

16 октября 2024

# Как правильно организовать работу с бизнес-гlossарием в сложных предметных областях

Важной задачей управления данными (Data Governance, DG) является поддержание однозначности и согласованности понятий и определений, используемых в организации. Без механизма, который позволяет различным подразделениям находиться в одном семантическом поле и договариваться о методологии, могут возникать разночтения в понимании и интерпретации данных. Это приводит к типичным ошибкам: расхождению значений показателей в отчетах, некорректным значениям на визуализации.

Рассмотрим, как организовать работу с бизнес-гlossарием, чтобы он начал приносить пользу бизнесу.

## Как найти ошибки в методологии

В отличие от, скажем, упавшего пайплайна, ошибки, связанные с неправильным использованием методологии, найти гораздо сложнее. Они часто возникают при взаимодействии бизнес-подразделений и технических специалистов — особенно если последние не работают в организации заказчика и не погружены в предметную область.



Ключевым инструментом, упрощающим работу с методологией, является бизнес-гlossарий. В соответствии со «Сводом знаний по управлению данными», бизнес-гlossарий содержит в себе согласованные определения бизнес-терминов и описывает их связь с данными. Он помогает бизнесу иметь единый источник сведений о методологии, а также аккумулирует в себе знания экспертов предметной области. Это может быть полезно при онбординге новых сотрудников, но не только.

Наибольшую ценность бизнес-гlossарий приносит на стыке деятельности бизнес- и ИТ-подразделений, так как позволяет в явном виде связать понятия, которыми оперирует бизнес, и их техническое представление. Например, в гlossарии могут быть описаны ключевые метрики (ответственность бизнес-подразделений) и SQL-запросы, с помощью которых эти метрики вычисляются техническими командами.

Кроме того, единый гlossарий упрощает взаимодействие бизнес-подразделений. Если основным источником методологии в организации являются доменные эксперты или нормативные документы, то поиск информации занимает много времени. Гlossарий позволяет сделать эти процессы прозрачными и (при правильной организации работы) более эффективными. Их важность сложно переоценить: нам даже известны случаи, когда подразделение неожиданно для себя проваливало КПЭ, так как контролирующая инстанция использовала другие правила округления.

Подытожим: люди — не самый надежный способ хранения и передачи информации, а документы — не самый доступный. С помощью бизнес-гlossария можно избежать «сломанного телефона» и существенно упростить взаимодействие в компании.

## Как обычно внедряют глоссарий: стандартные проблемы

На рынке существует много решений для ведения бизнес-глоссария, но часто их внедрение заканчивается разочарованием. Стандартный процесс в таких ситуациях выглядит следующим образом:

- Выбирают решение для ведения бизнес-глоссария.
- Производственным подразделениям в директивном порядке поступает указание по набору в бизнес-глоссарий понятий и определений, за которые они ответственны — например, используемых на дашбордах или соответствующих системах-источниках.
- Далее начинается наполнение глоссария. Часто без проработки концепции того, что мы считаем бизнес-термином, так как это принимается за плюс-минус «интуитивно очевидное»: разве бизнес-подразделения не знают, какие понятия они используют? Вот они и должны оказаться в глоссарии.
- В отсутствие понимания процесса связи определений глоссария с данными, обновления и согласования терминов, данная работа оседает в глоссарии мертвым грузом, а также стремительно устаревает.
- Энтузиазм организации в части внедрения DG исчезает, а бизнес-ценность проекта оказывается по итогу нулевой или отрицательной (учитывая трудозатраты на наполнение глоссария).

Может ли результат выглядеть по-другому? Наш опыт в области управления данными показывает, что может; однако, успех зависит от правильной постановки целей. Во многих случаях дата-офис изначально ставит цели, достижение которых не приближает компанию к ожидаемым результатам.

## **Четыре мифа об управлении данными и почему они не работают**

Далее я приведу примеры неверных представлений о том, как должна быть организована работа по управлению данными и на чем имеет смысл сфокусироваться вместо этого.

### **Миф № 1. Самое важное — это правильно выбрать инструмент**

**На самом деле: для управления данными процессы важнее инструментов.**

На начальном этапе для управления данными достаточно всего двух инструментов: текстового редактора для того, чтобы писать регламенты, а также Excel (или аналога) для всего остального.

Потребность во внедрении DG может появиться у организации по разным причинам, и это может повлиять на процессы внедрения глоссария. Например, для организаций в некоторых секторах характерно наличие нормативно утвержденных методологий расчета показателей; соответственно, жизненный цикл бизнес-терминов оказывается связанным с документооборотом организации, а соответствующие системы должны быть

интегрированы. Для других заказчиков таких требований может и не быть, зато для них критически важно отражать связь показателей с метаданными хранилища, и, соответственно, набор желаемых функций бизнес-гlossария будет скорее ближе к дата-каталогу или MDM.

В любом случае, эти организационные особенности довольно сложно определить до начала пилотного проекта, а скоринг доступных продуктов скорее сделает картину более размытой, чем, наоборот, внесет ясность. Необходимо подбирать функции продуктов, исходя из реальных потребностей.

Поэтому вместо скоринга продуктов на начальной стадии имеет смысл определиться с процессными вопросами:

- Какие у нас есть нормативные источники методологии?
- Кто сейчас отвечает за методологию? Кто должен отвечать за методологию?
- Где используются те или иные показатели? Кто их формирует? Кто их использует? Где они хранятся?
- Как часто устаревает информация?
- Кому сейчас не хватает информации о методологии?
- Как обеспечено распространение информации о методологии?

Ответы на эти вопросы позволят вам определить требования к целевому процессу управления данными, а также к самому гlossарию. Таким образом можно понять, какие подразделения более заинтересованы во внедрении

гlossария и какой именно информации им не хватает. Это определит требования к самому гlossарию и к методологии. О ней дальше.

## **Миф № 2. Главное — начать описывать бизнес-термины**

**На самом деле: следует заранее определиться с методологией.**

Как я отмечал выше, существует множество способов описать бизнес-термины, формирующие предметную область организации, особенно если эта организация большая и бизнес-процессов также много.

В связи с этим следует уже на начальном этапе договориться о том, что мы считаем бизнес-термином, как мы определяем бизнес-термин. Например, что бизнес-термин — это любой показатель дашборда, или бизнес-термин — только то, что нормативно закреплено в документах, или бизнес-термин — просто отдельное значение бизнес-ключа в таблице с показателями.

Кроме того, необходимо понимать следующее:

- Какими атрибутами может обладать бизнес-термин.
- Как определяется его формула расчета.

- Что является источником методологии по данному бизнес-термину (а также, кто принимает решение в случае отсутствия таких источников).
- Наконец — кто является ответственным за данный бизнес-термин.

В отсутствие этих критериев существует риск, что ответственные принесут принципиально различную и несопоставимую информацию.

## **Миф № 3. Необходимо описать как можно больше бизнес-терминов**

**На самом деле: процесс управления данными необходимо выстраивать итерационно.**

Предположим, что мы провели исследование и хотим выстроить пилотный процесс. Что мы должны делать теперь? С чего нам начать?

Для определения бизнес-терминов есть три основных подхода. В первом случае мы идем от сущностей, используемых в BI-системах, во втором — от метаданных хранилища, в третьих — от нормативных требований и документов, которые используются в работе.

Нет особой разницы в том, с чего мы начинаем, в каждом отдельном случае есть своя бизнес-ценность и свои сложности. Например, если начать с описания бизнес-терминов нормативных документов, можно столкнуться с

последствиями, так как окажется, что нормативное описание расходится с тем, как в действительности рассчитываются или описываются показатели.

Гораздо важнее выстраивать управляемый процесс постепенных улучшений. Бизнес-гlossарий находится на пересечении интересов бизнес-пользователей и технических команд, так как позволяет перекинуть мостик между бизнес-показателями и их технической реализацией. Но в то же время не стоит подключать к работе над glossарием сразу всех, кто работает с данными (от бизнес-подразделений до BI или КХД). Если подключить сразу всех заинтересованных, то большая часть не увидит той информации, которую хотелось бы иметь. В этой ситуации дата-офис должен четко понимать, в каком порядке реализовывать приоритеты; подключая всех сразу, можно попросту «потеряться» в различных контекстах.

Например, одна крупная компания реализует проект по внедрению практик управления данными. Отчетность у них жестко регламентирована, при этом в ландшафт входит более 100 систем-источников, а также более 500 различных дашбордов. По исходным оценкам, только верхнеуровневых бизнес-сущностей в компании было несколько тысяч.

Амбициозное желание бизнеса — создать среду «абсолютной видимости» для данных, в которой для каждого показателя доступна информация из сразу нескольких контекстов: нормативные основания его расчета; системы, в которых показатель формируется; формы, на которых он используется; ответственные подразделения; связанные документы.

В этот процесс, таким образом, оказываются вовлечены самые разные по своей специализации агенты; для каждого из них бизнес-термин, его атрибуты, представляют собой нечто свое. Так, например, для методологов формулой расчета показателя будет считаться бизнес-описание того, как показатель связан с другими показателями. Для пользователей из числа работающих над хранилищем, формула расчета же — это набор конкретных процедур над таблицами, ссылающихся на конкретные поля, названия которых ничего не скажут методологам. Информация, жизненно важная для одних, оказывается абсолютно бессмысленной другим, и с этим ничего не сделать.

Соответственно, ожидания от того, что хотелось видеть в глоссарии, отличаются у разных категорий пользователей.

Важно иметь в виду, что на проектах такого масштаба, скорее всего, мы не удовлетворим всех пользователей сразу; более того, нет смысла подключать их к работе одновременно: у дата-офиса не хватит capacity для того, чтобы вместить в себя столько различных контекстов. Гораздо важнее выстроить итеративный процесс, когда на каждом этапе мы получаем определенную, пусть и небольшую ценность, и постепенно подключаем новых пользователей, адаптируя результат под их потребности.

В приведенном выше примере компания начала с описания в бизнес-глоссарии моделей данных отчетных форм и дашбордов: как есть перенесли всю ту информацию, которая уже была. Речь здесь идет о существенном объеме данных, модели данных велись в Excel, совокупное число строк составило около 40 000. На этом этапе было важным не застрять на валидации данных, так как, во-первых, даже проверенная информация могла

достаточно быстро начать устаревать, а, во-вторых — в процессе валидации возникали методологические вопросы, решение которых отнимало много времени, большинство из которых были связаны со справочниками.

Один из типичных вопросов: в утвержденных формах отчетности определенный показатель измеряется в штуках, в системе-источнике — в единицах, визуализируется на дашбордах — в тысячах штук. Один ли это показатель? С точки зрения бизнес-пользователей — очевидно, что один; для технических пользователей это, разумеется, не так: для того, чтобы получить значение другого показателя, его необходимо преобразовать. Логика, которой мы будем придерживаться в данном случае, будет определять наполнение глоссария.

По итогам первого этапа было получено описание бизнес-сущностей, однако к нему оставалось много вопросов: где-то информация была устаревшей или неполной; в каких-то случаях из-за отсутствия достаточной базы показатели были заведены дважды, или даже трижды. Тем не менее, поставленная цель была достигнута, а именно — обеспечена централизация информации о показателях в одной системе, вместо ее распределения по многочисленным эксель-таблицам.

Второй этап — это непосредственно организация использования глоссария бизнес-пользователями, тех, кто является (или будет являться) владельцами данных. Основная задача здесь — это установить процесс, в рамках которого владельцы данных начинают отвечать за содержание, корректность описания, а также использование бизнес-терминов на дашбордах. Это было одним из самых главных требований — в глоссарии должна храниться не только информация о самих бизнес-терминах, но и быть отражена связь с

визуализацией.

Ключевая черта данного процесса — это правильное целеполагание; важно с одной стороны не застрять на промежуточном этапе, с другой — регулярно поставлять ценность для компании. Для этого необходимо выделить ключевые приоритеты каждого этапа, а также те подразделения, которые будут подключены. На первом этапе имеет смысл наполнить глоссарий хотя бы той информацией, которая уже есть; на втором этапе — подключать первых бизнес-пользователей и устанавливать процессы владения данными и использования данных, на третьем этапе — подключать технических пользователей и анализировать использование данных на уровне хранилища.

Альтернативный подход — это сделать модельное описание одной небольшой предметной области, одного домена и потом двигаться экстенсивно. Стоит иметь в виду, однако, что при анализе предметной области всплывают технические и концептуальные вопросы, решение которых требует значительно больше времени, чем закладывалось изначально; поэтому избравшим этот путь рекомендуем сразу готовить ответы на вопросы типа «почему прошло полгода, а до сих пор описаны только продажи?»

## Миф № 4. Чем больше пользователи работают в системе — тем лучше

**На самом деле: однотипные задачи должны быть автоматизированы.**

Аналитические и методологические работы при внедрении практик управления данными занимают очень много времени. Поэтому важно уметь автоматизировать все, что может быть автоматизировано, так как большое количество однотипных задач поглотит все ресурсы дата-офиса.

Один из примеров — это непосредственно ручная работа по наполнению глоссария. Во многих случаях у бизнес-подразделений есть наработки, которые могут быть использованы для наполнения бизнес-глоссария. Это могут быть эксель-таблицы с описанием и атрибутами показателей или стандартные отчетные таблицы. В большинстве случаев эти материалы будут основой формирования глоссария. Значительный ресурс может быть сэкономлен, если заранее обеспечить автоматизацию процессов переноса этой информации в глоссарий: с одной стороны, обеспечив в глоссарии функцию импорта структурированных файлов, с другой стороны — подготовив скрипты для валидации, дедупликации и приведения сведений к нужному атрибутивному составу.

## Заключение

Бизнес-глоссарий — это важный инструмент, который позволяет упростить понимание между различными бизнес-подразделениями организации, а

также между бизнесом и техническими специалистами.

Мы рекомендуем начинать внедрение бизнес-гlossария только после предварительной проработки целевого процесса, определения тех, кто «владеет» методологией и тех, кто испытывает в ней потребность. Должны быть понятны кейсы использования glossария, а также список необходимых пользователям сведений: правила расчета, нормативные основания, владельцы данных или их местоположение. Чем более подробно будет описан процесс ведения glossария, методология описания, тем лучше.

В то же время, надо быть готовым к тому, что даже самая качественная методология не будет покрывать всех юзкейсов. Поэтому логично двигаться постепенно, учитывая фидбэк пользователей. Это проще контролировать при постепенном подключении производственных подразделений.

Наконец, следует помнить, что основная задача glossария — это упрощение и развитие коммуникации внутри компании. Именно качество коммуникации, а не само внедрение информационной системы, должно быть ключевой метрикой успеха проекта.