



УДК 004.03

DOI: 10.31799/2007-5687-2020-4-13-20

РОЛЬ MDM-СИСТЕМЫ В ДЕЯТЕЛЬНОСТИ ОРГАНИЗАЦИИ И ТИПЫ ДАННЫХ, ИСПОЛЬЗУЮЩИЕСЯ В MDM-ПРОЕКТАХ

В. Д. Шарипова

Санкт-Петербургский университет аэрокосмического приборостроения

В статье рассказывается о том, какую роль играет управление средой MDM в BI-проекте и в деятельности организации в целом. Представляется обзор типов данных, которые часто встречаются в проектах внедрения MDM-систем. Объясняется, почему мастер-данные, справочные и транзакционные данные занимают критическое положение в подобных проектах.

Ключевые слова: BI, MDM-система, MDM-проект, типы данных, HСИ, мастер-данные, транзакционные данные.

Для цитирования:

Шарипова В. Д. Роль MDM-системы в деятельности организации и типы данных, используемые в MDM-проектах // Системный анализ и логистика: журнал.: выпуск №4(26), ISSN 2077-5687. – СПб.: ГВАП., 2020 – с. 13-20. РИНЦ.

THE ROLE OF MDM-SYSTEM IN THE ACTIVITIES OF ORGANISATION AND THE DATA TYPES USED IN MDM-PROJECTS

V. D. Sharipova

Saint-Peterburg State University of Aerospace Instrumentation

This article discusses the role of managing the MDM in a BI-project and in the activities of an organization as a whole. Provides an overview of the data types that are often found in MDM implementation projects. Explains why master-data, reference data, and transactional data are critical in such projects.

Key words: BI, MDM-system, MDM-project, types of data, reference data, master-data, transactional data.

For citation:

Sharipova V. D. The role of MDM-system in the activities of organization and the data types used in MDM-projects // System analysis and logistics.: №4(26), ISSN 2077-5687. – Russia, Saint-Petersburg.: SUAI., 2020 – p. 13-20.

Необходимость внедрения MDM-систем

Переход на инновационный путь развития – это магистральное направление модернизации экономики. Более того, доказано, что в основе любого предпринимательства должна лежать систематическая инновационная деятельность [1]. И важным элементом мотивационного механизма, побуждающего предприятия к инновационной активности, призван стать бизнес-анализ. Его целевая направленность заключается в информационном обеспечении эффективного корпоративного управления, которое понимается как система взаимодействия между акционерами и менеджментом компании, а также с другими заинтересованными сторонами с целью реализации интересов собственников и законных прав остальных стейкхолдеров [2]. Поэтому неудивительно, что на сегодняшний день системы класса Business Intelligence (BI) стали настоящим must have для любого бизнеса вне зависимости от сферы деятельности и отрасли.

Согласно исследованию, проведенному аналитиками компании Gartner, около 70-80% реализованных проектов внедрения систем класса BI являются провальными [3]. Статистика, конечно, имеет под собой основу. Неуспешность проектов объясняется рядом факторов, которые оказывают непосредственное влияние на весь процесс. Здесь помимо общих факторов вроде необходимости исхода инициативы внедрения системы от топ-менеджмента компании, необходимости единого понимания цели внедрения и т.п. отдельное внимание уделяется другому немаловажному условию успешного BI-проекта: наличию достоверных и качественных данных в



учётной системе [4]. Процесс контроля качества должен производиться как можно ближе к исходным данным и непрерывно производиться на всех этапах сбора и обработки информации.

Напрашивается вопрос: что, если объединить проекты мастер-данных и программ бизнес-аналитики? Ведь успех программы бизнес-аналитики зависит от того, как организовано управление средой MDM. При правильно поставленной и решённой задаче нужные данные (чёткие, точные и однородные) могут быть использованы для анализа и принятия решения. Это касается не только для BI или BA (Business Analysis), ведь критически важно, чтобы информация об основных данных (мастер-данных) была верной, поскольку небольшие ошибки в ней могут привести к серьёзным последствиям при использовании в транзакциях [5].

В этой статье постараемся рассмотреть, что такое основные данные (мастер-данные) и почему важно различать типы данных в MDM-проектах. Стоит сразу отметить, что приводимая в данной работе категоризация данных относится преимущественно к области управления данными и к области проектов внедрения MDM-систем – в частности. Т.е. проектов создания в организации IT-системы, поддерживающей процессы управления мастер-данными (Master Data Management – MDM).

Идея MDM

Основной задачей MDM-системы является создание и поддержание достоверной, актуальной непротиворечивой информации, которая часто описывается термином «золотая запись» или «эталонная запись» («Single Version of The Truth», «golden record»). Эталонная запись – очищенная (т.е. избавленная от шума, приведённая к стандартному виду), дедуплицированная (т.е. не имеющая дублей, уникальная) и консолидированная (т.е. объединённая из нескольких версий) версия мастер-данных. Соответственно, эталонные данные имеют большую ценность для приложений в BI (Business Intelligence), BA (Business Analysis), операционных или других.

Обычно архитектура MDM-проекта представляет собой схему типа «звезда», в центре которой находится MDM-система, к которой подключены IT-системы организации (рис. 1).

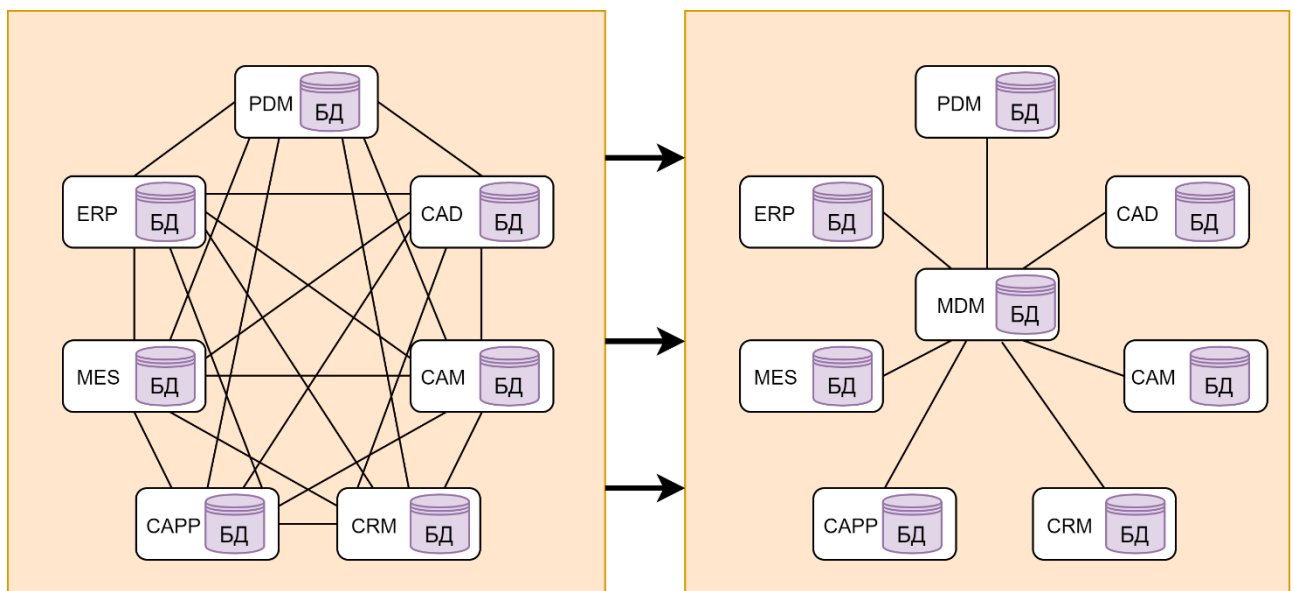


Рис. 1. Переход к модели типа «звезда», использующийся в MDM-проектах

Типы данных

В MDM-проектах особенно заметно наличие различных типов данных. Это связано с несколькими моментами:



- MDM-проект имеет распределенную структуру, потому что данные содержатся в различных системах;
- центром проекта является новая модель данных MDM;
- стоит задача обмена данными в гетерогенных IT-системах, имеющих собственные модели данных.
- характеристики интегрируемых данных зачастую совершенно различны [6].

Существует три основных типа данных, использующихся в MDM-проектах: справочные, мастер-данные и транзакционные данные.

1. Справочные (ссылочные) данные (НСИ – нормативно справочная информация) – базовые данные для всех информационных систем, часто представляющие собой нормативные документы, международные стандарты, сокращения, словари и пр. Их также называют референс-Типичные Ссылочные Данные – справочники счетов, тематические классификаторы. Справочные данные имеют также синонимы «Reference Data», «Lookup Data», «Dictionaries».

Это данные с низкой волатильностью, которые обычно формируются за пределами организации, но актуальны для деятельности организации. Эти данные обычно содержат классификаторы и их описания (например, КЛАДР, ОКП, ОКВЭД, коды валют). Управление справочными данными как специальным типом мастер-данных, с использованием функциональности концентратора управления мастер-данных, общепризнанно лучшей практикой.

2. Мастер-данные (Master data) – основные данные предприятия, основной актив, представляющие собой ключевые объекты деятельности предприятия. Типичные Мастер Данные – информация о клиентах, поставщиках, номенклатуре. Мастер-данные используются во всём предприятии, а иногда и между предприятиями для упрощения обмена информацией.
3. Транзакционные данные (Transactional data) – данные, отображающие результат выполнения транзакции и относящийся непосредственно к бизнес-операции. Типичный пример – банковская операция, счет за использование услуг, накладная к отгрузке и пр. Часто называются также как «Transactional Data», «Application Specific Data», «Operational Data».

Одним из критериев, по которому можно различить типы – это их жизненный цикл. Упрощенная временная диаграмма жизненного цикла для каждого типа приведена на рисунке 2.

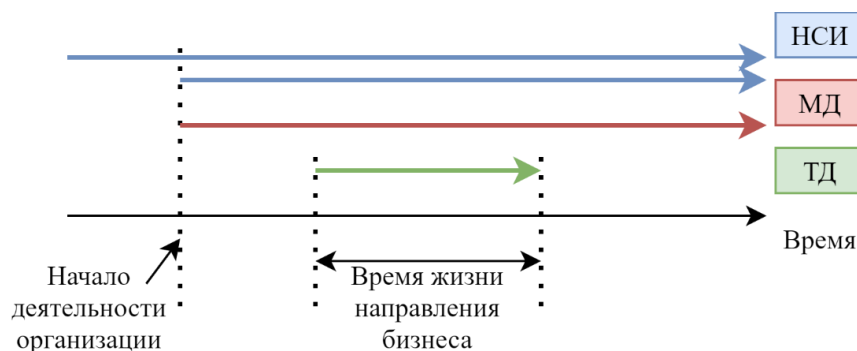


Рис. 2. Диаграмма жизненного цикла для каждого типа

Рассмотрим распределение данных на условном примере. Допустим, пассажир АВС приобрёл билеты на несколько рейсов, чтобы добраться из пункта назначения А в пункт прибытия В на определённый день и соответственно, на определённое время. Пример описывает бизнес-процесс авиакомпании, которая продала билеты, т. е. данная операция является транзакцией. В таблице 1 приведено распределение данных.



Таблица 1 – Пример распределения данных

Тип данных	Атрибут 1	Атрибут 2	Атрибут 3
Транзакция	Код пассажира	Код вылета аэропорта	Код аэропорта прибытия
Мастер-данные	Код пассажира	ФИО	Паспортные данные
НСИ	Код аэропорта	Название аэропорта	Страна

Перечисленные типы используются в MDM-проектах, однако это не обязательно означает то, что все типы данных будут использоваться непосредственно в MDM-системе. Набор используемых типов зависит от выбранного стиля внедрения: реестрового (registry), сосуществующего (coexistence) или транзакционного (transactional). Например, при сосуществующем стиле внедрения в MDM-систему «подтягиваются» только НСИ и мастер-данные, а остальные вводятся распределённо в нескольких других источниках.

На практике, выбор той или иной стратегии внедрения MDM определяется многими факторами: целями предприятия в области управления мастер-данными, степенью зрелости предприятия, степенью готовности IT-инфраструктуры, наличием инвестиций на реализацию проекта и многими другими параметрами [7].

Наряду с вышеописанными типами данных в MDM-проектах используются и другие. Например, метаданные, исторические, неструктурированные данные и проч. Они могут реализовываться в виде внутренних справочников, таблицы с настройками, отчетных таблиц и др.

Исторические данные (Historical data) – это данные, которые включают в себя исторические транзакционные и мастер-данные. Чаще всего такие данные аккумулируются в ODS и DWH системах и служат для решения различных аналитических задач и поддержки принятия управленческих решений [7]. Изменения и обновления данных записываются при помощи сервисов аудита/ведения журналов (логирования). Соответственно, на основе исторических данных строятся Аналитические данные (Analytical Data). Примером аналитических данных в описанном случае может являться общая стоимость всех приобретённых билетов.

Метаданные (Metadata) — это данные о других данных, как правило, о данных Домена, такие как каталоги баз данных или XML схемы. Это описание может включать информацию как о структуре данных, так и об отдельных записях или файлах. Метаданные также используются для проверки, очистки, поиска, сопоставления данных.

На рисунке 3 представлена возможная высокоуровневая архитектура нескольких типов данных в организации.

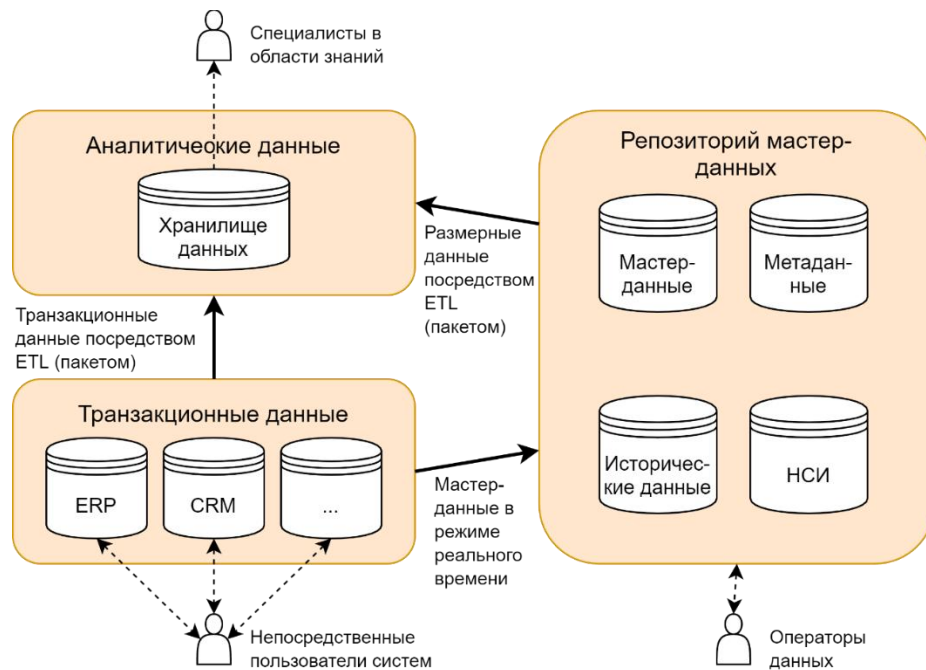


Рис. 3. Возможная высокоуровневая архитектура нескольких типов данных в организации

Преобразование типов данных друг в друга

Из примера выше можно заметить, что одни и те же данные (например, код пассажира) могут относиться в одно и то же время к разным типам, в зависимости от вариантов их использования. Например, справочник «Общероссийский классификатор территорий муниципальных образований» (ОКТМО) ведется централизованно РОССТАНДАРТОм России. Для РОССТАНДАРТа работа со справочником ОКТМО является одним из основных направлений деятельности. Таким образом, для внутренних систем РОССТАНДАРТа справочник ОКТМО является мастер-данными, т.к. только внутри этой организации производится сбор, обработка, согласование и в конечном счете формирование единого справочника ОКТМО. Для любой другой организации, импортировавшей данный справочник и использующей его в своей работе, ОКТМО является ссылочными данными (НСИ) и используется в качестве справочной информации. Соответственно, типы данных взаимосвязаны друг с другом в различных формах.

Транзакционные данные чаще всего расположены в своих системах, Мастер данные консолидированы в МДМ системе, частично в Транзакционных системах, Ссылочные данные растражированы по IT системам в виде полных копий.

Критерии для идентификации типов данных

Далее приведем сравнение типов данных по различным критериям согласно IBM Developer (табл. 2). Эти критерии были выбраны как наиболее отличительные с целью показать характерные и уникальные характеристики каждого типа данных. Список не претендует на полноту и может быть расширен для каждого конкретного проекта [4].

Таблица 2 – Сравнительная характеристика типов данных по различным критериям

№	Критерий	Ссылочные данные (НСИ)	Мастер-данные	Транзакционные данные
1	Типичные данные	Справочники, нормативы, сокращения, кодификаторы, акронимы, стандарты, словари, «бизнес-правила» в виде диапазонов, пороговых значений, констант,	Основные объекты организации, прошедшие процедуры «очистки» и	Данные, отражающие результат транзакции в системе. Часто представлены комбинацией Мастер-



		последовательностей и пр. Могут использоваться для нормализации в БД. Часто являются связующим звеном между мастер-данными и транзакционными. Выражают «факт». Самодостаточны.	согласования. Определяются типом МДМ системы (CDI, PIM). Выражают «базовые объекты».	данных и НСИ и дополнительных специфичных для данной транзакции атрибутов. Выражают «действие-результат».
2	Скорость изменения модели данных	Модель данных практически стабильна на протяжении использования.	Модель данных меняется редко и, в основном, как ответ на адаптацию к новым бизнес-требованиям и расширению моделей в транзакционных системах в части, относящейся к основным данным предприятия.	Модель данных меняется часто в зависимости от новых бизнес-требований к системе.
3	Скорость изменения данных	Очень низкая. Практически не изменяются.	Низкая, данные меняются в момент изменения мастер-данных в одной из ИТ-систем (или MDM-системе) и формирования «золотой записи».	Высокая. Меняются в каждой бизнес-транзакции.
4	Критичность к качеству данных	Очень высокие требования к качеству данных. Недопустима путаница, разночтения и дублирование данных.	Высокие требования к качеству.	Высокие требования. Качество данных определяется транзакционной системой, в которой, как правило, заданы механизмы "компенсации" ошибок.
5	Время жизни	В случае внешних для организации данных - существовали по факту до начала создания/эксплуатации ИТ-систем организации. В случае внутренних данных - существуют все время существования бизнеса предприятия.	Появляются и поддерживаются на протяжении всего времени существования бизнеса предприятия.	Чаще всего время жизни определяется временем жизни соответствующей транзакционной системы.
6	Количество данных	Небольшое, часто отражает реальные "факты", количество которых заведомо конечно.	Изначально достаточно большое. Конечное множество.	Количество данных не ограничено.
7	Взаимосвязь с другими типами данных	Не включают в себя и не ссылаются ни на какие другие типы данных.	Включают в себя или ссылаются на Ссылочные данные	Включают в себя или ссылаются на любые другие типы данных.
8	Типичные способы	Распространяется централизованно и однонаправленно - чаще списками,	Диктуется "мощностью" интеграции между	Определяется спецификой транзакционной



	распространения	файлами, выгрузками. Реже - репликацией между БД.	MDM и транзакционной системой. В общем случае должны быть построены двунаправленные каналы связи, работающие в режиме "онлайн" и "офлайн".	системы. Возможно использование всех типов интеграции.
9	Места появления данных внутри Предприятия.	- Транзакционные системы. - MDM-система. - Часто "прошиваются в код" исполняемых модулей ИТ-систем. - Глоссарий Предприятия.	- МДМ система. - Транзакционные системы.	Транзакционные системы
10	Особенности модели	Объекты с небольшим количеством атрибутов. Неглубокий уровень вложенности в классификаторах. Наличие натуральных ключей. Не развитые механизмы поддержки историчности.	Объекты с большим количеством атрибутов и сложными взаимосвязями. Развитые механизмы для хранения истории изменений.	Определяется спецификой транзакционной системы. В общем случае модель содержит много объектов с различным количеством атрибутов.

Понимание типов данных и их характеристик требуется для:

- проектирования интеграционных взаимодействий в MDM-проекте;
- проектирования потоков данных внутри ландшафта ИТ;
- определения границ жизненного цикла с целью поддержки актуальности данных для избежания необоснованных затрат по обслуживанию;
- разделения зон ответственности при контроле состояния;
- поддержания целостности и непротиворечивости основной модели данных предприятия, консолидирующей все типы данных в единой представлении;
- адекватной утилизации ИТ-систем, работающих с данными;
- помощи в расчете экономических параметров, созданию методик оценки стоимости хранения и обслуживания одного объекта в МДМ системе, стоимости транзакции с участием МДМ и пр.;
- создания внутренних архитектурных стандартов и принципов, принятых и используемых в организации для поддержки и развития ИТ. Качественная поддержка инициатив Data Governance, Enterprise Information Management и пр. [4].

Заключение

Управление основными данными позволяет организациям улучшать качество и согласованность ключевых активов, включая данные клиентов, географические данные, информацию о продуктах и данные об активах. Фактически, MDM-система становится для компаний, которые её внедряют, основным источником данных для других, не менее важных источников. В результате это приводит к возможности выполнения более качественного анализа и оценки затрат организации, её эффективности, оптимизации существующих процессов, прогнозирования дальнейшего развития, создания отчётности, а также к снижению избыточности



прогнозирования данных, их неструктурированного хранения и др.

При внедрении MDM-системы этапы анализа, создания модели данных и, соответственно, определения типов используемых данных должна уделяться большая часть внимания, ведь от их успешной реализации зависит успех проекта MDM в целом и эффективность его работы в будущем.

СПИСОК ЛИТЕРАТУРЫ

1. Друкер П. Классические работы по менеджменту. М.: "Альпина Бизнес Букс", 2008
2. Пенс И. Ш., Фурс С. А. Корпоративное управление в современной промышленности России: состояние и факторы совершенствования [Электронный ресурс]. — М.: ООО «Недра коммюникейшнс ЛТД», 2008. — 120 с. — URL: http://www.mirkin.ru/_docs/book050.pdf (дата обращения: 14.11.2020 г.).
3. Wayne Kernochan. Why Most Business Intelligence Projects Fail [Электронный ресурс] // enterprise APPS today, 2011. — URL: <http://www.enterpriseappstoday.com/business-intelligence/why-most-business-intelligence-projects-fail-1.html> (дата обращения: 14.11.2020 г.).
4. Ключек М. С., Парфенова А. С. Функциональность MDM-решений [Электронный ресурс] // Международный научный журнал «Инновационное развитие» : №1(18), январь, 2018. — URL: https://www.elibrary.ru/download/elibrary_35351259_81086709.pdf (дата обращения: 14.11.2020 г.).
5. April Reeve. Managing Data in Motion [Электронный ресурс] // ScienceDirect : Chapter 15 - Master Data Management, p. 105-111. — URL: <https://www.sciencedirect.com/science/article/pii/B9780123971678000157> (дата обращения: 14.11.2020 г.).
6. Карпов А. Типы данных в МДМ проектах [Электронный ресурс] // IBM Developer : 08.02.2012. — URL: <https://www.ibm.com/developerworks/ru/library/dm-karpov/mdm/index.html> (дата обращения: 14.11.2020 г.).
7. Власов М. Что такое «система управления мастер-данными» и зачем она нужна [Электронный ресурс] // DATAREON : март, 2017. — URL: https://www.datareon.ru/know/press/detail_47098/ (дата обращения: 14.11.2020 г.).

ИНФОРМАЦИЯ ОБ АВТОРЕ

Шарипова Виктория Дамировна –

студент кафедры системного анализа и логистики

Санкт-Петербургский государственный университет аэрокосмического приборостроения

190000, Санкт-Петербург, ул. Большая Морская, д. 67, лит. А

E-mail: vika_sharipova@mail.ru

INFORMATION ABOUT THE AUTHOR

Sharipova Viktoria Damirovna –

student of system analysis and logistics department

Saint-Petersburg State University of Aerospace Instrumentation

67, Bolshaya Morskaya str., Saint-Petersburg, 190000, Russia

E-mail: vika_sharipova@mail.ru