сверки. Кроме того, необходимо сформировать критерии и предоставить инструменты, позволяющие сотрудникам самостоятельно валидировать достоверность данных, например, проверять отчетность на корректность и соответствие данным из учетной системы. Важно при этом превратить валидацию в интересную задачу с понятными целями и результатами. Самым активным пользователям данных можно предложить варианты с геймификацией.

«Доверие к данным формируется благодаря их последовательности, воспроизводимости и прозрачности, — продолжает Андрей Бойко, коммерческий директор B2B-Center. — Чтобы повысить ценность данных для бизнеса и облегчить их использование в различных процессах, мы



рекомендуем связывать данные с источниками, проверять их корректность, применять количественные метрики для оценки качества и надежности данных, а также гарантировать безопасность и конфиденциальность информации».

Согласно замечанию Андрея Андриченко, директора по развитию компании «ЭсДиАй Солюшен», **подключать внешние сервисы для обеспечения автоматической верификации и обогащения данных помогают интеграционные возможности MDM-систем:** «Например, для верификации данных о контрагентах может использоваться облачный сервис “Контур.Фокус”, позволяющий по ИНН организации оперативно получать и обновлять информацию о компании в справочнике контрагентов. В этом случае доверие к данным строится на доверии к внешнему источнику».

Николай Скворцов, руководитель направления методологии компании «IS: Логика Данных», **особое значение** придает **информированию заинтересованных лиц о взаимосвязях работ в области управления данными:** «Владение такой информацией — важная предпосылка повышения доверия к данным. Не случайно в области управления качеством продукции стали популярны методологии бережливого производства и шести сигм, ориентированные на учет связей. Более того, один из ключевых инструментов методологии шести сигм — диаграммы SIPOC (Suppliers, Inputs, Processes, Outputs, Consumers — “поставщики, входы, процессы, результаты, потребители”) — взяты на вооружение в DAMA DMBOK2, где они используются как основа контекстных диаграмм областей знаний. Ценность их в том, что они позволяют представить управление данными в виде набора

взаимосвязанных элементов. Такое целостное описание работ, составляющих управление данными, могло бы стать важным пограничным объектом (на языке социологии) — инструментом коммуникации между разнородными рабочими группами, обеспечивающим функционирование единого корпоративного информационного пространства в области качества данных».

ДЕЙСТВИЯ И ИНИЦИАТИВЫ, СПОСОБСТВУЮЩИЕ ПОВЫШЕНИЮ ДОСТОВЕРНОСТИ ДАННЫХ

Наши эксперты в целом поддержали **комплексный подход, охватывающий не только технологические мероприятия, но и инициативы организационного характера.**

Так, Кудрявцева обращает внимание на недавние **исследования, показавшие, что главными препятствиями в работе с данными компании считают именно людей и процессы:** «Следовательно, для повышения достоверности данных важны в первую очередь разработка общих для всей компании регламентов и формирование корпоративной культуры, ориентированной на высокое качество работы с информацией. Люди должны быть обучены, знакомы с лучшими практиками, должны знать, где найти точные, непротиворечивые данные и как извлекать из них ценную информацию. Кроме того, важно выбирать ИТ-инструменты, соответствующие требованиям бизнеса в области работы с данными. Так, компаниям,

внедрившим в ходе импортозамещения множество разрозненных ИТ-систем, полезно выстроить процессы централизованного сбора и унификации данных на основе MDM-систем. Наконец, важно добиться прозрачности процессов управления данными и четко распределить зоны ответственности и контроля изменений».



Алиса Школьникова,
Руководитель направления
Data Governance
департамента аналитических
решений ГК «КОРУС
Консалтинг»

Главное — это люди. Поэтому основные инициативы должны быть связаны с поиском ответственных за данные. Чем больше данных будет охвачено владельцами — формальными и неформальными, теми, кто реально отвечает за ввод данных, понимает, как определять их достоверность, проактивно минимизирует ошибки ввода и участвует в том, чтобы предоставлять данные в нужном для бизнес-целей виде, тем больше будет доля достоверных данных.

Игорь Моисеев, директор по развитию DataCatalog (входит в Группу Arenadata), также рекомендует сочетать технологические и организационные

меры: «Ключевыми действиями являются разработка стандартов и политик, определяющих приемлемые для организации нормы качества данных, определение ролей и обязанностей сотрудников, отвечающих за управление данными, и контроль качества данных». Моисеев уверен, что достоверность данных нужно поддерживать на всех этапах их жизненного цикла. В частности, надо встроить механизмы проверки вводимых данных на корректность и соответствие заданным форматам, внедрить автоматизированную проверку качества данных в таблицах с применением инструментов Data Quality, сформировать глоссарий, описывающий онтологию качества данных и обеспечивающий унификацию форматов данных, показателей, метрик и описаний объектов. Также важно задействовать опыт экспертов — представителей российских консалтинговых компаний и интеграторов.

Миронов, соглашаясь с коллегами, **среди необходимых действий и инициатив** отмечает **стандартизацию процессов, включая разработку регламентов, создание единого глоссария терминов и форматов данных**, применение таких методологий управления качеством, как бережливое производство и шесть сигм, а также обучение и развитие персонала. В комплекс мер также рекомендуется включить автоматизацию различных этапов работы с данными и применение машинного обучения для выявления аномалий и прогнозирования возможных проблем с данными. Важно также не забыть про мониторинг и аудит данных и, конечно, формирование культуры работы с данными, в том числе стимулирование открытого доступа к данным внутри организации. «Внедрение этих действий и инициатив не только улучшит достоверность данных, но и повысит эффективность бизнес-

процессов в целом», — добавляет Миронов.

Скворцов в качестве наиболее важной инициативы выделяет **разработку и реализацию корпоративной стратегии работы с данными**: «Она должна охватывать мероприятия по решению основных проблем, связанных с качеством данных, и созданию единого информационного пространства в этой области. Создание пространства подразумевает разработку и сопровождение пограничных объектов (моделей, диаграмм, глоссариев и т.п.) для обеспечения коммуникации между рабочими группами, вовлеченными в процессы управления и использования данных. Важно также предусмотреть деятельность по обеспечению единой интерпретации таких объектов всеми группами. Выработка общего понимания — ключ к повышению доверия, в том числе и к данным».

Кудзиев предлагает **в первую очередь выявить наиболее критичные для бизнеса области данных**: «Для их определения имеет смысл обратить внимание на ситуации, когда из-за низкого качества данных организация может столкнуться с серьезными проблемами и финансовыми издержками, например, штрафами со стороны регуляторов или просчетами в операционном планировании. Разделив задачи на меньшие по объему, мы ускорим решение бизнес-проблемы и улучшим качество данных для конкретной задачи. В дальнейшем добавление новых доменов будет способствовать формированию и развитию подхода Data Mesh».

«Сбор данных, обеспечение их достоверности и качества должны исходить из решаемой бизнес-задачи, — продолжает Полехин. — Поэтому прежде чем

пытаться повысить достоверность данных, нужно четко сформулировать задачу, которую нужно решить на их основе. Кроме того, необходимо периодически проводить аудит данных, чтобы уточнять, какие именно данные собираются, откуда доставляются и с какой периодичностью. Например, данные, которые использовались для описания и контроля какого-то бизнес-процесса, спустя год могут оказаться бесполезными или неполными из-за того, что в процессе произошли изменения или поменялись правила расчета показателей его эффективности. Обеспечение достоверности данных — непрерывный процесс, который должен быть строго соотнесен с пониманием бизнес-задач, решаемых на основе этих данных».

Усилия по обеспечению достоверности данных оправданы, поскольку решения, принимаемые на основе достоверных данных, будут более точными, отмечает Гиацинтов: «Для повышения достоверности, конечно, требуется систематизация процессов, нацеленных на повышение качества данных, и их регламентация. Кроме того, нужна отдельная организационная структура для управления достоверностью данных. Инструменты, применяемые для подобных работ, должны соответствовать требованиям пользователей в части повышения достоверности данных».

По мнению Мастерова, **описание подходов, термины и понятия, изложенные в DAMA DMBOK2, возможно принять за нормативную основу**, что также позволяет минимизировать вероятность некорректных интерпретаций. Следует подчеркнуть, что достоверность относится к тем параметрам качества данных, для обеспечения и контроля которых еще большее методологическое значение имеет подход, основанный на

адаптации цикла Шухарта–Деминга, состоящего из повторяющейся последовательности фаз: планирование — реализация — контроль — доработка — планирование и т.д. В число методов, важных для обеспечения достоверности, Мастеров советует включить контроль входных данных, контроль качества внешних данных, формализованный контроль изменений, автоматизированное (с ручным подтверждением) или ручное исправление данных, а также применение эффективных метрик качества данных и использование программных средств для его проверки и аудита. Разумеется, потребуется и обучение персонала, занимающегося подготовкой данных и особенно обработкой данных из внешних источников.

Андриченко акцентирует внимание на то, что **обеспечение качества данных в MDM-системах базируется на специализированных функциях контроля и верификации вводимых данных**, включая механизмы ограничительных таблиц, содержащих сведения о допустимых комбинациях значений атрибутов информационных объектов, связанных с определенным классом и извлеченных из соответствующей нормативно-технической документации: «Мастер формирования номенклатурных позиций, реализованный в MDM на основе механизма ограничительных таблиц, обеспечивает создание множества качественных записей без необходимости отслеживания запрещенных комбинаций значений атрибутов и без ручного ввода — основного источника ошибок».

«Наибольшее влияние на качество данных оказывают системные действия: проверка источников на надежность и прозрачность, валидация и фактчекинг, критический анализ с точки зрения логики и доказательной базы,

проверка актуальности данных», — продолжает Бойко.

Рощин среди действий и инициатив по повышению достоверности данных советует **реализовать меры по введению метрик качества данных, регулярных его проверок и минимизации ошибок, связанных с человеческим фактором при вводе данных.** Также, конечно, нужно описать бизнес-процессы работы с данными и жизненный цикл данных.

Учаев исходит из того, что только **комплексный подход к организации процесса управления данными как цифровым активом может обеспечить высокую достоверность обрабатываемых данных.** В качестве примера Учаев предлагает рассмотреть сервис обеспечения достоверности транзакций «1С:Сверка 2.0». Среди его функций — идентификация и проверка достоверности данных подключаемых организаций, сверка атрибутов контрагентов и сопоставление их с организациями, идентификация и сверка версий учетных систем, данных, а также принципов учета операций и документов, поиск и сопоставление корреспондирующих документов (включая их реквизиты и аналитические данные), выявление разногласий, их консолидация и хранение.

«К задаче повышения достоверности данных мы подходим комплексно, — делится опытом своей компании Бокштейн. — В первую очередь стремимся обеспечить высокие стандарты качества — для этого используем автоматизированные инструменты валидации и очистки данных, проверяем их на ошибки и аномалии. Также разрабатываем четкие стандарты для сбора, хранения и обработки данных. Кроме того, стремимся обеспечить

прозрачность процессов работы с данными — ведем подробную документацию об источниках, методах сбора и обработки данных, предоставляем метаданные для объяснения контекста данных, их значений и ограничений. Разумеется, обучаем команды, задействованные в работе с данными, собираем и анализируем обратную связь с ними. Наконец, развиваем культуру работы с данными: для этого важно, с одной стороны, заручиться поддержкой руководства, чтобы сотрудники рассматривали инициативу как значимую, и, с другой, — использовать достоверные данные для обоснования принимаемых решений, что также способствует росту доверия и уверенности в их качестве. Важно также регулярно проводить аудит данных (в том числе внешний) и предоставлять пользователям понятные прозрачные метрики качества данных. Не следует забывать и об обеспечении безопасности и конфиденциальности данных».

Согласно видению Вахмянина, **данные становятся действительно достоверными только тогда, когда руководители регулярно используют их для принятия решений** — тем самым они демонстрируют, что эти данные приоритетны, и поддерживают тем самым культуру управления на основе данных. Повысить качество данных помогает внедрение современной BI-платформы, благодаря чему данные становятся наглядными и понятными. «За счет их визуализации становятся видны их изъяны, специалисты получают возможность их исправить, — поясняет Вахмянин. — Проект внедрения BI может показаться довольно сложным, но он значительно упрощается при использовании лучших практик. Учитывая высокий уровень интереса к ним, мы взяли за их систематизацию и разработку первой российской методологии внедрения BI-технологий и управления на основе данных.

Первый релиз методологии выйдет в начале 2025 года».

Таким образом, **ядром мероприятий и инициатив по повышению достоверности данных выступают действия организационного и методологического характера**. Технологические проекты помогают довести их до логического завершения и переложить выполнение рутинных процедур на ИТ-инструменты.