



Universidades Lusíada

Augusto, Nascimento Joaquim, 1957-

Implementação de um sistema de MDM como meio de integração de aplicações

<http://hdl.handle.net/11067/1757>

Metadados

Data de Publicação	2015-12-10
Resumo	As empresas que pretendem tirar proveito dos dados que possuem para aumentar a eficiência competitiva do negócio podem investir em tecnologias de integração de sistemas, mas os objectivos, os conceitos e a terminologia das soluções disponíveis no mercado podem ser confusos. Tecnologias como Data Warehouse (DW), Service-Oriented Architecture (SOA), Enterprise Application Integration (AEI) e Master Data Management (MDM) são alternativas de integração de aplicações (Butler, 2011, p. 6) mas têm pro...
Palavras Chave	Bases de dados - Gestão, Relação com o cliente - Gestão, Sistemas de informação para a gestão
Tipo	masterThesis
Revisão de Pares	Não
Coleções	[ULL-FCEE] Dissertações

Esta página foi gerada automaticamente em 2024-05-01T06:07:17Z com informação proveniente do Repositório



UNIVERSIDADE LUSÍADA DE LISBOA

Faculdade de Ciências da Economia e da Empresa

Mestrado em Ciências da Computação

Implementação de um sistema de MDM como meio de integração de aplicações

Realizado por:

Nascimento Joaquim Augusto

Orientado por:

Prof. Doutor Marco Bruno Correia Costa

Constituição do Júri:

Presidente:	Prof. Doutor Mário Caldeira Dias
Orientador:	Prof. Doutor Marco Bruno Correia Costa
Arguente:	Prof. Doutor Paulo Jorge Gonçalves Pinto

Dissertação aprovada em: 10 de Dezembro de 2015

Lisboa

2015



UNIVERSIDADE LUSÍADA DE LISBOA
Faculdade de Ciências da Economia e da Empresa
Mestrado em Ciências da Computação

Implementação de um sistema de MDM como meio
de integração de aplicações

Nascimento Joaquim Augusto

Lisboa

Maio 2015



UNIVERSIDADE LUSÍADA DE LISBOA
Faculdade de Ciências da Economia e da Empresa
Mestrado em Ciências da Computação

Implementação de um sistema de MDM como meio
de integração de aplicações

Nascimento Joaquim Augusto

Lisboa

Maio 2015

Nascimento Joaquim Augusto

Implementação de um sistema de MDM como meio de integração de aplicações

Dissertação apresentada à Faculdade de Ciências da
Economia e da Empresa da Universidade Lusíada de
Lisboa para a obtenção do grau de Mestre em Ciências
da Computação.

Orientador: Prof. Doutor Marco Bruno Correia Costa

Lisboa

Maio 2015

Ficha Técnica

Autor Nascimento Joaquim Augusto
Orientador Prof. Doutor Marco Bruno Correia Costa
Título Implementação de um sistema de MDM como meio de integração de aplicações
Local Lisboa
Ano 2015

Mediateca da Universidade Lusíada de Lisboa - Catalogação na Publicação

AUGUSTO, Nascimento Joaquim, 1957-

Implementação de um sistema de MDM como meio de integração de aplicações / Nascimento Joaquim Augusto; orientado por Marco Bruno Correia Costa. - Lisboa: [s.n.], 2015. - Dissertação de Mestrado em Ciências da Computação, Faculdade de Ciências da Economia e da Empresa da Universidade Lusíada de Lisboa.

I - COSTA, Marco Bruno Correia, 1971-

LCSH

1. Bases de dados - Gestão
2. Relação com o cliente - Gestão
3. Sistemas de informação para a Gestão
4. Universidade Lusíada de Lisboa. Faculdade de Ciências da Economia e da Empresa - Teses
5. Teses - Portugal - Lisboa

1. Database management
2. Customer relations - Management
3. Management information systems
4. Universidade Lusíada de Lisboa. Faculdade de Ciências da Economia e da Empresa - Dissertations
5. Dissertations, Academic - Portugal - Lisbon

LCC

1. QA76.9.D3 A94 2015

AGRADECIMENTOS

À universidade Lusíada de Lisboa, pela oportunidade.

Ao Professor Marco Bruno Costa, pela disponibilidade, saber e orientação.

À minha querida mãe, pelo exemplo de persistência.

Aos meus filhos Beatriz, Júlio e Gabriela, pelo carinho.

À Fátima Silveira e à sua filha Syndy, pelo incentivo.

Ao meu colega Luís Damas, pelo imprescindível contributo.

Aos meus irmãos Pascoalina e Tony, pelo apoio.

Aos meus sobrinhos Graça e Augusto Pupulo, pelo carinho.

Ao meu saudoso Pai, amigo e companheiro, Augusto Joaquim Taco.

APRESENTAÇÃO

As empresas que pretendem tirar proveito dos dados que possuem para aumentar a eficiência competitiva do negócio podem investir em tecnologias de integração de sistemas, mas os objectivos, os conceitos e a terminologia das soluções disponíveis no mercado podem ser confusos.

Tecnologias como *Data Warehouse (DW)*, *Service-Oriented Architecture (SOA)*, *Enterprise Application Integration (AEI)* e *Master Data Management (MDM)* são alternativas de integração de aplicações (Butler, 2011, p. 6) mas têm propósitos diferentes conforme iremos ver ao longo deste trabalho.

Em ambientes com múltiplas aplicações, as pesquisas como: “qual é a morada correcta de determinado cliente?”, “que produtos o cliente adquiriu?” ou “quantos clientes tem a empresa?”, requerem uma visão íntegra e holística dos dados.

MDM é uma disciplina tecnológica de negócio que ajuda as organizações a alcançarem uma “visão única da verdade” dos dados em áreas de domínio importantes como clientes, produtos e contas (Cervo e Allen, 2011, p. 395).

Existem soluções de MDM específicas para dados de clientes e produtos. As soluções para os dados de clientes são designadas por *Customer Data Integration (CDI)*. As que se focalizam na gestão dos dados de produtos designam-se por PIM - *Product Information Management* (Dreibelbis et al, 2008, p. 665). Os sistemas de MDM que suportam múltiplos domínios de dados, variados estilos de implementação e diversos métodos de utilização designam-se por *MDM Multi-Form* (Dreibelbis et al, 2008, p. 665).

Este trabalho tem como objectivo o estudo do estado da arte do MDM e, como prova dos conceitos, apresentar uma implementação de um sistema de MDM num organismo público.

Palavras-chave: *Master Data Management, MDM, Master Data, Data Governance, Data Quality.*

PRESENTATION

Companies seeking to take advantage of data to increase competitive business efficiency can invest in systems integration technologies, however the objectives, concepts and terminology of the solutions available on the market can be confusing.

Technologies such as Data Warehouse (DW), Service-Oriented Architecture (SOA), Enterprise Application Integration (AEI) and Master Data Management (MDM) are applications integration alternatives (Butler, 2011, p. 6) but have different purposes as will see throughout this work.

In environments with multiple applications, searches like: "what is the correct address of a particular customer?", "what products the customer purchased?" or "how many customers the company does have?" require a full and holistic view of the data.

MDM is a technology business discipline that helps organizations achieve a "single view of the truth" data in key domain areas such as customers, products and accounts (Cervo and Allen, 2011, p. 395).

There are specific MDM solutions for customer data and product. Solutions for customer data are referred as Customer Data Integration (CDI). Those focusing on the management of product data are called PIM - Product Information Management (Dreibelbis et al, 2008, p 665.). MDM systems that support multiple data fields, different styles and different implementation methods of use are called Multi-Form (MDM Dreibelbis et al, 2008, p. 665).

This work aims to study the state of the MDM art and, as proof of concepts, presents an implementation of a MDM system in a government agency.

Keywords: *Master Data Management, MDM, Master Data, Data Governance, Data Quality.*

LISTA DE ILUSTRAÇÕES

Ilustração 1 - Exemplos de Inconsistência de dados.....	1
Ilustração 2 - Pesquisa e Estudo.....	4
Ilustração 3 - Espaço de solução do MDM.....	20
Ilustração 4 - MDM estilo Consolidação.....	23
Ilustração 5 - MDM estilo Registo.....	24
Ilustração 6 - MDM estilo Coexistência.....	26
Ilustração 7 - MDM estilo <i>Hub</i> Transaccional.....	27
Ilustração 8 - MDM Analítico.....	28
Ilustração 9 - MDM Operacional.....	29
Ilustração 10 - MDM Empresarial.....	30
Ilustração 11 - Abordagem por tipo de dados.....	31
Ilustração 12 - Uma Arquitectura de Referência.....	32
Ilustração 13 - Serviços de MDM.....	34
Ilustração 14 - MDM é um sistema.....	38
Ilustração 15 - Programa de Governança de Dados.....	39
Ilustração 16 - Estilo de implementação MDM da APA.....	45
Ilustração 17 - Planeamento do projecto MDM da APA.....	46
Ilustração 18 - Modelo conceptual do MDM da APA.....	49
Ilustração 19 - Estudo dos Dados.....	51
Ilustração 20 - Infra-estrutura Tecnológica.....	54
Ilustração 21 - Tarefas de Limpeza e Consistência de dados.....	55
Ilustração 22 - Domínio Composto: Validar morada.....	57
Ilustração 23 - Cleansing do <i>Emails</i>	58
Ilustração 24 - Matching Policy.....	59
Ilustração 25 - Lista das entidades depois do Matching.....	60
Ilustração 26 - A <i>Subscription View</i> das entidades.....	61
Ilustração 27 - Lista de entidades no MDS Datastore.....	61
Ilustração 28 - <i>Interface Master Data</i> com MDS Addin-Excel.....	62
Ilustração 29 - Estado dos dados depois do Cleansing e Matching.....	63

LISTA DE TABELAS

Tabela 1 - Domínios de <i>Master Data</i>	9
Tabela 2 - Definição de <i>Master Data</i>	9
Tabela 3 - Tipos de dados.....	10
Tabela 4 - Dimensões de qualidade de dados.	11
Tabela 5 - Definição de MDM.....	17
Tabela 6 - MDM vs. Projectos tradicionais.	17
Tabela 7 - SOR vs. SR.....	19
Tabela 8 - Padrão de utilização de dados.	22
Tabela 9 - Estilos de implementação.	27
Tabela 10 - <i>Master Data</i> de Entidades da APA.....	47
Tabela 11 - Regras de Matching Policy.....	58

LISTA DE ABREVIATURAS

APA	-	Agência Portuguesa do Ambiente
BI	-	<i>Business Intelligence</i>
BPM	-	<i>Business Process Management</i>
CDI	-	<i>Customer Data Integration</i>
CRM	-	<i>Customer Relationship Management</i>
DG	-	<i>Data Governance</i>
DQS	-	<i>Data Quality Services</i>
DSS	-	<i>Decision Support Systems</i>
DW	-	<i>Data Warehouse</i>
EAI	-	<i>Enterprise Application Integration</i>
ERP	-	<i>Enterprise Resource Planning</i>
ETL	-	<i>Extract, Transform and Load</i>
GUI	-	<i>Graphical User Interface</i>
INESC	-	Instituto Nacional de Engenharia e Sistemas de Computadores
KB	-	<i>Knowledge Base</i>
MD	-	<i>Master Data</i>
MDM	-	<i>Master Data Management</i>
MDS	-	<i>Master Data Services</i>
ODBC	-	<i>Open DataBase Connectivity</i>
OLAP	-	<i>On-line Analytical Processing</i>
OLTP	-	<i>On-line Transaction Processing</i>
PIM	-	<i>Product Information Management</i>
ROI	-	<i>Return On Investment</i>
TCO	-	<i>Total Cost of Ownership</i>
SGBD	-	Sistema de Gestão de Base de Dados
SGE	-	Sistema de Gestão de Entidades
SOE	-	<i>System of Entry</i>
SOR	-	<i>System of Record</i>
SQL	-	<i>Structured Query Language</i>
SR	-	<i>System of Reference</i>
TI	-	Tecnologias de Informação

ÍNDICE

1. Introdução	1
1.1. Objectivos de Estudo	3
1.2. Contexto do Estudo	3
1.3. Questão de Investigação	3
1.4. Relevância de Estudo	3
1.5. Metodologia de Investigação	4
2. Estudo da Literatura	7
2.1. Tipos de dados	7
2.2. Qualidade de Dados	10
2.3. Integração de Dados	12
2.4. Master Data Management - MDM	15
2.4.1. Definição	15
2.4.2. Sistema de Registo vs. Sistema de Referência	18
2.4.3. Abordagens de Implementação	19
2.4.4. MDM: Abordagem por Dimensões	19
2.4.5. MDM: Abordagem por Tipo de Dados	28
2.4.6. Arquitectura de Referência	31
2.5. Governança de Dados	37
3. Implementação do Conceito	43
3.1. Agência Portuguesa do Ambiente	44
3.2. Enquadramento do Projecto	44
3.3. Implementação do Projecto	46
3.3.1. Estudo sobre Produtores e Consumidores dos <i>Master Data</i>	46
3.3.2. Identificação dos Administradores dos <i>Master Data</i>	52
3.3.3. Governança dos <i>Master Data</i>	52
3.3.4. Infra-estrutura Tecnológica	53
3.3.5. Processos de Limpeza, Eliminação e Testes	55
3.3.6. Partilha dos Dados	60
4. Conclusão	63
4.1. Conclusões	63
4.2. Limitações	64
4.3. Recomendações	64
4.4. Investigação Futura	65
5. Referências Bibliográficas	67

1. INTRODUÇÃO

É um cenário bastante comum encontrar organizações onde os dados de determinadas aplicações não serem partilhados na organização, fazendo com que os mesmos dados se encontrem replicados por outras aplicações. Por exemplo, o nome, a morada e os contactos dos mesmos clientes podem estar sucessivamente replicados no sistema de facturação (ERP), no sistema de gestão de relacionamento (CRM) etc.

A ilustração seguinte mostra o que acontece frequentemente nesses ambientes:

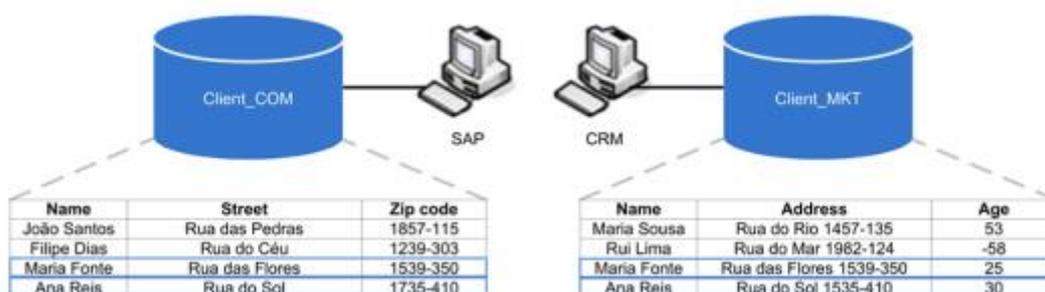


Ilustração 1 - Exemplos de Inconsistência de dados

Fonte: Galhardas, Torres e Damásio, 2007, (Galhardas, Torres e Damásio, 2007, p. 2)

Nesta ilustração podemos observar que há dados que se repetem no sistema de SAP e no CRM. De salientar também que o formato dos dados são diferentes: os códigos postais (C.P.) no sistema de CRM estão junto às moradas, enquanto no outro estão armazenados num campo autónomo. Observa-se igualmente que o C.P. da cliente “Ana Reis” no sistema de CRM é diferente do sistema de SAP.

A consequência, nos processos de negócio de uma organização, com dados organizados desta forma podem ser muitas. Os dois exemplos a seguir descrevem duas consequências reais:

Exemplo 1

Um morador residente nos arredores de Lisboa aderiu ao serviço de fornecimento de energia eléctrica de uma nova empresa no mercado nacional que se apresentou com promessas de melhores preços e melhores serviços (ENDESA, S.A., 2013). Para este cliente não foi bem assim que aconteceu. Para a formalização do contrato, o futuro

cliente forneceu electronicamente os dados de identificação, localização e contactos conforme foram solicitados.

Sem mais nenhuma comunicação, decorridos dois meses depois, recebeu com espanto um aviso de eminente corte de energia eléctrica da parte do novo serviço por falta de pagamento (ENDESA, S.A., 2013). Numa linguagem típica, é informado igualmente que o incumprimento já teria sido comunicado à Entidade Reguladora dos Serviços Energéticos (ERSE) para as devidas consequências.

Após reclamar que não recebera nenhuma factura, uma semana depois, o cliente recebeu as facturas em falta e uma nota informativa dizendo que as mesmas não tinham sido entregues por motivos técnicos relacionados com o código postal que não estava correcto. Isto leva a concluir que nessa empresa deveriam existir pelo menos duas listas de contactos de clientes com dados diferentes: uma seria usada para as facturas e a outra para informar os clientes de eminente corte de energia eléctrica.

A seguir vamos apresentar um exemplo que envolve uma integração aplicacional ao nível dos dados que não correu bem:

Exemplo 2

Citius é a plataforma do sistema de justiça que permite aos agentes judiciais entregar por via electrónica os processos judiciais nos tribunais portugueses e fazer toda a tramitação processual. O *Citius* foi alvo de um processo de consolidação ao nível dos dados no âmbito de um programa de Reforma do Mapa Judiciário.

De acordo com um comunicado público emitido pelo Instituto de Gestão Financeira e Equipamento da Justiça (IGFEJ), entidade responsável pela plataforma *Citius*, o processo de fusão das 300 bases de dados das comarcas fez com que no final muitos processos judiciais tivessem ficado mal classificados devido a um número muito elevado de incongruências nos dados de origem (IGFEJ, 2014), isto é, os dados de indexação nos sistemas de origem não tinham qualidade suficiente para permitir uma classificação correcta nas 23 bases dados finais.

O professor catedrático José Tribolet, presidente do INESC, afirmou através da Rádio Notícias TSF (2014) que “O *Citius* não vai ao sítio em menos de dois anos”; segundo ainda a Rádio Notícias TSF (2014), “O *Citius* está sem funcionar há várias semanas o que levou a ministra da Justiça a admitir a existência de «transtornos» pelos quais assumiu a responsabilidade política”.

Entretanto, todos os dias surgiam notícias dando conta de enormes constrangimentos no funcionamento dos tribunais. O presidente da Associação Sindical de Juízes pediu na rádio TSF (18/9/2014) medidas de excepção nomeadamente a aprovação de legislação que suspendesse todos os prazos processuais impedindo que muitos processos prescrevessem.

Através destes dois exemplos, pretendemos demonstrar algumas consequências resultantes da insuficiente qualidade nos dados e da necessidade de implementação de soluções de garantia de qualidade nas organizações.

1.1. OBJECTIVOS DE ESTUDO

Os objectivos deste trabalho consistem em:

- Descrever os principais conceitos sobre *Master Data* e *Master Data Management* (MDM);
- Implementar um projecto de MDM num organismo público como comprovação dos conceitos.

1.2. CONTEXTO DO ESTUDO

O âmbito deste estudo está limitado aos conceitos fundamentais sobre MDM nomeadamente: *Master Data*, *Data Quality*, *Data Governance*, Estilos de implementação e Arquitectura de Referência.

1.3. QUESTÃO DE INVESTIGAÇÃO

- MDM, o que é?
- Porquê o MDM e não outras tecnologias?
- MDM, mito ou realidade?

1.4. RELEVÂNCIA DE ESTUDO

A relevância deste estudo é fundamentada nas seguintes constatações:

1. As tradicionais arquitecturas de integração de aplicações e Sistemas de Informação (SI) tais como o DW, SOA e EAI não foram desenhadas para resolver os problemas de qualidade de dados nas aplicações operacionais;
2. Um sistema de MDM não só fornece serviços de garantia de qualidade de dados mas também possui serviços de partilha de dados a toda a organização;
3. Actualmente existem muito poucas opções de literatura nas bibliotecas e nas livrarias portuguesas que abordam a temática do MDM.

Consideramos que estas constatações justificam este estudo tendo por objectivo:

- Destacar e divulgar os conceitos sobre MDM;
- Destacar os aspectos técnicos que diferenciam o MDM das soluções similares de integração;
- Realçar os benefícios do MDM para as organizações que procuraram tirar vantagens competitivas dos activos que possuem nos seus repositórios de dados.

1.5. METODOLOGIA DE INVESTIGAÇÃO

A metodologia usada nesta investigação consistiu, primeiramente, em fazer um levantamento bibliográfico sobre o MDM e tecnologias similares, seguido de um processo cíclico de estudo da literatura e de nova selecção da literatura.

A técnica aplicada para a comprovação dos conceitos foi a de implementar o sistema de MDM da Agência Portuguesa do Ambiente, I.P. (APA).



Ilustração 2 - Pesquisa e Estudo

A selecção da literatura foi obtida fazendo pesquisas preliminares na internet e junto dos fornecedores de soluções de MDM. Por exemplo, a pesquisa da frase “*Master Data*

Management” no Google Académico permitiu identificar um conjunto de livros electrónicos e trabalhos realizados sobre este tema. Com os documentos consultados nessas pesquisas foi possível identificar e seleccionar nova literatura que, por sua vez, permitiu obter novas referências bibliográficas, e assim sucessivamente, até obter as referências bibliográficas que compõe este estudo.

A documentação consultada encontra-se nas bibliotecas universitárias de Lisboa, na internet e essencialmente em plataformas de livros electrónicos. De salientar que as páginas dos livros electrónicos consultados não estavam numeradas o que impossibilitou a indicação exacta do número das páginas nas citações conforme é solicitado no *template* fornecido pela universidade. No lugar das páginas colocamos a posição mais próxima do material citado.

O principal tema deste trabalho é o *Master Data Management* tendo como principais fontes de estudo autores muito referenciados cujas publicações cobrem os principais conceitos aqui apresentados. Sendo alguns destes autores:

Loshin, Cervo, Allen, Dreibelbis, Dudov e Berson.

A bibliografia seleccionada cobre os seguintes conceitos: *Master Data*, *Master Data Management*, *Data Governance*, *Data Integrity*, *Data Quality* e Arquitectura de Referência.

A pesquisa através da internet fez-se com frequência com a utilização das palavras-chaves: MDM, *Master Data*, *Data Governance* e *Data Quality*.

2. ESTUDO DA LITERATURA

2.1. TIPOS DE DADOS

INTRODUÇÃO

Consideramos que para uma boa compreensão deste estudo é importante conhecer antes de mais o tipo de dados que existem nas organizações e o papel que cada tipo desempenha nas operações quotidianas.

Os dados de uma empresa podem ser classificados como operacionais e não operacionais (Cervo e Allen, 2011, p. 486):

- Dados “operacionais” - são dados usados no suporte das actividades quotidianas das organizações;
- Dados “não operacionais” - são dados normalmente capturados em Datawarehouse (DW) para sistemas de Business Intelligence (BI), *Reporting*, Dashboards etc.

No dia-dia as organizações lidam também com, pelo menos, mais quatro tipos de dados (Dreibelbis et al., 2008, p. 1217) e (Butler, 2011, p. 4):

- Dados de referência (*Reference Data*) - são dados usados na categorização de outros dados. Este tipo de dados define e distribui um conjunto de simples listas de valores usadas em “view lookups” para assegurar o uso consistente de códigos como por exemplo distritos e códigos postais (Dreibelbis et al., 2008, p. 1252) ou para assegurar a consistência de valores de *Master Data* tais como países ou cores (Dreibelbis et al., 2008, p. 1270).
- Dados históricos - são dados usados para guardar as alterações feitas nos *Master Data* e nos dados transaccionais ao longo do tempo (Dreibelbis et al., 2008, p. 1288);
- Metadados - São informações descritivas úteis para ajudar as pessoas e sistemas a entender outros dados (Dreibelbis et al., 2008, p. 15042).
- *Master Data* - são considerados como sendo os principais dados numa organização e representam objectos sobre os quais são realizadas as transacções.

Iremos descrever resumidamente os dados tipos transaccionais (dados operacionais), analíticos (dados não operacionais), *Master Data* e metadados.

DADOS TRANSACCIONAIS

Este tipo de dados é usado para realizar as funções de venda, compra, transferência financeiras, entre outras. Estes dados representam objectos envolvidos nas transacções bem como as próprias transacções. Quando um cliente compra um produto, os objectos da transacção são o cliente e o produto. Os dados da transacção são o tempo, o lugar, preço, desconto, métodos de pagamento, etc. (Butler, 2011, p. 4).

DADOS ANALÍTICOS

São dados que entram em processos de tomada de decisão descrevendo os padrões de compra dos clientes, identificando os comportamentos dos mercados, etc. Estes dados são guardados em estruturas de tabelas projectadas para suportar operações de consultas constantes (Butler, 2011, p. 5).

MASTER DATA

Master Data são nomes críticos numa empresa: representam objectos de negócio que são partilhados entre as aplicações transaccionais (Dreibelbis et al., 2008, p. 1270). Descrevem os atributos das principais entidades de uma organização: clientes, fornecedores, produtos, materiais, empregados, etc. Encontram-se em todas as áreas de negócio. Por exemplo, os dados dos fornecedores são usados simultaneamente em vários departamentos e guardados em múltiplos sistemas de informação (Otto e Ebner, 2010, p. 102).

Para Wolter e Haselden (2006), os *Master Data* podem ser agrupados em quatro classes designadas, invariavelmente, por áreas de assunto, áreas de domínio, tipos de entidades ou domínios de dados (Cervo e Allen, 2011, p. 353) e que são:

- Pessoas (clientes, funcionários, fornecedores, vendedores, etc.);
- Coisas (produtos, peças, catálogos, etc.);
- Localização (países, regiões, cidades, locais, etc.);
- Conceitos (contractos, garantias, licenças, etc.).

Aos *Master Data* de pessoas e empresas designam-se colectivamente por “Party Data” (Dreibelbis et al, 2008, p. 821).

De acordo com Dreibelbis et al. (2008, p. 839) os *Master Data* podem ser agrupados por **Party Data, Produtos e Contas**. Para facilmente identificar os elementos que os compõem, estes autores sugerem a formulação das seguintes perguntas: “quem?”, “o quê?” e “como?”:

Tabela 1 - Domínios de *Master Data*.

Quem? (Party)	O quê? (Produtos)	Como? (Account)
Clientes	Produtos	Contas
Fornecedores	Materiais	Contractos
Empregados	Activos	Acordos
Organizações	Serviços	Contractos

Fonte: De Dreibelbis et al. (2008, p. 868)

Para os autores Otto e Ebner (2010, p. 102), as perguntas do tipo “onde é que o cliente vive?” e “onde é que o produto foi comprado?” referem-se à localização de outros domínios. Ou seja a localização está associada a todos os outros domínios e por essa razão o domínio “Onde?” deve ser visto como um subdomínio de *Master Data* e não como um domínio normal.

Para o *Master Data* foram encontradas muitas definições. O quadro seguinte sintetiza as diferentes definições da literatura consultada. De destacar a utilização das palavras “fundamentais”, “principais” e “críticos” em todas as definições.

Tabela 2 - Definição de *Master Data*.

Definição	Fonte
<i>Master data</i> são itens de dados que descrevem as entidades fundamentais de uma organização.	Dahlberg, Heikkilä, Heikkilä (2011, type of data and process).
<i>Master data</i> é o conjunto consistente e uniforme de identificadores e atributos que descreve as principais entidades de uma empresa tais como clientes, catálogos, cidadãos, fornecedores, locais, hierarquias e contas.	IT Glossary da Gartner (search for <i>master data</i>)
<i>Master data</i> é a informação crítica de suporte às operações e BI das empresas. Esta informação é normalmente categorizada por áreas de entidades de <i>master data</i> , também designadas por domínio de dados, tais como clientes, produtos, fornecedores, parceiros, empregados, materiais etc.	Cervo e Allen (2011, p. 353)

METADADOS

Estes dados descrevem a estrutura, origem e o significado das coisas (Dreibelbis et al., 2008, p. 1252). Por exemplo, o catálogo de uma base de dados contém informação sobre os dados geridos pelo Sistema de Gestão de Base de Dados (SGBD) indicando o domínio de valores para um atributo, indicando se um atributo é chave primária, quais os valores por predefinição para uma coluna, se um atributo é obrigatório, entre outros.

Tabela 3 - Tipos de dados.

Tipo	Tipo de informação	Exemplo
Transaccional	Informação com detalhes sobre a transacção de negócio.	Dados de uma factura: custo total, valor de desconto.
Analítico	Informação que suporta tomada de decisão.	Total de vendas de um produto numa determinada região e num determinado tempo.
<i>Master data</i>	Informação que indica objectos sobre os quais são feitas transacções.	Clientes, produtos, contractos, licenças.
Metadados	Informação sobre outras informações	Valor por defeito para um atributo.

2.2. QUALIDADE DE DADOS

O termo qualidade é associado, geralmente, às expressões comuns como “qualidade das peças”, “qualidade dos serviços”, “qualidade dos produtos”, “qualidade dos materiais” e por aí adiante, refere-se, por definição, ao “conjunto de atributos e características de uma entidade ou produto que determinam a sua aptidão para satisfazer necessidades e expectativas da sociedade” (Decreto-Lei nº 140/2004 de 8 de Junho).

Embora não tenham cheiro, cor nem acabamento, os dados digitais possuem características que podem influenciar o comportamento das pessoas. Por exemplo, segundo Dreibelbis et al (2008, p. 10895) se os clientes de uma empresa não podem ser identificados sem ambiguidade ou se as facturas são enviadas sistematicamente para moradas erradas, esta situação pode criar um clima de insatisfação e afastar os clientes por esse mesmo motivo.

Portanto, para os dados digitais “qualidade” significa igualmente possuir um conjunto de características de satisfação. Segundo Osion (2003, p. 24), diz-se que “um dado tem qualidade se satisfaz os requisitos para os quais é predefinida a sua utilização, faltando-lhe qualidade se não satisfizer esses requisitos”.

As características que determinam a qualidade dos dados digitais não se esgotam no entanto numa lista, segundo Loshin (2009, p. 1882) as dimensões que indicam a qualidade dos dados são: *consistency*, *accuracy*, *completeness* e *timeliness*. Dreibelbis et al (2008, p. 15042) considera que um *Master Data* deve possuir, como garantia de qualidade, as seguintes características: *timely*, *relevant*, *complete*, *valid*, *accurate* e *consistente*. O quadro seguinte descreve os requisitos de qualidade mais importantes e desejáveis de acordo com a literatura consultada.

Tabela 4 - Dimensões de qualidade de dados.

Dimensões	Definição
Precisão (<i>accuracy</i>)	Medida que indica se os dados representam correctamente, uma acção ou objecto do mundo real.
Perfeição (<i>completeness</i>)	Medida que indica se todos os valores estão presentes num processo de recolha de dados.
Oportunidade (<i>timeliness</i>)	Medida que indica se os dados representam o mundo real num determinado momento.
Consistência (<i>consistency</i>)	Medida que indica se os valores num conjunto de dados são coerentes com os valores de um outro conjunto Loshin (2009, p. 1906).
Relevância (<i>relevancy</i>)	Medida que indica se o dado é útil e aplicável para o que é destinado.
Acessibilidade (<i>accessibility</i>)	Medida que indica se os dados estão disponíveis num determinado momento.

Fonte: Adaptado de Otto e Ebner, 2010, p. 104

Apesar de todos os requisitos aqui referidos serem muito importantes como indicadores de qualidade, Loshin (2009, p. 1836) afirma que “numa empresa um dos factores de motivação para a instituição de um programa de MDM é a necessidade de se obter a consistência (*consistency*) e precisão (*accuracy*) dos dados”. Estes devem ser os requisitos de qualidade mais importantes num projecto de MDM.

A “Garantia de Qualidade” de dados é uma disciplina focalizada em garantir dados aptos para uso numa organização que de acordo com o grupo Gartner (Gartner, 2013), ela não é só tecnologia, inclui também as funções e as estruturas organizacionais, processos de monitorização, notificação, medição, *reporting* e alertas. Dada a dimensão e a complexidade do universo dos dados numa organização, as ferramentas para ajudar a automatizar as tarefas de limpeza de dados continuam a atrair cada vez mais fabricantes de *software* de qualidade (Gartner, 2013).

O mercado das ferramentas de qualidade de dados inclui fornecedores que disponibilizam produtos de *software* independentes que empregam as seguintes técnicas comuns de limpeza (Gartner, 2013):

- Criação de perfis de dados (*Profiling*) e medição de qualidade - Avaliação dos dados e procura entender o nível de precisão dos mesmos. Através do exame de todos os valores de uma coluna, procura inferir as propriedades da coluna tais como o domínio dos valores, valor máximo e mínimo, tipo de dado, comprimento, etc.;
- Análise (*Parsing*) e padronização de dados - Decomposição dos campos de texto em partes e formatação dos valores em esquemas (*layouts*) condizentes com os padrões da empresa;
- Enriquecimento - Comparação dos dados internos com fontes externas para corrigir os dados de códigos postais, descritores geográficos, etc.;
- Comparação e vinculação (*Matching*) – Método que identifica, compara, liga ou funde entradas relacionadas num conjunto de dados;
- Monitorização - Acompanhamento da qualidade dos dados com correcção automática das variações com base em regras de negócios predefinidas.

2.3. INTEGRAÇÃO DE DADOS

“Não podemos entender o impacto das novas tecnologias sem olharmos para o passado” (Cummins, 2002, p. 2).

A diversidade de arquitecturas de aplicações e tecnologias, em conjunto com as especificidades das aplicações, fragmentaram as empresas criando grandes barreiras à captura, comunicação e integração a informação de gestão necessária à operação efectiva e melhoria das empresas (Cummins, F. 2002).

Por definição *Middleware* é um componente de *software* que permite a uma aplicação comunicar com outra aplicação dentro de um mesmo computador ou entre dois computadores ligados em rede (Silva, M., 2003). Um exemplo de *Middleware* é *Open Database Connectivity* (ODBC), um *software* que permite que uma aplicação possa aceder a uma base de dados de diferente tecnologia ou vendedor. O *software* que permite a interligação entre sistemas de informação (SI) designa-se por “Integração de

SI". A integração de SI pode ser feita ou ao nível dos dados, linguagens de programação ou ao nível de interface com o utilizador GUI (Silva, M., 2003).

Para ultrapassar a fragmentação nas empresas causada pela diversidade de aplicações, surgiram algumas tecnologias de "Integração de SI" como por exemplo a Enterprise Application Integration (EAI) e o Service Oriented Architecture (SOA) (Butler, 2011, 2011, p. 6). Mas antes deste surgimento, as empresas integravam as aplicações por meios "ad hoc" designados actualmente por Middleware Tradicional. Era um meio expedito de interligação baseado em ligações "ponto-a-ponto". Sempre que era necessário interligar duas ou mais aplicações, a cada aplicação era adicionada uma camada de *Middleware* que criava, enviava, recebia e lia as mensagens contidas em registos (Ruh et al, 2001).

A camada de *Middleware* Tradicional fazia a abstracção de toda a complexidade de lidar com os diferentes programas, linguagens de programação, formatos de dados e protocolos de redes permitindo desta forma as aplicações diferentes pudessem ligar-se entre si. Ligar apenas duas aplicações podia ser simples mas interligar ponto-a-ponto um conjunto de aplicações já necessitava de muitas alterações. A arquitectura final podia ser um emaranhado de ligações complexas e dispendiosas, de difícil manutenção e, como consequência (Linthicum, 2008, p. 8), a integração podia acabar por ser mais dispendiosa do que fazer a manutenção da própria aplicação a que se pretendia ligar.

Atendendo à complexidade e à escala de esforço necessários para manter esta forma expedita de integrar, as empresas foram pressionadas a encontrar alternativas (Ruh et al, 2001) como são as que a seguir são descritas.

DATA WAREHOUSE (DW)

DW emergiu para suportar a tomada de decisão nas organizações. A informação transaccional contida nas aplicações operacionais das empresas, designadas por *Online Transaction Processing (OLTP)*, são muitas vezes de difícil acesso na forma adequada para análise do ponto de vista da sua consistência (Cummins, 2002, p. 4). Por outro lado, os dados nas aplicações OLTP estão em constante alteração.

O DW captura e integra nas suas bases de dados a informação proveniente das OLTP. A informação é transformada para adquirir a consistência necessária antes de ser carregada na DW. A transformação dos dados faz-se através de um processo intermédio suportado por ferramentas de ETL (Extract, Transform and Load). O ETL

extrai os dados de várias fontes (data source), transforma-os para adquirirem consistência e integridade, e finalmente carrega os dados a DW.

Os dados são carregados no DW e consumidos apenas pelos sistemas a jusante, isto é, não são retornados para as aplicações de origem. Portanto, os dados das aplicações de origem permanecem no mesmo estado de inconsistência e imprecisão como estavam antes de serem transformados e carregados na DW.

ENTERPRISE APPLICATION INTEGRATION (EAI)

Esta tecnologia também captura e integra dados mas com um objectivo diferente. EAI surgiu para integrar aplicações comerciais específicas designadas por COTS¹ para as quais, na década de 1990, requeriam a integração com sistemas mais antigos (Cummins, F. 2002, p. 6). Para resolver o problema das ligações ponto-a-ponto, característica do middleware tradicional, foi criado um intermediário (broker) com a função de fazer a recepção e envio de mensagens entre as aplicações, e desde então, surgiram novos produtos para suportar esta prática.

Actualmente a integração por EAI é feita num ambiente composto por um repositório de metadados e um conjunto de conectores de aplicações usando os seguintes mecanismos (Butler, 2011, p. 7):

- As aplicações disponibilizam as funções sob forma de serviços que são acedidos por outras aplicações como por exemplo: funções de criação de contas bancárias;
- Partilhar informação entre aplicações, como por exemplo informação sobre clientes, fornecedores, etc.;
- Coordenar processos de negócio, como por exemplo fornecer funcionalidades para gerir workflow de negócio.

Os mecanismos da EAI integram as aplicações a vários níveis, mas os dados nos sistemas de origem permanecem inalterados tal como no DW.

¹ Commercial Off-The-Shelf (COTS). Unidade de software independente que encapsula, dentro de si, o seu projecto e implementação, e oferece serviços, por meio de interfaces bem definidas, para o meio externo (Gerosa, M em "O Desenvolvimento Baseado em Componentes).

SERVICE ORIENTED ARCHITECTURE (SOA)

SOA é uma arquitectura de integração em que os componentes são acedidos através de serviços. O SOA baseia-se na ideia de que a lógica necessária para resolver um problema maior pode ser melhor alcançada se este for decomposto numa colecção de pequenas peças relacionadas entre si” (Erl, 2006, p. 3). Cada serviço pode encapsular uma tarefa de uma etapa individual ou de um subprocesso composto por um conjunto de etapas.

Dreibelbis et al, (2008, p. 15042) definem o SOA como:

- Uma arquitectura de integração em que os componentes são disponibilizados através de serviços. Estes serviços são disponibilizados através de protocolos de interfaces e comunicações independentes de plataformas que encapsulam funcionalidades a serem disponibilizados aos serviços consumidores.
- Do ponto de vista técnico, um SOA disponibiliza funções que podem ser usadas em processos de negócio, como por exemplo, retornar a morada de um cliente Dreibelbis et al, (2008). A informação retornada pode provir de várias fontes. Se os dados das fontes não tiverem qualidade será necessário uma solução intermédia para fornecer consistência e integridade aos dados do SOA à semelhança dos sistemas analíticos.

Esta tecnologia, tal como nas soluções anteriores, os dados nas aplicações de origem permanecem no mesmo estado de inconsistência.

Resumo

As tecnologias comuns de integração que vimos até ao momento, integram soluções intermédias para dar consistência aos dados de que necessitam. Esses dados são consumidos por outras aplicações mas não são devolvidos aos repositórios de origem. É aqui portanto que entra o papel do MDM como se verá nos próximos capítulos.

2.4. MASTER DATA MANAGEMENT - MDM

2.4.1. DEFINIÇÃO

No capítulo anterior, foi referido que as soluções de integração como SOA, DW e EAI lidam bem com a fragmentação dos dados mas deixam os dados dos sistemas de

origem no mesmo estado. Para resolver este problema, de acordo com Cervo e Allen (2011, p. 362), “as organizações chegaram à conclusão que a forma efectiva de o resolver seria através da criação de uma abordagem de fonte única na gestão dos *Master Data* com base em altos padrões de qualidade e governança ao nível de toda a organização”. Nesse sentido Dreibelbis et al, (2008, p. 15042), afirma que “num mundo ideal, haveria um único lugar onde todos os *Master Data* comuns da organização seriam armazenados e geridos. Os dados seriam precisos, consistentes e mantidos de forma coerente e segura. Todas as alterações teriam lugar nessa única cópia do *Master Data* e todos os diferentes consumidores interagiriam com essa fonte oficial de informação”. Passaria a haver uma fonte única de *Master Data*. De acordo com Dreibelbis et al, (2008, p. 15042), uma única fonte de *Master Data* significa:

- Haver uma fonte de informação oficial;
- Possibilidade de usar a informação de forma consistente;
- Suportar as mudanças de negócio necessárias envolvendo os *Master Data* e a gestão dos *Master Data*.

Para Loshin (2009, p. 448) o sistema de MDM é a materialização desse mundo ideal que é alcançado através da “colecção das melhores práticas de gestão, a conjugação dos principais interessados e clientes do negócio, incorporando aplicações de negócio, métodos e ferramentas de gestão, políticas, serviços e infra-estruturas”.

Neste estudo foram muitas e diferentes as definições encontradas para descrever o que é, e para serve, o MDM. Apresentamos algumas definições:

Tabela 5 - Definição de MDM.

Definição	Fonte
MDM é uma disciplina tecnológica de negócio que ajuda as organizações a alcançarem uma “visão única da verdade” em áreas de domínio importantes como clientes, produtos e contas.	Cervo e Allen (2011, p. 365)
MDM refere-se às disciplinas, tecnologias e soluções usadas para criar e manter os <i>master data</i> de uma empresa consistentes e precisos para todos os <i>stakeholders</i> internos e externos.	Dreibelbis et al. (2008, p. 363)
MDM é uma disciplina tecnológica em que o negócio e as Tecnologias de Informação trabalham em conjunto para garantir a uniformidade, precisão, gestão, consistência semântica e a prestação de contas (<i>accountability</i>) sobre os principais activos de <i>master data</i> partilhados da empresa.	IT Glossary da Gartner
MDM é uma plataforma de processos e tecnologias para criar e manter um ambiente de dados de qualidade, confiável, sustentável, preciso e seguro. Representa a “versão única e holística da verdade” para os <i>master data</i> e relações, assim como uma referência usada na empresa num vasto e disperso conjunto de sistemas, áreas de negócios, canais e comunidades de utilizadores.	Dubov e Berson (2011, p. 888)

EQUÍVOCOS SOBRE MDM

Para esclarecer alguns equívocos o quadro abaixo apresenta aquilo que é essencialmente a mais-valia do MDM nas organizações face às outras iniciativas de gestão de informação que já existem no mercado, tais como, *Business Intelligence* (BI) e *Data Warehouse* (DW).

Tabela 6 - MDM vs. Projectos tradicionais.

Projectos Tradicionais	A Mais-Valia do MDM
BI, DW, Operacional	Fornecer <i>master data</i> consistentes e precisos aos sistemas de BI, DW e às aplicações operacionais.
Consistência de dados nas aplicações transaccionais	Contribuir para a consistência dos <i>master data</i> que suportam as transacções passando a ser a fonte oficial de <i>master data</i> para todos os sistemas, aplicações e processos.
Integração de dados	Os projectos de MDM envolvem colaboradores, regras e políticas de governança e administração de dados, tecnologias e aplicações informáticas para garantir a consistência, consolidação e a integração dos dados.
Integração e gestão dos dados de toda a empresa.	O MDM é só para os <i>master data</i> . Os outros dados transaccionais são mantidos nos sistemas operacionais e analíticos.

Fonte: Imhoff, C. e White, C., 2006, p. 7

Nos próximos capítulos serão apresentadas as diferentes abordagens, os diferentes tipos de implementação e outros aspectos técnicos importantes que caracterizam os sistemas de MDM.

2.4.2. SISTEMA DE REGISTO VS. SISTEMA DE REFERÊNCIA

O objectivo de um sistema de MDM numa organização, como já foi referido, consiste em fornecer *Master Data* com qualidade para toda a organização. Dreibelbis et al (2008, p. 1025) afirmam que para isso o ideal é que:

- Haja apenas uma versão do *Master Data* gerida no MDM;
- Todas as aplicações, processos e colaboradores que necessitam do *Master Data* recorram ao MDM;
- Todas as actualizações sejam feitas sob controlo do MDM.

Desta forma fica garantido que os *Master Data* ficam protegidos de alterações indevidas, passando o MDM a ser a “única fonte da verdade”. Isto é, o *Hub* MDM é o único lugar em toda a organização onde os *Master Data* são garantidos serem de qualidade e actualizados atempadamente (Dreibelbis et al, 2008, p. 1025). Ao conjunto dos *Master Data* neste cenário idealizado designa-se por System of Record (SOR).

Quando as aplicações querem estar seguras de que estão a lidar com a informação mais recente e com a mais elevada qualidade, consultam esta fonte da verdade (Dreibelbis et al, 2008, p. 1025).

Mas a necessidade de se fazerem profundas alterações nas infra-estruturas tecnológicas da organização para implementar e manter um SOR conjugada com, segundo Dreibelbis et al (2008, p. 1025), a complexidade técnica, financeira, aplicacional e legal, torna difícil conseguir atingir este desiderato. Todas estas dificuldades fazem com que tenham de existir cópias dos *Master* nas aplicações de origem onde podem ser geridas, integradas e sincronizadas com o SOR para garantir a qualidade das mesmas cópias.

Este conjunto de cópias nas aplicações operacionais sincronizadas com o SOR do *Hub* MDM designa-se por Sistema de Referência (SR). Uma vez que estas cópias estejam sincronizadas, o SR passa a funcionar também como uma fonte oficial (with authority) dos *Master Data* (Dreibelbis et al, 2008, p. 1042).

O SR é usado preferencialmente como fonte “read-only” de informação com todas as alterações a serem feitas no SOR (Dreibelbis et al, 2008, p. 1042). O quadro abaixo descreve as vantagens e as desvantagens do SOR e do SR.

Tabela 7 - SOR vs. SR.

	Vantagem	Desvantagem
SOR	Tem os <i>master data</i> sempre limpos e actualizados. É a fonte única da verdade.	Há necessidade de se modificar as aplicações operacionais para deixarem de fazer alterações aos <i>master data</i> . Os custos financeiros para alterar as aplicações podem ser elevados se estas forem de terceiros.
SR	A complexidade de implementação do MDM pode ser menor porque os <i>master data</i> são copiados do MDM para as aplicações não sendo necessário fazer grandes alterações nas aplicações. As alterações podem ficar-se ao nível das bases de dados.	As cópias para as aplicações são feitas em diferido por esta razão os dados não estão sempre actualizados face ao SOR.

2.4.3. ABORDAGENS DE IMPLEMENTAÇÃO

Existem muitos aspectos técnicos a ser considerados quando se pretende desenhar uma solução de MDM. Onde devem residir os *Master Data*? Que dados devem ser abrangidos? Que solução MDM deve ser escolhida? Qual o âmbito do MDM face aos dados?

No estudo da literatura foram identificadas duas abordagens que ajudam a responder a estas e a outras questões:

- Abordagem por dimensões - nesta abordagem o MDM é visto como um espaço de solução com as seguintes dimensões: domínio dos dados, tipo de utilização e estilo de implementação;
- Abordagem por tipo de dados - nesta abordagem o MDM é visto exclusivamente na perspectiva dos “dados operacionais” e “dados não operacionais” alvos do projecto.

2.4.4. MDM: ABORDAGEM POR DIMENSÕES

É grande e variada a quantidade de livros sobre MDM bem como as diferentes definições em torno dos seus conceitos. Por exemplo, segundo Cervo e Allen (2011, p. 516), para alguns autores os conceitos “Abordagem” ou “Arquitectura” significam o mesmo que “Plataforma” ou “Implementação” para outros. Ao que alguns autores chamam “Desenho e Desenvolvimento” de MDM outros designam-no por “Estilo de

Implementação de MDM”. Para Cervo e Allen (2011, p. 516), quando se pretende escolher a abordagem de implementação mais adequada, é particularmente importante “entender de que forma o *Master Data* será integrado, usado, gerido, aperfeiçoado e governado”, pois são estas questões que permitem definir o âmbito e o tipo de operações a efectuar no *Hub* MDM e nas aplicações operacionais, etc. Desta forma, Dreibelbis et al (2008, p. 819) definem o MDM como um espaço de solução nas seguintes dimensões:

- Domínio de Dados - Esta dimensão classifica o MDM com base no tipo de dados a ser gerido: clientes ou produtos;
- Padrão de Utilização - É a dimensão que classifica o MDM com base no papel que o *Master Data* desempenha na organização (Dreibelbis et al, 2008, p. 882);
- Estilo de Implementação - Dimensão que caracteriza o MDM com base no tipo de operações (leitura e escrita) a realizar nas aplicações e no *Hub* Central para garantir a consistência dos *Master Data*.

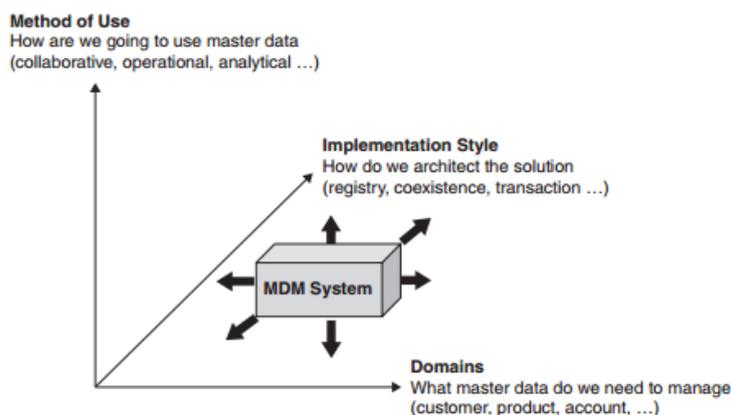


Ilustração 3 - Espaço de solução do MDM

Fonte: Dreibelbis et al, 2008 (Dreibelbis et al, 2008, p. 819)

DOMÍNIO DOS DADOS

Conforme foi atrás referido, esta classificação baseia-se no tipo de *Master Data* que é foco na solução. Nesta perspectiva Berson e Dubov (2011, p. 911 e p. 1040) e Dreibelbis et al (2008, p. 821) classificam o MDM em CDI (clientes) e PIM (produtos).

CUSTOMER DATA INTEGRATION (CDI)

CDI são soluções desenhadas especificamente para fazerem a gestão da informação de pessoas e empresas. O CDI tem por objectivo gerir a “party data” proveniente de várias aplicações e partilhar a informação aos sistemas a jusante (Dreibelbis et al, 2008, p. 821).

Product Information Management (PIM)

PIM são soluções específicas em fazer a gestão de *Master Data* de produtos e serviços. Estes sistemas recolhem a informação proveniente de múltiplas fontes, conseguindo consenso na sua definição, e publicam a informação consensual na web, nos sistemas de marketing, etc. (Dreibelbis et al, 2008, p. 821).

CDI e PIM são as soluções mais comuns porque, segundo o grupo Gartner (Gartner, 2013), no mercado há relativamente ainda pouco interesse em outros domínios de dados que justifique o desenvolvimento de MDM específicos para outras áreas de assunto.

Estas soluções por serem específicas podem ser menos complexas de implementar e por isso mais vantajosas, mas quando a organização prevê evoluir para outros domínios de dados, pode ser mais vantajoso implementar soluções de MDM *Multi-Form*.

MÉTODO DE UTILIZAÇÃO

Como foi dito anteriormente, a classificação por Método de Utilização tem como base o papel que o *Master Data* desempenha nos processos e aplicações da organização e que podem ser: criação colaborativa, operacional e analítica (Dreibelbis et al, 2008, p. 882) conforme descrito no quadro seguinte:

Tabela 8 - Padrão de utilização de dados.

Método de utilização	Descrição	Papel do MDM
Criação Colaborativa	Modelo de utilização que permite os utilizadores colaborarem no processo de criação e manutenção dos <i>master data</i> e metadata associados (Berson, Dubov 2011, p. 1106).	Lidar com os processos que suportam as operações de criação, definição, melhorias e aprovação de <i>master data</i> ajudando a obter consenso entre os participantes dos processos (Dreibelbis et al, 2008, p. 902).
Operacional	Modelo de utilização que permite que os <i>master data</i> sejam recolhidos, alterados e utilizados nas transacções de negócio (Berson, Dubov, 2011, p. 1106).	Actuar como um sistema OLTP respondendo aos pedidos de múltiplas aplicações e utilizadores (Dreibelbis et al, 2008, p. 945) suportando as transacções.
Analítico	Modelo de utilização que suporta processos de negócio e aplicações que usam <i>master data</i> essencialmente para analisar a performance do negócio (Berson, Dubov, 2011, p. 1083).	Fornecer <i>master data</i> limpos e consistentes aos sistemas analíticos (Dreibelbis et al, 2008, p. 972).

ESTILOS DE IMPLEMENTAÇÃO

Os sistemas de MDM são implementados para aumentar a qualidade dos *Master Data* e fornecer uma utilização consistente desta informação naquele que é frequentemente um ambiente confuso e complexo. Há várias formas de implementar o MDM para ajudar a suportar estes dois requisitos de qualidade e partilha e ao mesmo tempo poder acomodar o conjunto de métodos de utilização.

Diferentes combinações de implementação e requisitos de utilização levaram à evolução de um número variado de estilos de implementação do MDM. São comuns as implementações híbridas que combinam múltiplos estilos. Alguns estilos são mais simples que outros. As organizações podem começar por implementar estilos simples para problemas urgentes e ir alargando a solução a outros estilos para resolver problemas mais complexos.

Os estilos de implementação mais comuns são: consolidação, registo, coexistência e *Hub* transaccional. À medida que os estilos progridem de Consolidação para *Hub* Transaccional as funcionalidades e a complexidade de implementar tendem a aumentar (Dreibelbis et al, 2008, p. 1090).

ESTILO CONSOLIDAÇÃO

Neste estilo, os *Master Data* provenientes das diversas bases de dados e sistemas de ficheiros são mantidos num único *Hub* MDM. Ao longo do percurso os dados são depurados, transformados e integrados para se obter um “golden record” para um ou mais domínios de *Master Data*. O “golden record” serve de fonte da verdade para os

sistemas analíticos de *Reporting* e como Sistema de Referência (SR) para outras aplicações.

Neste estilo, os dados limpos e consolidados fornecidos aos sistemas à jusante não podem ser alterados (Dreibelbis et al, 2008, p. 1108). A sincronização dos dados com os sistemas consumidores pode ser feita em *batch*.

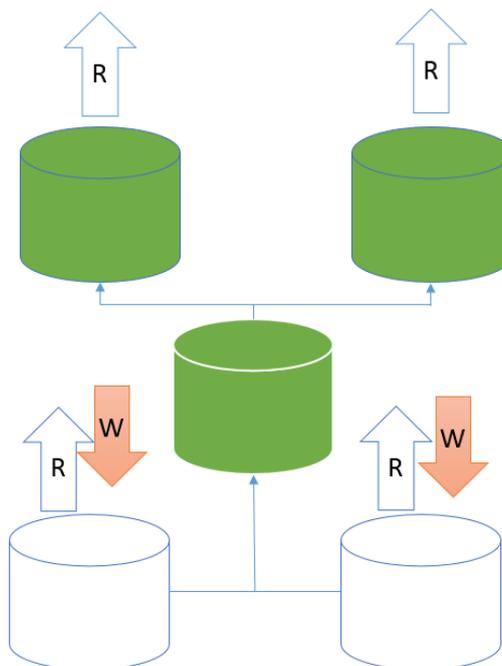


Ilustração 4 - MDM estilo Consolidação.

Fonte: Dreibelbis et al, 2008. (Dreibelbis et al, 2008, p. 1.114).

As alterações dos dados neste estilo são feitas nos sistemas de origem e, conseqüentemente, o MDM nem sempre tem os *Master Data* actualizados. Esta é a sua grande desvantagem. Este estilo pode ser o ponto de partida para um projecto MDM de várias fases (Dreibelbis et al, 2008, p. 1120).

ESTILO REGISTO

Neste estilo, apenas um conjunto mínimo de informação dos Master Data, por exemplo nome, primeira morada e NIF, é colocado no Hub central e a restante informação é mantida e actualizada nas aplicações fonte. Este conjunto mínimo de identificação no Hub central é usado para localizar os respectivos registos dos Master Data nas aplicações fonte (Dreibelbis et al, 2008, p. 1129). As principais características são:

- O MDM fornece aos consumidores as referências cruzadas para a informação detalhada que é mantida nas aplicações fonte;
- O MDM assegura a identificação e a qualidade da informação que possui e assume que a informação correlacionada residente nas aplicações é consistente.

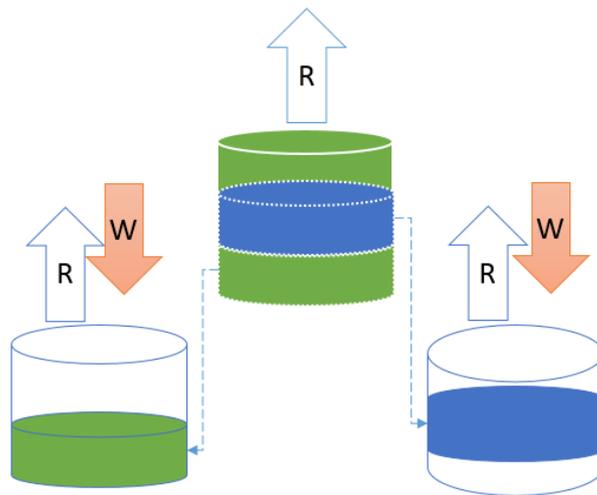


Ilustração 5 - MDM estilo Registro

Fonte: Dreibelbis et al, 2008. (Dreibelbis et al, 2008, p. 1136).

Neste estilo de implementação, a recolha de informação é feita em dois passos:

1. A informação de identificação dos registos é pesquisada no *Hub* MDM;
2. Usando a informação de identificação e a informação cruzada, é recolhida a informação relevante nas fontes.

As desvantagens deste estilo são:

1. As pesquisas podem envolver muitas fontes fazendo com que sejam muito complexas de implementar;
2. A entrada de mais uma aplicação na infra-estrutura, pode implicar a reconfiguração do modelo de dados do *Hub* MDM para poder acomodar os registos da nova aplicação.

Vantagens são as seguintes:

1. A informação detalhada, sendo fornecida pelas aplicações fonte, está sempre actualizada;
2. Em ambientes complexos, um grupo de trabalho não é obrigado a fornecer os todos os dados a outro grupo porque a informação pode ser acedida remotamente;
3. Pode ser relativamente fácil de implementar porque a responsabilidade da maior parte dos dados é mantida nas aplicações fonte.

ESTILO COEXISTÊNCIA

O estilo Coexistência envolve *Master Data* que podem ser criados e armazenados em vários locais e possui um “golden record” fisicamente instanciado no sistema de MDM que é sincronizado com os sistemas fonte. O “golden record” é construído da mesma forma como no estilo Consolidado, tipicamente através de importações batch (Dreibelbis et al, 2008, p. 1157).

Este estilo faz a combinação dos modelos Consolidação e Registo mantendo as características de cada um (Dreibelbis et al, 2008, p. 1165). As particularidades deste estilo são:

- Mantém a informação de identificação com a qual faz a referência aos registos nas aplicações fonte (tal como no modelo Registo);
- Replica os atributos mais importantes de cada entidade fonte no MDM;
- Uma cópia do *Master Data* é mantida nas aplicações fonte sincronizada com o “golden record” do MDM;
- As novas inserções e alterações de registos são feitas nas aplicações e enviadas ao MDM onde são transformadas em “golden record” e devolvidas novamente às aplicações fonte.
- A sincronização pode ser feita em “batch” (tal como no estilo Consolidação);
- Os dados podem ser pesquisados e actualizados directamente no *Hub* MDM;
- As actualizações feitas no MDM podem ser enviadas simultaneamente aos sistemas a montante e a jusante.

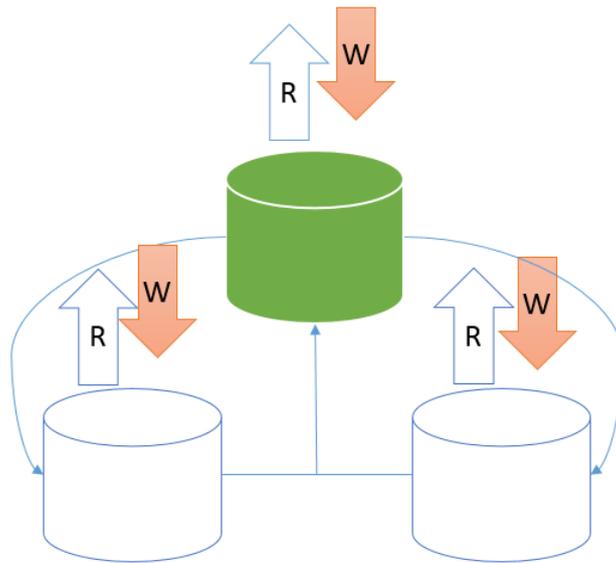


Ilustração 6 - MDM estilo Coexistência.

Fonte: Dreibelbis et al, 2008. (Dreibelbis et al, 2008, p. 1165).

O conflito entre actualizações é um dos problemas que este estilo pode enfrentar uma vez que um mesmo dado pode estar a ser actualizado simultaneamente no MDM e numa aplicação fonte (Dreibelbis et al, 2008, p. 1172).

ESTILO HUB TRANSACCIONAL

Este estilo centraliza um conjunto completo de *Master Data* para uma ou mais áreas de domínios. Actua como SOR dos *Master Data* que gere. As cópias dos *Master Data* do MDM são mantidas sincronizadas com os sistemas aplicativos tal como no modelo Coexistência (Dreibelbis et al, 2008, p. 1172). As particularidades deste estilo são:

- As aplicações fonte não inserem e não alteram os *Master Data*;
- Apenas o MDM pode criar novos dados e fazer alterações;
- O fluxo dos dados alterados faz-se num só sentido: do MDM para as aplicações.

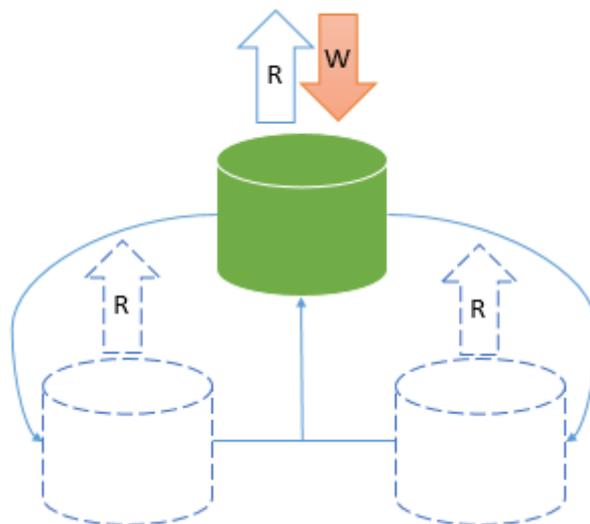


Ilustração 7 - MDM estilo *Hub Transaccional*.

Fonte: Dreibelbis et al, 2008. (Dreibelbis et al, 2008, p. 1187)

O *Hub Transaccional* pode centralizar exclusivamente a gestão dos *Master Data* de um ou mais domínios (Dreibelbis et al, 2008, p. 1172). Ao centralizar exclusivamente a gestão dos *Master Data* de todos os sistemas de entrada (*Systems of Entry*), o *Hub Transaccional* torna-se simultaneamente num SOR e SR (White, 2007. p. 8).

Tabela 9 - Estilos de implementação.

Estilo	Consolidação	Registro	Coexistência	Hub Transaccional
Objectivo	Manter os <i>master data</i> concisos e precisos num único repositório para <i>Reporting</i> .	Proporcionar um SOR mínimo no MDM com link para informação detalhada residente nas aplicações.	Manter uma visão única dos <i>master data</i> no MDM sincronizado com outros sistemas.	Fazer a gestão exclusiva dos <i>master data</i> fornecendo acesso através de serviços.
Vantagens	Bom para preparar e manter uma lista limpa de <i>master data</i> para BI e <i>Reporting</i> .	Proporciona uma visão mais detalhada dos dados apenas quando é necessário; É de rápida implementação.	Intrusão mínima nas aplicações fonte quando é necessário fazer alterações.	Suporta aplicações transaccionais actuais e futuras.
Desvantagens	É “ <i>read-only</i> ” e nem sempre está actualizado porque os dados já podem ter sido modificados nas aplicações de origem.	Pode levar a construção de pesquisas distribuídas muito complexas.	Há possibilidade de haver conflitos durante a actualização dos dados porque são modificados no MDM e nas aplicações.	Pode exigir alterações de funcionalidades nos sistemas abrangentes.

Fonte: Adaptado de Dreibelbis et al. (2008).

2.4.5. MDM: ABORDAGEM POR TIPO DE DADOS

A outra forma de identificar e caracterizar os sistemas de MDM é através dos dados (operacionais ou não operacionais) sobre os quais recaem as operações. Nesta perspectiva, de acordo com o Cervo e Allen (2011, p. 491), o MDM pode ser classificado em:

1. MDM Analítico - Abrange os dados não operacionais e suporta as operações de apoio à decisão;
2. MDM Operacional - Abrange os dados operacionais e suporta as actividades operacionais;
3. MDM Empresarial - Este modelo suporta simultaneamente as operações de apoio à tomada de decisão e as actividades operacionais da organização.

MDM ANALÍTICO

Historicamente, o MDM Analítico foi a abordagem mais adoptada pela sua relativa simplicidade de impulsionar projectos de DW. Tendo como peça central a DW, não são necessários implementar todos os serviços típicos do MDM, nomeadamente *Data Quality*, administração e governança de dados (Cervo e Allen, 2011, p. 516). Sendo esta a grande vantagem desta abordagem.

A desvantagem desta abordagem reside no facto de os sistemas operacionais não recolherem os benefícios do MDM (Cervo e Allen, 2011, p. 548) porque este não devolve os dados corrigidos aos sistemas operacionais de origem.

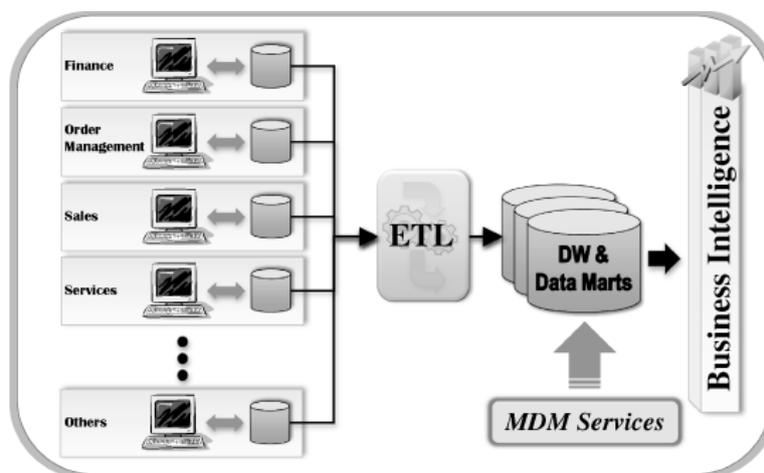


Ilustração 8 - MDM Analítico.

Fonte: Cervo e Allen, 2011. (Cervo e Allen, 2011, p. 543).

MDM OPERACIONAL

O MDM Operacional é implementado para garantir a consistência semântica dos *Master Data* das actividades transaccionais (Berson e Dubov, 2011, p. 1087) com as seguintes vantagens e desvantagens:

- Como vantagem podemos referir que permite a oportunidade de consolidar os dados operacionais dispersos na empresa e tornar-se num Sistema de Referência (SR) (Cervo e Allen, 2011, p. 564).
- Tem a desvantagem de poder ser necessário fazer alterações na infra-estrutura tecnológica da empresa cujas dificuldades crescem com a dimensão do volume dos dados e dos sistemas dispersos a integrar (Cervo e Allen, 2011, p. 564).

A implementação pode ser feita gradualmente em etapas previamente definidas. Em cada etapa é migrado um sistema para dentro do repositório do MDM até que todos os sistemas abrangidos no âmbito do projecto sejam integrados (Cervo e Allen, 2011, p. 572).

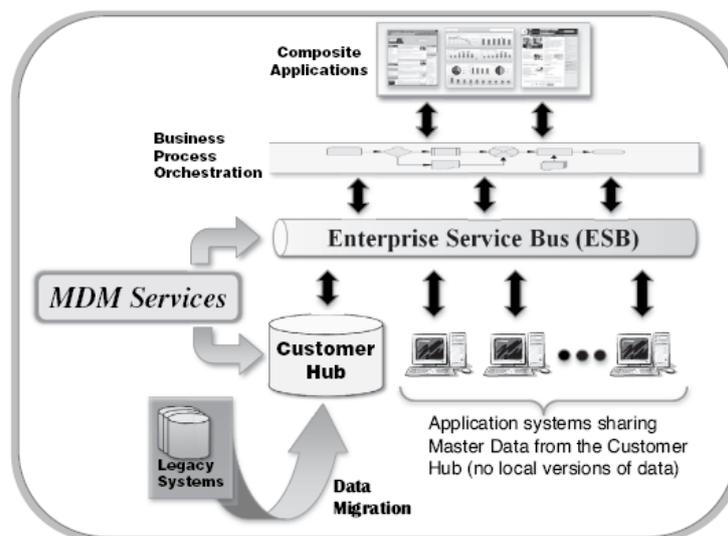


Ilustração 9 - MDM Operacional.

Fonte: Cervo e Allen, 2011. (Cervo e Allen, 2011, p.610).

Depois do sistema central (central *Hub*) estar pronto, com todos os sistemas actuais (*legacy systems*) integrados, os novos sistemas não necessitam de ter versões locais dos *Master Data* porque passam a consumir directamente do central *Hub* através de um serviço de integração de dados (Cervo e Allen, 2011, p. 610).

MDM EMPRESARIAL

O MDM Empresarial é a combinação do MDM Operacional com o MDM Analítico e como tal, pode ser implementado através da combinação destas duas arquiteturas. Este modelo tem duas vantagens:

- A manutenção da DW torna-se mais fácil porque a maior parte dos serviços de MDM necessário na DW já existem no sistema operacional;
- Com a integração de grande parte das aplicações no MDM operacional, o funcionamento do ETL do sistema analítico fica mais simplificado porque há poucas aplicações para extrair dados.

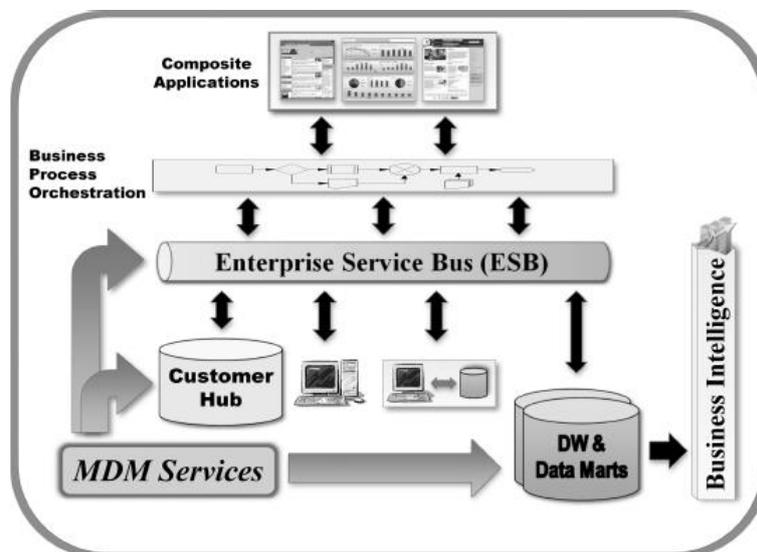


Ilustração 10 - MDM Empresarial.

Fonte: Cervo e Allen, 2011. (Cervo e Allen, 2011, p. 665)

RESUMO

Para implementar um sistema de MDM pode ser necessário fazer alterações nas infra-estruturas e/ou nas aplicações o que pode levar a interrupção temporária dos serviços da empresa.

Por um lado, os dados operacionais são actualizados e consumidos em tempo real pelos processos e aplicações que suportam as actividades da empresa, ao passo que os sistemas de apoio à decisão consomem dados diferidos que se encontram armazenados em sistemas intermédios. Por outro lado, o MDM Analítico não altera os dados e, por este motivo, o impacto das paragens das aplicações na actividade da

empresa é menor no MDM Analítico e muito maior no MDM Empresarial, conforme ilustrado no gráfico abaixo:

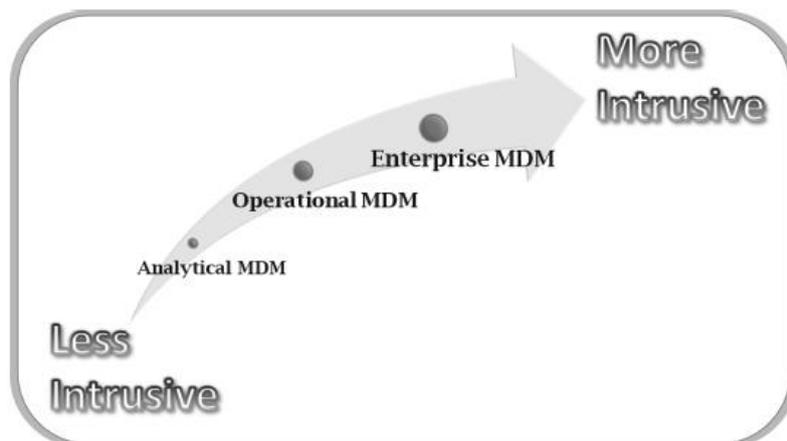


Ilustração 11 - Abordagem por tipo de dados.

Fonte: Cervo e Allen, 2011. (Cervo e Allen, 2011, p.502)

2.4.6. ARQUITECTURA DE REFERÊNCIA

Uma Arquitectura de Referência (*Reference Architecture*) serve para disponibilizar uma perspectiva global do conjunto de componentes, serviços, processos e interfaces de um sistema, organizada em camadas funcionais, onde cada camada fornece serviços à camada de cima e consome serviços da camada de baixo (Berson e Dubov, 2011, p. 3103).

Segundo Dreibelbis et al (2008, p. 2586), a arquitectura de referência de MDM descreve a estrutura através da apresentação das camadas que a compõem, da hierarquia dos componentes e da forma como os componentes trabalham em conjunto para fornecer capacidades técnicas e de negócio.

A importância de uma arquitectura de referência consiste em descrever as funcionalidades, a localização de todos os componentes e as suas interações para qualquer sistema (Berson e Dubov, 2011, p. 3103).

Retirando algumas funcionalidades muito específicas e padrões de implementação próprios, a arquitectura de um *Hub* de dados de um MDM de clientes é muito similar ao do domínio de produtos e de outros mais (Berson e Dubov, 2011, p. 3096).

De acordo com (Dreibelbis et al, 2008, p. 3154), uma arquitectura de referência de MDM é desenhada para:

- Suportar múltiplos padrões de utilização do MDM (Criação Colaborativa, Operacional e Analítica);
- Suportar o CDI e PIM e MDM de outros domínios;
- Manter as relações entre os domínios de dados e fornecer funcionalidades para manter uma fonte oficial dos *Master Data* da empresa;
- Ser escalável com alta disponibilidade e extensível;
- Fornecer flexibilidade para integrar tecnologias de várias fabricantes e para integrar-se com futuras iniciativas ainda desconhecidas.

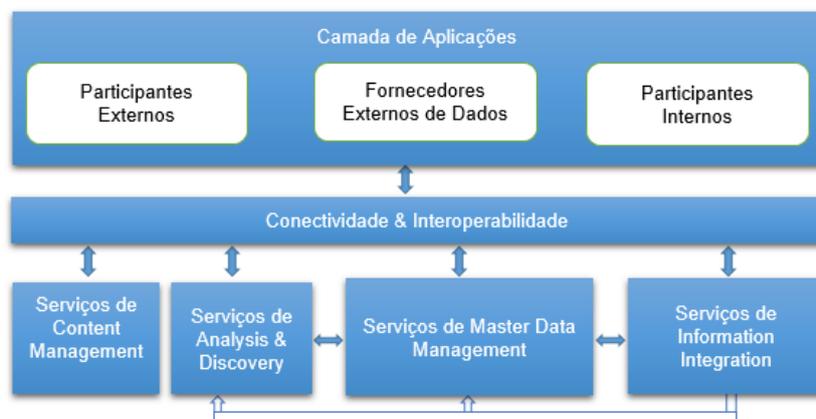


Ilustração 12 - Uma Arquitectura de Referência.

Adaptação de Dreibelbis e Oberhofer (2008)

Nesta arquitectura de referência as aplicações da camada de aplicação podem requisitar os serviços de *Content Management*, *Analysis & Discovery*, *Master Data Management* e *Information Integration* através dos serviços de *Conectividade & Interoperabilidade*. Esta funcionalidade permite que qualquer iniciativa na organização possa requisitar os serviços da arquitectura (Dreibelbis et al, 2008, p. 3154).

De forma análoga os serviços do MDM podem requisitar os serviços de *Analysis & Discovery* bem como outros serviços para assegurar a qualidade dos *Master Data* manipulados no *Hub MDM*.

Os serviços de *Information Integration* inicializam os repositórios do MDM e podem também ser solicitados pelos serviços de *Analysis & Discovery* para fornecerem dados externos durante o processo de descoberta de identidades relacionados com os *Master Data*.

BLOCO: APLICAÇÕES DE NEGÓCIO

Este bloco é composto pelos seguintes serviços:

- **PARTICIPANTES EXTERNOS**

Serviços que disponibilizam funcionalidades que permitem os clientes e agentes externos acederem e actualizarem os seus *Master Data*. Por exemplo, os clientes de uma empresa podem actualizar os *Master Data* das compras *online* ou acederem aos serviços de *homebanking* (Dreibelbis e Oberhofer, 2008).

- **PARTICIPANTES INTERNOS**

Nesta estrutura encontram-se as aplicações e os sistemas da organização produtores e consumidores dos *Master Data*.

- **FORNECEDORES EXTERNOS DE DADOS**

Esta camada disponibiliza os serviços de *Reference Data* que podem ser usados pelas aplicações. Por exemplo, os serviços governamentais podem fornecer listas actualizadas de combate ao terrorismo e combate ao branqueamento de capitais (Dreibelbis e Oberhofer, 2008) que podem ser usadas em aplicações de negócio.

CAMADA: CONECTIVIDADE E INTERACTIVIDADE

É a camada que disponibiliza os serviços de troca de dados entre empresas, sistemas dentro da empresa e fornecedores externos. Pode ser implementada usando as arquitecturas de integração de dados e aplicações, como por exemplo Enterprise Application Integration (EAI) ou Enterprise Service Bus (ESB) (Dreibelbis e Oberhofer, 2008).

BLOCO: SERVIÇOS DE INFORMATION INTEGRATION

Os serviços de *Information Integrity* têm como objectivo resolverem aspectos relacionados com a integridade dos *Master Data* usando técnicas de *profiling*, *Matching* e padronização entre outras (Dreibelbis et al, 2008, p. 3120).

BLOCO: SERVIÇOS DE MASTER DATA MANAGEMENT

Constitui o núcleo principal da arquitectura e é composto por serviços que garantem a qualidade dos dados e os disponibilizam a toda a organização.

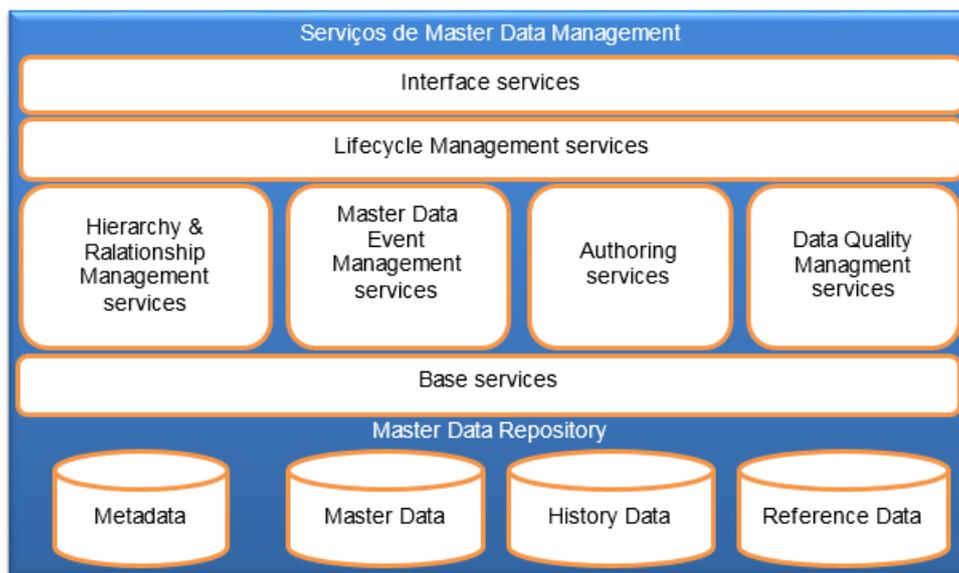


Ilustração 13 - Serviços de MDM.

Adaptação de Dreibelbis e Oberhofer, 2008)

- **SERVIÇOS DE INTERFACE**

Constituem os pontos de entrada através dos quais outros serviços requisitam os serviços desta camada usando técnicas tais como web services, processamento em batch e chamadas de métodos.

- **SERVIÇOS DE *LIFECYCLE MANAGEMENT***

A forma como os *Master Data* são criados, lidos, actualizados, pesquisados e excluídos descrevem o ciclo de vida dos *Master Data* e é conhecida por CRUD (*Create, Read, Update e Delete*). Os serviços de *Lifecycle Management* têm por objectivo suportar o CRUD dos *Master Data* do MDM (Dreibelbis et al., 2008, p. 2757).

Os serviços de *Data Quality Management* são invocados pelos serviços do *Lifecycle Management* para garantir a limpeza, padronização e reconciliação dos dados.

- **SERVIÇOS DE *MASTER DATA EVENT MANAGEMENT***

Estes serviços fornecem as funcionalidades que definem as regras de negócio para detectar, dentro do MDM, os eventos, as condições dos eventos e as respectivas acções a serem tomadas quando ocorrerem (Dreibelbis et al., 2008, p. 3318).

- **SERVIÇOS DE *HIERARCHY & RELATIONSHIP MANAGEMENT***

As listas dos materiais que compõem os produtos ou a estrutura de vendas de uma empresa são exemplos de hierarquia de dados geridos no MDM.

Os Serviços de *Hierarchy & Relationship Management* permitem criar e gerir as relações entre os *Master Data*, organizar as entidades dos *Master Data* em hierarquias e agrupamentos e criar múltiplas *view* de hierarquias de *Master Data* (Dreibelbis et al., 2008, p. 3311). Estes serviços podem ser também requisitados pelos serviços de *Identity Analysis* (bloco de *Analysis & Discovery*) e utilizados para descobrir relações não óbvias que possam existir entre *Master Data*, tais como ligações familiares (Dreibelbis et al., 2008, p. 3341).

- **SERVIÇOS DE *AUTHORING***

Serviços usados para criar e ampliar as definições das entidades dos *Master Data*, criar e actualizar as entidades no MDM, definir as hierarquias, os relacionamentos e os agrupamentos de *Master Data*.

Estes serviços fornecem funcionalidades que podem ser usadas por outros serviços para implementar interfaces de utilizadores para sistema de MDM ou para administração de dados mais eficientes e fáceis de implementar porque as funcionalidades são suportadas por serviços que criam, pesquisam, actualizam e apagam os *Master Data* no MDM (Dreibelbis et al., 2008, p. 3290).

- **SERVIÇOS DE *DATA QUALITY MANAGEMENT***

São serviços que aplicam técnicas para validar e reforçar as regras de governança e automatização dos processos de limpeza de dados. Estes serviços podem requisitar os serviços de *Information Integrity* da estrutura *Information Integration Services* (Dreibelbis et al., 2008, p. 3243).

- **SERVIÇOS DE *BASE***

São serviços que têm como objectivo disponibilizar aos componentes IT externos, através de interfaces de MDM, funcionalidades para implementar a privacidade, pesquisa, auditoria de acesso e *workflow*. Estes serviços são implementados para se integrarem com componentes comuns que suportam o *workflow*, segurança e auditoria de acesso (Dreibelbis et al., 2008, p. 3365).

- ***MASTER DATA REPOSITORY***

Um conjunto de repositórios que suportam os metadados dos *Master Data*, os dados dos *Master Data*, os dados de *Reference Data*, os históricos das alterações efectuadas nos *Master Data* e dados transaccionais ao longo do tempo (Dreibelbis et al., 2008, p. 3414).

A inicialização destes repositórios é feita através dos Serviços de *Information Integration* que garantem a integridade dos dados.

BLOCO: SERVIÇOS DE ANALISYS & DISCOVERY

Este bloco é composto pelos seguintes serviços:

- **SERVIÇOS DE *IDENTITY ANALISYS***

Serviços que têm como objectivo descobrir a verdadeira identidade de uma pessoa, descobrir as relações não óbvias entre pessoas, tais como daquelas que vivem na mesma casa mas têm nomes e endereços diferentes ou descobrir relações entre pessoas e organizações (Dreibelbis et al., 2008, p. 3488).

- **SERVIÇOS DE *OPERATIONAL INTELLIGENCE***

Estes serviços fornecem funcionalidades baseadas em eventos analíticos que permitem as empresas ficarem dotados de capacidades de formulação de cenários de análises para mais facilmente se adaptarem às alterações de mercado (Dreibelbis et al., 2008, p. 3144).

- **SERVIÇOS DE *QUERY, SEARCH E REPORTING***

Estes serviços fornecem funcionalidades para a realização de pesquisas “*ad hoc*”, *Reporting* e criação de modelos multidimensionais de dados de negócio (Dreibelbis e Oberhofer, 2008).

- **VISUALIZATION**

Esta componente fornece funcionalidades de construção de mapas, gráficos, serviços de análise espacial e serviços de *rendering* para a interacção com os componentes que fornecem serviços de apresentação ao nível do utilizador (Dreibelbis, Oberhofer e 2008).

- ***DATA & METADATA REPOSITORY***

Repositório para onde são carregados dados de fontes externas que, em conjunto com os dados internos da empresa, permitem os serviços de *Identity Analytics* descobrirem a verdadeira identidade de pessoas e suas relações. Estes dados externos são fornecidos através dos Serviços de *Information Integration* cujos serviços garantem a integridade dos dados (Dreibelbis e Oberhofer, 2008).

BLOCO: SERVIÇOS DE CONTENT MANAGEMENT

Os serviços aqui contidos fornecem funcionalidades para capturar, agregar, pesquisar, catalogar, garantir a segurança, aprovar e gerir conteúdos não estruturados em vários

formatos (imagens, documentos de textos, páginas web, folhas de cálculo, gráficos, *email*, vídeo entre outros).

Os Serviços de MDM podem solicitar os Serviços de *Content Management* para acederem aos conteúdos não estruturados, como por exemplo imagens associadas aos *Master Data* (Dreibelbis e Oberhofer, 2008).

2.5. GOVERNANÇA DE DADOS

“MDM without Data Governance is ‘data integration’, not MDM!”

(Gartner, 2012).

Um componente fundamental na implementação, com sucesso, de um sistema de MDM é a Governança de Dados (GD). A GD estabelece as regras para a utilização dos dados numa organização. Exemplos de regras de governança de dados são: campos obrigatórios, valores padrão, encriptação, requisitos de segurança, políticas de retenção de dados, etc. (Sql Server 2008 R2).

Berson e Dubov (2011, p. 9511) definem a GD como um processo de “orquestração formal de pessoas, processos e tecnologias numa organização para transformar os dados em activos”. A GD é a cola dentro do MDM segundo Cervo, Allen (2011, p. 1605), e que, segundo Berson e Dubov (2011, p. 9465) sem ela “é impossível implementar e operar com sucesso uma solução de MDM”.

FRAMEWORK DE GOVERNANÇA DE DADOS

Loshin (2009, p. 680) afirma que sendo o MDM uma iniciativa que abrange toda a organização, integrando pessoas, ferramentas, processos e políticas que moldam a forma como as organizações procuram obter mais-valias dos activos de informação de que dispõem, é necessária uma framework de Governança de Dados forte que ajude a organização a transitar para uma cultura proactiva de garantia de qualidade.

Por outro lado, a implementação não atempada de um programa de governança de dados é um dos primeiros e mais sérios erros que se comete quando se desenvolve um sistema de MDM (Cervo e Allen, 2011, p. 1644).

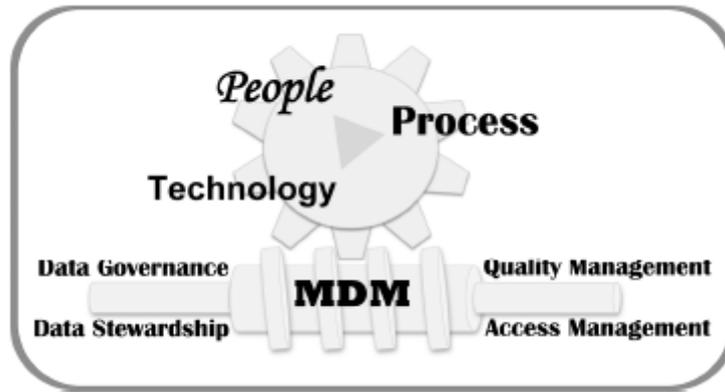


Ilustração 14 - MDM é um sistema.

Fonte: Cervo e Allen, 2011, (Cervo e Allen, 2011, p. 1095)

Do que foi dito no parágrafo anterior resulta que o primeiro desafio para implementar um programa de GD deve portanto ser instituir uma framework de governança de dados que de acordo com Berson e Dubov (2011, p. 9531) deve ser repartida nas seguintes áreas:

- Estratégia de GD - Área encarregue de definir a missão, o planeamento e os objectivos estratégicos da GD;
- Organização da GD - Tem a responsabilidade de criar um “Conselho de Governança” para dirigir o programa de GD e garantir o seu alinhamento com os objectivos de negócio da organização;
- Políticas de GD - Área com a responsabilidade de definir as políticas e as regras de GD;
- Processos de GD - Área encarregue de definir os processos e procedimentos que dirigem e controlam a distribuição das estratégias e das políticas do programa de GD;
- Monitorização e Investigação - Área responsável por controlar os dados que provocam falhas no cumprimento das políticas e métricas de GD.
- Arquitectura e Tecnologia - Área responsável pelas ferramentas e arquitecturas tecnológicas para a gestão operacional da GD.

PROGRAMA DE GOVERNANÇA DE DADOS

Um programa de GD tem por missão definir os processos, as métricas e as ferramentas que permitem melhorar continuamente a qualidade, segurança, privacidade, pontualidade, conformidade e a adaptabilidade dos *Master Data* numa organização (Berson e Dubov, 2011, p. 9608). O seu desenvolvimento deve compreender as duas fases seguintes: Planeamento e Implementação (Cervo e Allen, 2011, p. 1640):



Ilustração 15 - Programa de Governança de Dados.

Adaptação de: Cervo e Allen (2011, p. 1640)

Na fase de Planeamento e Desenho são definidos os propósitos fundamentais, que consistem em (Loshin, 2009, p. 1512):

- Clarificar a arquitectura de informação da organização;
- Mapear as funções das aplicações aos objectivos de negócio;
- Instituir uma estrutura de processos para as políticas de informação.

De acordo com Loshin (2009, p. 1521), para a clarificação da arquitectura de informação é necessário documentar, através de um levantamento, os dados que existem na organização, a forma como são geridos, usados e como suportam os sistemas aplicativos. É um inventário que deve identificar também os locais na organização onde a existência de insuficiências ou redundâncias de dados pode criar obstáculos à supervisão da GD.

Este levantamento serve para fornecer uma visão global e real do estado dos *Master Data* e das aplicações operacionais. É um exercício que pode permitir revelar a existência de repositórios de dados desconhecidos ao departamento das TI (Wolter e Haselden, 2006).

O programa de GD deve ser dinâmico. Segundo Cervo e Allen (2011, p. 1821), ao longo do tempo é necessário ajustar o plano e o papel dos elementos que integram o programa para ficar alinhado com as alterações que possam tipicamente ocorrer na organização.

POLÍTICAS, PADRÕES E CONTROLO:

Na definição das políticas, regras e padrões de qualidade é necessário obter a participação do Conselho de GD. Este envolvimento é necessário para garantir a influência do Conselho de GD nos processos de recolha dos dados, para assegurar os padrões de qualidade dos dados, as regras de validação e as expectativas de controlo dos dados (Cervo e Allen, 2011, p. 1832).

FASE DE IMPLEMENTAÇÃO:

Nesta fase deve ficar definido um plano de comunicação que assegure a difusão dos objectivos e o lançamento do programa de GD. Neste plano o patrocinador executivo (*executive sponsor*), os membros do Conselho de GD e a principal equipa do MDM devem ficar bem identificados (Cervo e Allen, 2011, p. 1960):

- Os pedidos ao Conselho de GD devem ser sólidos para que as respostas sejam apropriadas e rápidas (Cervo e Allen, 2011, p. 1971);
- As principais equipas de MDM, os administradores de dados e analistas devem fazer parte da rede do Conselho de GD (Cervo e Allen, 2011, p. 1971).

No processo de validação é necessário assegurar que o programa de GD está pronto para ser implementado e usado de acordo com as necessidades da organização. É necessário também assegurar que todas as equipas e os principais intervenientes estão suficientemente familiarizados com a terminologia e com as práticas de MDM (Cervo e Allen, 2011, p. 2002).

Na etapa de execução é necessário:

- Estar atento ao evoluir da qualidade dos dados e aos eventos operacionais que podem afectar negativamente ou positivamente os *Master Data* (Cervo e Allen, 2011, p. 2012);
- Manter continuamente informadas as partes interessadas das acções e eventos na GD (Cervo e Allen, 2011, p. 2012);
- Manter os membros das equipas de trabalho activamente engajados na execução do programa (Cervo e Allen, 2011, p. 2012).

Com o decorrer do tempo, a GD atinge um nível elevado de maturidade em que a aceitação do MDM na organização será um facto. A partir desse momento o Conselho

de Governança não precisará de se reunir com frequência mas deverá continuar a manter um controlo sobre o MDM (Cervo e Allen, 2011, p. 2021).

Na etapa de Manutenção e Melhorias:

- O Conselho de GD deve rever periodicamente as métricas e os indicadores de performance na perspectiva de qualidade e quantidade (Cervo e Allen, 2011, p. 2043);
- É necessário manter o processo de GD num estado de equilíbrio permanente, introduzindo melhorias na qualidade dos dados e aumentando a eficácia na gestão dos dados;
- A entrada dos dados deve ser monitorizada para detectar e corrigir hábitos negativos nos processos de introdução (Cervo e Allen, 2011, p. 2106).

3. IMPLEMENTAÇÃO DO CONCEITO

Este capítulo descreve a implementação do sistema de MDM na Agência Portuguesa do Ambiente, I.P. (APA, I.P.) como prova dos conceitos de MDM. A prova materializou-se com a instalação e configuração da infra-estrutura tecnológica e com a implementação dos serviços da solução. Com este projecto foi possível responder às Questões de Investigação colocadas no início deste trabalho.

Começamos por descrever resumidamente o que é a APA I.P, mas antes devemos salientar que a informação aqui apresentada foi obtida consultando a documentação interna, a informação disponível no sítio de internet da APA e com a realização de entrevistas. As entrevistas foram dirigidas às seguintes pessoas e entidades:

- O chefe de Divisão de Tecnologias de Informação (DTI), Dr. Luís Duarte (luis.duarte@apambiente.pt), que apresentou uma visão geral dos sistemas e dos projectos em curso. Descreveu detalhadamente o problema das entidades no Sistema de Licenciamento Ambiental (SILiAmb); descreveu a nova arquitectura de informação da APA; destacou a importância do projecto MDM para a APA no contexto do programa de reestruturação dos Sistemas de Informação da APA;
- A Directora do Departamento de Recursos Hídricos (DRH), Dr^a Felisbina Quadrado (maria.quadrado@apambiente.pt) que descreveu as consequências da má qualidade dos dados nos processos de negócio do departamento, nomeadamente no processo de cobrança das taxas de recursos hídricos (TRH);
- Os gestores internos da aplicação de gestão documental (FileDoc), bem como os operadores e utilizadores que descreveram as dificuldades na utilização desta aplicação como consequências da má qualidade dos dados. Estes elementos contribuíram também na produção de um Manual de Procedimentos no âmbito do Programa de Governança de Dados;
- Com o apoio da *Webuild*, a empresa que desenvolveu o FileDoc, foi possível entender as regras de negócio, os webservices disponíveis e o modelo entidade-associação desta aplicação.

3.1. AGÊNCIA PORTUGUESA DO AMBIENTE

A Agência Portuguesa do Ambiente, I.P. (APA) resulta da fusão de vários organismos regionais e nacionais.

É um organismo da administração central do estado que tem como missão: “propor, desenvolver e acompanhar a gestão integrada e participada das políticas do ambiente e do desenvolvimento sustentável, de forma articulada com outras políticas sectoriais e em colaboração com entidades públicas e privadas que concorrem para o mesmo fim, tendo em vista um elevado nível de protecção e de valorização do ambiente e a prestação de serviços de elevada qualidade aos cidadãos” (Decreto-Lei n.º 56/2012, de 12 de Março).

A APA é autoridade nos domínios do ambiente, água, seguranças das barragens, resíduos e poluição. Tem a sua sede em Lisboa e delegações no Porto, Coimbra, Évora e Faro. Conta com a participação de mais de 800 colaboradores.

A APA é dirigida por um Conselho Directivo e tem como dirigente máximo um presidente, coadjuvado por uma vice-presidente e duas vogais. O Departamento de Tecnologias e Sistemas de Informação (DTSI) está sob direcção da vice-presidente. O DTSI é composto por duas divisões: a Divisão de Tecnologias de Informação (DTI) e a Divisão de Sistemas de Informação (DSI).

3.2. ENQUADRAMENTO DO PROJECTO

Na APA, o termo **entidade** é usado para designar genericamente os registos de cidadãos e empresas. Neste trabalho **entidade** pode significar registo de clientes e empresas e também pode significar tabelas de base de dados.

A existência de dados de entidades de má qualidade e redundantes na APA, era uma herança dos antigos sistemas regionais. O processo de consolidação mal sucedido nas entidades agravou a situação. Como solução para resolver este problema a APA optou por implementar um projecto de gestão centralizada para as entidades, baseado nos conceitos de MDM. Ao projecto designou-se por Sistema de Gestão de Entidades (SGE) escolhendo MDM de **Coexistência** como estilo de implementação.

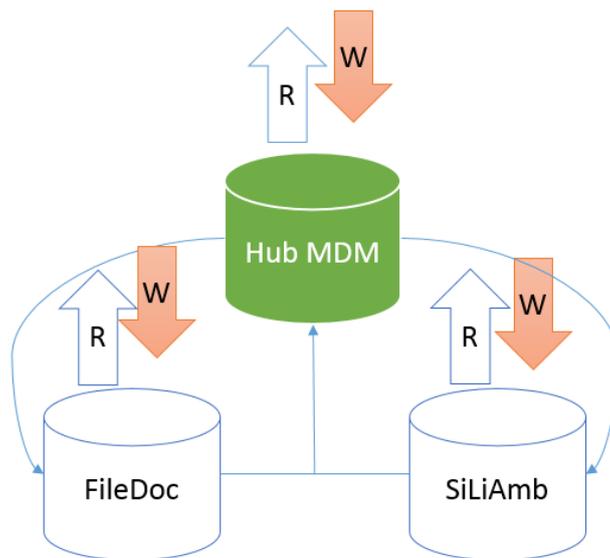


Ilustração 16 - Estilo de implementação MDM da APA

(Coexistência).

O SGE tinha como objectivos principais e imediatos:

- (1) Limpar e eliminar os registos redundantes das duas aplicações operacionais de referência (FileDoc e SiLiAmb);
- (2) Implementar um sistema de garantia de qualidade dos *master data* baseado num Hub Central;
- (3) Implementar um mecanismo de partilha de dados.

Sendo a APA uma organização de grandes dimensões com alguma complexidade orgânica e atendendo que a sua infra-estrutura tecnológica encontram-se em fase de reestruturação, o SGE foi planeado para ir crescendo com a integração faseada dos variados *Master Data*.

Tendo como âmbito as entidades do FileDoc, o presente trabalho teve como objectivo:

1. Limpar e eliminar os registos redundantes;
2. Implementar o sistema MDM; e
3. Implementar uma solução para a partilha dos *master data* na APA;

3.3. IMPLEMENTAÇÃO DO PROJECTO

Este subcapítulo descreve as fases que compuseram a implementação do MDM cujo plano é apresentado na figura abaixo.

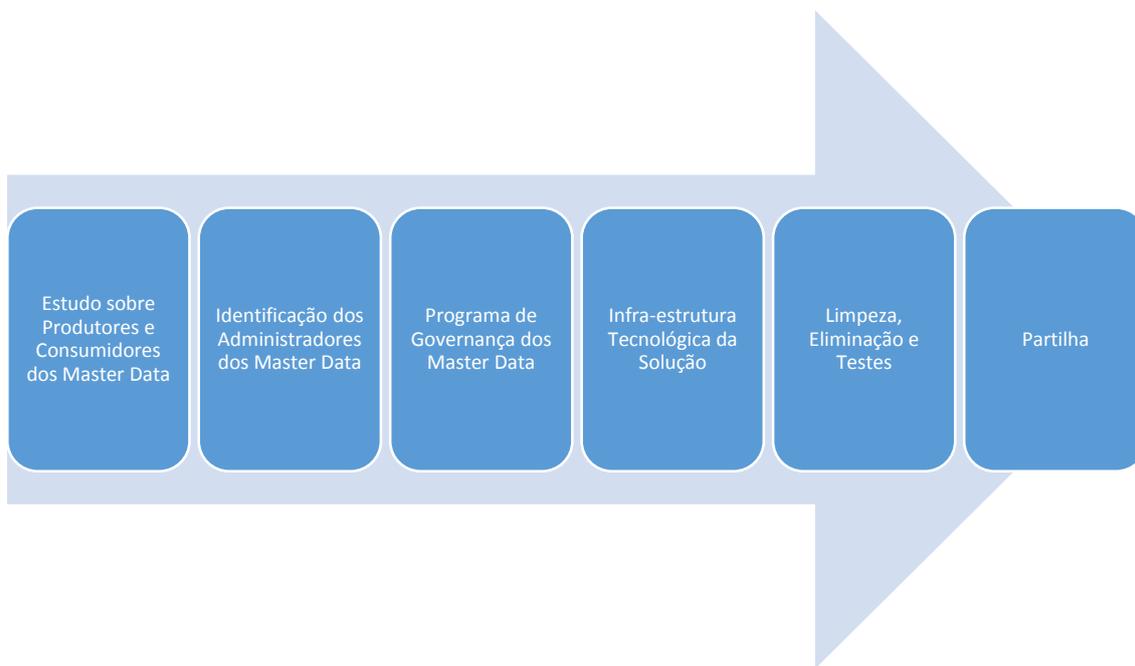


Ilustração 17 - Planeamento do projecto MDM da APA.

3.3.1. ESTUDO SOBRE PRODUTORES E CONSUMIDORES DOS *MASTER DATA*

Sendo esta a primeira fase, veio a tonar-se fundamental porque permitiu documentar a real situação sobre os principais dados da APA.

Durante o processo de inventariação foi frequente encontrar pequenas listas de entidades dispersas com nomes, moradas e contactos que não eram do conhecimento do Departamento de Informática.

As quatro principais tarefas desta fase foram:

- (1) Identificação e análise dos repertórios dos *Master Data*;
- (2) Identificação dos produtores e consumidores dos *Master Data*;
- (3) Análise ao estado dos *Master Data*;

(1) IDENTIFICAÇÃO E ANÁLISE DAS FONTES DOS MASTER DATA

FileDoc consiste no principal reportório de entidades, continha perto de 400.000 registos. Este reportório é suportado em Sql Server. Os atributos das entidades são: Nome, Morada, NIF e Contacto (telefone, fax e mail).

A base de dados do SILiAmb também tem centenas de milhares de entidades suportados em PostgreSQL. O esquema de dados do SILiAmb é tecnicamente mais complexo devido às suas regras de negócio. Os principais atributos das entidades são: Nome, Morada, NIF e Contacto (telefone, fax e mail)

Como *master data* de entidades da APA ficaram definidos os seguintes atributos e formatos:

Tabela 10 - *Master Data* de Entidades da APA.

Atributo	Tipo de Dado	Descrição
Name	Texto	Nome completo
Address	Texto	Endereço
PostalCode	Texto	Código Postal
PostalCodeDescription	Texto	Local Postal
TaxNumber	Texto	Número de contribuinte
Email	Texto	Endereço electrónico
Fax	Texto	Fax
Telephone	Texto	Telefone, telemóvel
City	Texto	Nome de Cidade

(2) IDENTIFICAÇÃO DOS PRODUTORES E CONSUMIDORES DOS *MASTER DATA*

Seguiram-se as seguintes perguntas: “que aplicações criam e modificam os *Master Data* identificados acima?”, “quais as aplicações que fazem uso destes *Master Data*?”

As respostas a estas questões ficaram facilitadas pela opção da APA em só integrar nesta etapa do SGE as aplicações que contribuem para os objectivos estratégicos da APA, para este trabalho o âmbito é o FileDoc.

Produtores:

- O sistema BPM para a desmaterialização dos processos de compras públicas em fase concursal de aquisição de serviços de implementação;

- Sistema integrado ERP em fase de concurso de aquisição de serviços de implementação;
- Sistema único de Business Intelligence em fase de concurso de aquisição de serviços de implementação;
- As aplicações fontes (FileDoc, SILiAmb);
- Todas as áreas de negócio da APA;
- Todas as áreas administrativas da APA

ANÁLISE AO FILEDOC

O FileDoc entrou em produção em setembro de 2013. A partir dessa data tornou-se obrigatório o uso deste sistema para registo de documentos (internos, de entrada ou saída), processos, circulação e expedição dos documentos associados à actividade da APA.

No FileDoc a tabela “Users” contém a lista de entidades (Internas e Externas) e a view “Entities” filtra as entidades externas (pessoas singulares e empresas) que serão alvo deste trabalho. A tabela Address contém as moradas das entidades. Cada entidade pode ter uma ou mais moradas na tabela Address.

Consumidores:

Os principais consumidores são toda a APA porque uma parte da actividade da APA é desenvolvida interagindo junto de um conjunto muito vasto de entidades públicas, privadas, empresas e cidadãos. A interacção consiste maioritariamente na recepção e envio de documentos. Os documentos incluem taxas de cobrança, contra-ordenações, ofícios, facturas, propostas, pedidos de informação, esclarecimentos, comunicações, etc.

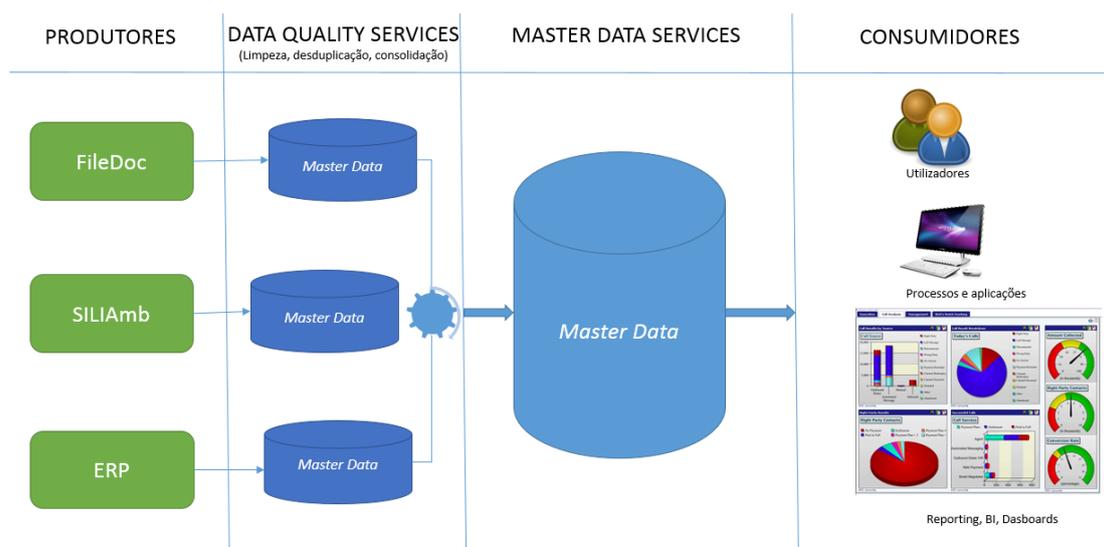


Ilustração 18 - Modelo conceitual do MDM da APA.

(3) ANÁLISE AO ESTADO DOS MASTER DATA

Como foi atrás referido, a APA resultou de uma fusão de vários organismos regionais e nacionais. Os antigos organismos possuíam aplicações de negócio cujas listas de entidades foram consolidadas num processo cujo resultado final não foi o melhor. O processo decorreu da seguinte forma:

1. Na 1ª fase as entidades associadas às utilizações dos recursos hídricos das aplicações SNITURH TRH (Sistema Nacional de Informação de Títulos de Utilização de Recursos Hídricos, Taxas de Recursos Hídricos) que estavam organizadas e validadas foram carregadas correctamente na aplicação SILiAmb.
2. Na 2ª fase foram associadas as entidades provenientes das aplicações de licenciamento das ARHs (Administração de Região Hídrica). Estes dados foram previamente validados e as repetições eliminadas. Até aqui tudo bem.
3. Na 3ª fase começou o caos. Foram carregados no SILiAmb os dados migrados da aplicação SIRAPA (Sistema Integrado de Registos da APA) sem terem sido validados e nem os registos repetidos eliminados.
4. Na 4ª e última fase, as entidades do SILiAmb, com todos os problemas da fase anterior, foram migradas para o FileDoc.

As consequências deste processo foram:

- Troca de entidades na emissão de ofícios externos;
- Existência de documentos com entidades certas, mas com moradas trocadas ou erradas ou até mesmo omissas;
- Necessidade de edição adicional mesmo quando eram utilizados os “*templates*”;
- Falta de confiança entre os utilizadores na lista oficial de entidades proveniente do FileDoc;
- Desperdício de tempo, recursos económicos e humanos para corrigir repetidamente os mesmos dados;
- Projecção de uma imagem externa negativa da APA, I.P.

Para além do processo final caótico de migração descrito nos parágrafos anteriores, nesta análise foram identificadas mais razões que contribuíram para o caos no FileDoc:

- Os dados de *Email*, NIF, Código Postal, Local Postal e Cidade não eram validados durante a criação de novos registos;
- Registos foram inseridos repetidamente nas aplicações;

O gráfico seguinte representa o universo de registos incompletos ou com dados errados antes da implementação do MDM:

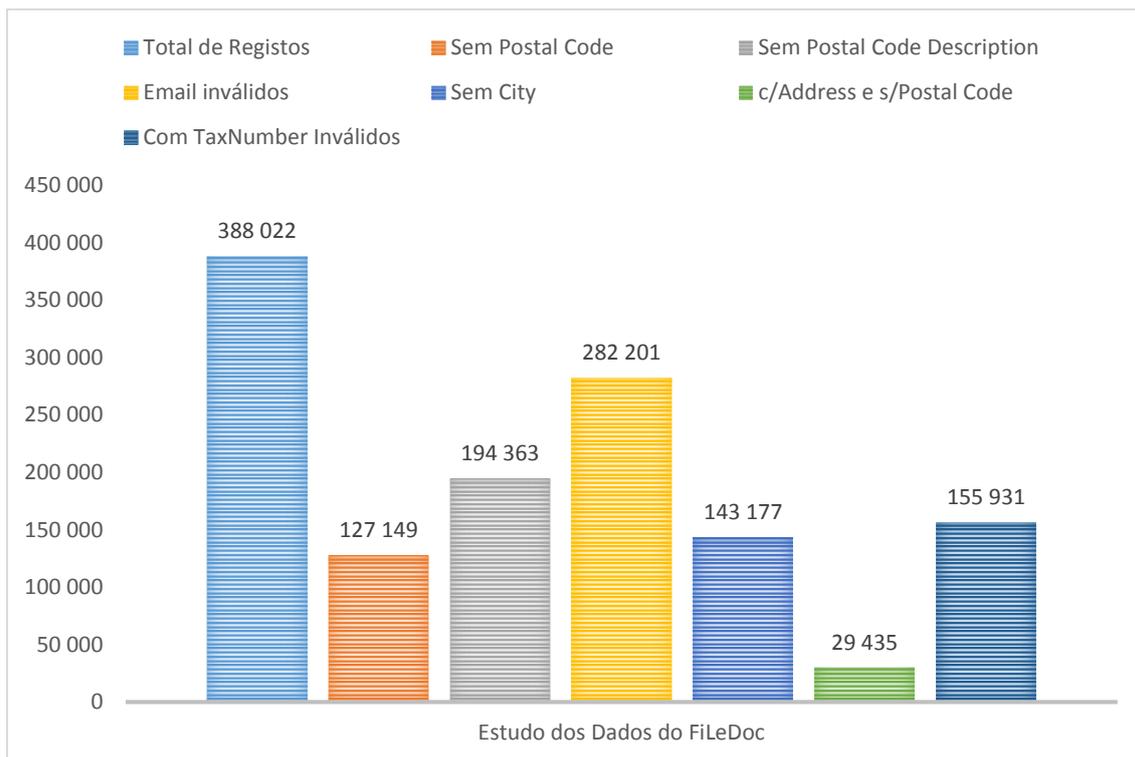


Ilustração 19 - Estudo dos Dados.

Análise ao SILiAmb

Nesta aplicação a criação dos novos registos é feita pelos próprios interessados via acesso *online* em <https://SILiAmb.apambiente.pt>.

O Número de Contribuinte é um campo obrigatório que impede a duplicação de registos. No entanto, as regras de negócio da aplicação ao permitirem a criação de múltiplos perfis (a empresa, o dono da empresa e o representante) para uma mesma entidade, dá lugar, com frequência, à criação de inconsistências nos dados. Um mesmo nome, morada ou contacto aparece escrito de formas diferentes em cada perfil. As principais consequências nos processos que dependem destes dados são:

- O NIF da empresa é frequentemente associado ao funcionário que fez o registo, resultando na emissão de licenças em nome do trabalhador, quando deveria ser em nome da empresa;
- As consultas de entidades no SILiAmb são muito pouco eficientes, devolvendo muitos registos e demorando demasiado tempo.

Como resultado final desta análise podemos associar os seguintes factores negativos aos *Master Data* de entidades do FileDoc e SILiAmb:

- Custos financeiros e económicos causados pelo desperdício de tempo, recursos materiais e humanos;
- Ineficiência operacional das aplicações;
- Projecção de uma imagem negativa da APA para o exterior.

3.3.2. IDENTIFICAÇÃO DOS ADMINISTRADORES DOS MASTER DATA

Para administradores dos *Master Data* foram propostos os dois actuais coordenadores de apoio de primeira linha do FileDoc: João Paulo Pereira (joao.pereira@apambiente.pt) e Pedro Nunes (pedro.nunes@apambiente.pt). Estes dois colaboradores acompanharam o processo de migração das antigas aplicações de gestão documental e por isso têm um profundo conhecimento sobre as entidades. Para além de garantirem apoio técnico às equipas de *helpdesk*, têm agora como missão controlar o estado dos *Master Data* no FileDoc.

3.3.3. GOVERNANÇA DOS MASTER DATA

Como foi anteriormente referido, um componente importante num projecto de MDM é a instituição de um Programa de Governança de Dados (GD).

Um Programa de GD de *Master Data* procura melhorar continuamente a qualidade dos dados numa organização (Berson e Dubov, 2011, p. 9608) e a sua implementação não atempada é um dos primeiros e mais sérios erros que se comete quando se desenvolve um sistema de MDM (Cervo e Allen, 2011, p. 1644).

Efectivamente, enquanto decorria a implementação deste projecto, o FileDoc continuou diariamente a criar novos dados errados e a duplicar registos. Foi portanto necessário antes de mais dar início a um processo de mudança na alteração dos dados para que se começasse a garantir, na fonte, uma qualidade melhor dos dados.

Nesse sentido foi instituído um Conselho de Governança dos *Master Data*. Para este Conselho foram propostos os seguintes elementos: os dois administradores dos *Master Data*, um administrador de base de dados e o chefe de divisão do DTI, como presidente do Conselho.

O CG deve reunir-se de 15 em 15 dias para avaliar o funcionamento do programa. O Conselho tem como objectivos:

- Definir novas políticas e regras de governança;
- Monitorizar a qualidade dos dados;
- Zelar pelo cumprimento das regras e normas de governança de dados.

As referidas regras e normas ficaram definidas num manual interno designado por Manual de Procedimentos (MP) e disponibilizadas via *online* a todos os utilizadores do FileDoc e a todos aqueles que interagem com os *Master Data* de entidades. Entre as regras constantes no manual destacamos as seguintes:

- «Ao serem criadas novas entidades, caso haja permissões no sistema para o fazer, deve-se colocar o nome completo da entidade com a primeira letra de cada palavra em maiúscula e o restante em minúsculas.»
- «O nome da entidade deverá ser escrito por extenso, excepto quando for constituído por uma sigla, neste caso, deve escrever-se a sigla, seguida de hífen e o nome por extenso, com iniciais maiúsculas. Ex: ICNF - Instituto de Conservação da Natureza e Florestas.»
- «Sempre que não seja indicado ou conhecido o NIF da entidade remetente deverá ser usado o seguinte nº: 111111111 (9 algarismos a 1).»

Para apoiar a implementação do MP foram propostas acções de formação específicas aos colaboradores da APA que criam e modificam os dados.

3.3.4. INFRA-ESTRUTURA TECNOLÓGICA

Como infra-estrutura tecnológica a APA optou pelo *software* Sql Server 2012 Enterprise Edition cujo licenciamento já detinha. Este *software* permite implementar os principais serviços de MDM descritos na Arquitectura de Referência e detalhados mais adiante.

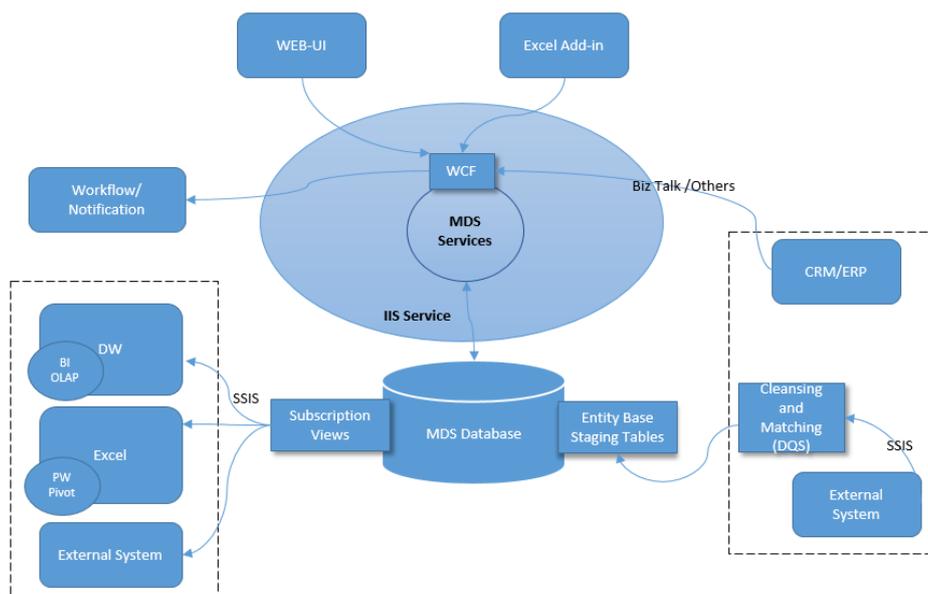


Ilustração 20 - Infra-estrutura Tecnológica.

Fonte: Sql Server 2012.

A componente informática do MDM da Microsoft designa-se globalmente por *Master Data Services* (MDS). Na realidade o MDS conjuga várias aplicações que a Microsoft já possui nomeadamente: o *Excel*, *Sql Server Data Tools*, *Sql Server Integration Services*, etc. Portanto, a solução MDM da Microsoft é construída conjugando estas aplicações.

A seguir são descritos os serviços desta arquitectura e a utilização que foi dada neste trabalho:

- O componente *Data Quality Services* (DQS) presente no lado direito da ilustração anterior permite limpar, fazer o match, padronizar e enriquecer os dados. Através do DQS fizemos as actividades descritas mais adiante;
- O *Master Data Services* (MDS) e o MDS Database compõe o núcleo da solução MDM do Sql Server. Destinam-se essencialmente para gerir e manter a estrutura de armazenamento de dados (modelos, entidades, hierarquias, relações, *Views*, etc.);
- Utilizando a interface web (WEB-UI) do MDS foi criado o modelo de dados para as entidades do FileDoc e com ela os administradores dos *Master Data* podem aceder ao MDS através de um *browser*;
- Com o MDS *Excel Add-in* e usando o Excel qualquer utilizador da APA, com as necessárias credenciais, pode criar e publicar novos *Master Data* no MDS sem

ajuda das TIs. Usando estes componentes foi publicada a lista inicial limpa e consistente de entidades no MDS, conforme descrito mais adiante;

- *Staging tables* é o local onde são guardados os *Master Data* e os atributos do modelo criado no MDS. Para partilhar os *Master Data* do *Staging tables* foi criada uma *Subscription view* ([mdm].[EntidadesFileDoc]). Utilizando queries SQL sobre as *Subscription Views* os consumidores podem consultar os *Master Data* no MDS Database. Com este modelo torna-se mais fácil construir um *Hub MDM* sob garantia das regras do DQS. Com recurso aos serviços de *Sql Server Integration Services*, o processo de enviar e receber *Master Data* do MDS pode ser automatizado implementando soluções de fluxos de dados em *Sql Data Tools*. No âmbito deste trabalho este mecanismo não foi implementado.
- O *External System* integra as aplicações consumidoras e produtoras. No âmbito deste trabalho foi integrado apenas o FileDoc.

3.3.5. PROCESSOS DE LIMPEZA, ELIMINAÇÃO E TESTES

Para as tarefas de Limpeza, Eliminação e Testes foram usados os serviços DQS de acordo com o gráfico que se apresenta:

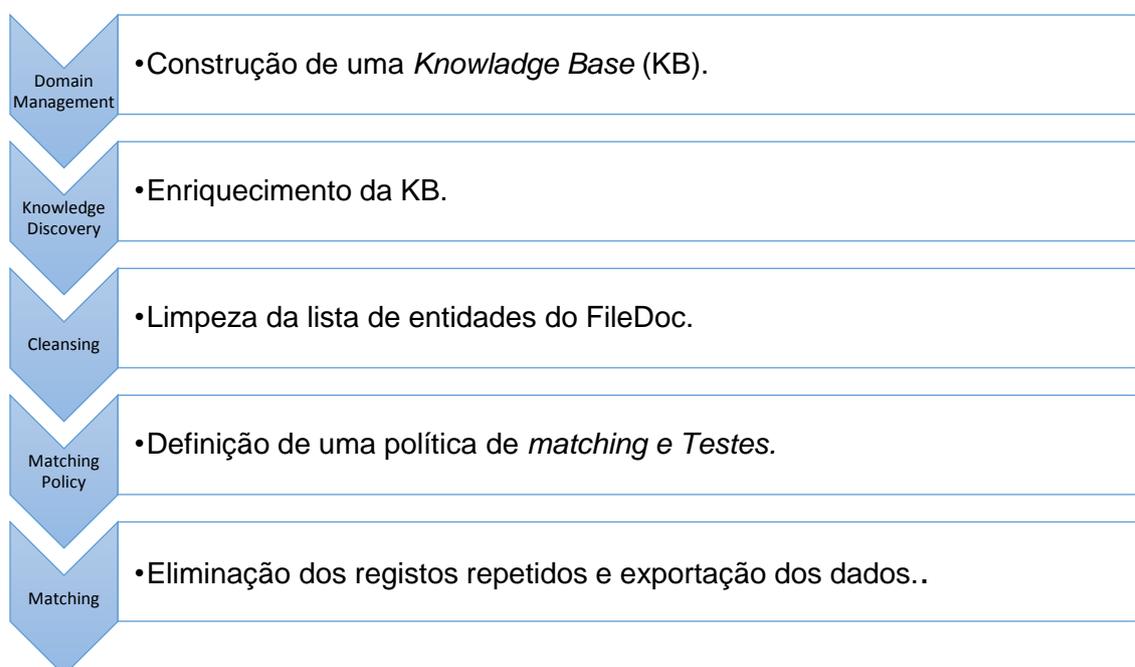


Ilustração 21 - Tarefas de Limpeza e Consistência de dados.

- **Domain Management**

As actividades desta tarefa destinaram-se a construir a base de dados do conhecimento (KB) da APA. Foram criados domínios para os *Master Data* identificados na fase de estudo dos Produtores e Consumidores. Para a validação de domínios foram usadas regras de domínio através de expressões regulares. Por exemplo, a expressão regular aplicada ao domínio *Email* foi a seguinte:

```
^[a-zA-Z0-9_\-\.]+@((\[[0-9]{1,3}\. [0-9]{1,3}\. [0-9]{1,3}\. )|([a-zA-Z0-9\-\ ]+\. ))([a-zA-Z]{2,4}[0-9]{1,3})(\ ?)+$
```

- **DOMÍNIO COMPOSTO**

Um domínio composto permite criar uma regra única que envolva múltiplos domínios. Uma outra característica deste tipo de domínio é permitir fazer *parsing* em múltiplos domínios. Por exemplo, um nome completo pode ser decomposto em vários domínios: primeiro nome, nome do meio e último nome. Para este projecto foi criado o domínio composto “**validar morada**” que, mesmo não tendo sido associada nenhuma regra pelos motivos que explicaremos adiante, foi útil na definição da regra de *Matching* conforme veremos também mais adiante.

REFERENCE DATA

Conforme explicado no estudo, a *Reference Data* permite, através de serviços externos, validar e enriquecer dados como por exemplo, moradas e contactos. Pudemos confirmar a eficácia destes serviços com testes realizados em moradas americanas utilizando o serviço “*Address Check - Verify, Correct, Geocode US and Canadian Addresses*” (<http://datamarket.azure.com/dataset/melissadata/addresscheck>). Infelizmente este serviço destina-se apenas às moradas dos Estados Unidos da América e do Canadá.

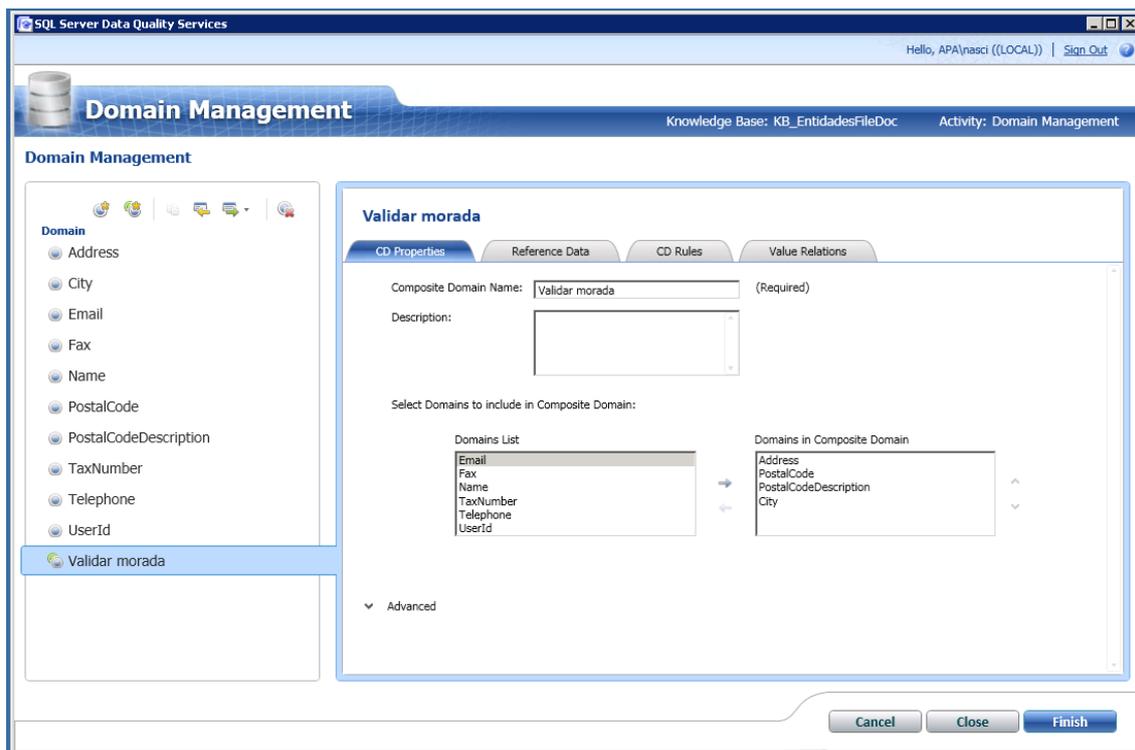


Ilustração 22 - Domínio Composto: Validar morada.

Não tendo sido possível encontrar serviços de *Reference Data* para moradas de Portugal, e uma vez que os CTTs disponibilizam listas dos Códigos Postais e Concelhos, optamos apenas por carregar estes dados nos domínios *PostalCode*, *PostalCodeDescription* e *City*. É um procedimento que deverá repetir-se sempre que os CTTs disponibilizarem novos dados.

KNOWLEDGE DISCOVERY

A Knowledge Base (KB), como já dissemos, é o conjunto dos domínios, valores dos domínios, políticas de *Matching*, regras de domínios e dados externos (*Reference Data*) da solução.

A Descoberta de Conhecimento (*Knowledge Discovery*) é o processo através do qual é carregado o conhecimento que permite a solução identificar e limpar dados incorrectos ou inválidos.

Na etapa anterior (*Domain Management*) importamos individualmente novos valores para os domínios *PostalCode*, *PostalCodeDescription* e *City*. Entretanto, poderíamos tê-lo feito nesta etapa através de uma interface que permitisse importar valores para múltiplos domínios a partir de uma base de dados.

CLEANSING

Depois de criada a KB nas actividades anteriores, passamos para a fase de limpeza dos dados. Esta actividade foi a mais árdua de todas. Foram necessárias várias iterações de limpeza. As iterações foram necessárias porque o volume de dados inválidos identificados pelo sistema e que necessitavam de intervenção humana era muito elevado. Foi necessário recorrer a comandos SQL para corrigir nomes de *Emails* ou eliminar milhares de entidades que na realidade eram nomes de processos documentais erradamente importados nas fases de consolidação dos antigos sistemas.

A limpeza foi feita mapeando os domínios da KB sobre Name, Address, PostalCode, PostalCodeDescription, City, Email, TaxNumber, fax e Telephone.

The screenshot shows the 'Data Quality Project' interface. On the left, there's a sidebar with 'Domains' and 'No. of values' for various fields like Email, Fax, Name, TaxNumber, and Telephone. The main area is titled 'Validar morada' and contains a table with columns: Value, # Records, Address, PostalCode, Correct to, PostalCodeDescription, City, Confidence, and Reason. A row for 'MONTE DA BARCA - APARTADO 104' is highlighted in orange, with its 'PostalCodeDescription' and 'City' fields circled in blue. Below the table, there's a section for 'Records containing the value:' with a similar table structure.

Ilustração 23 - Cleansing do *Emails*.

MATCHING POLICY E TESTES

Nesta fase definiu-se a regra para identificar os registos repetidos. A política de *Matching* obtida no final de vários testes resultou nas regras do quadro abaixo:

Tabela 11 - Regras de Matching Policy.

Domain	Similarity	Weight	Prerequisite
TaxNumber	Exact		Yes
Name	Similar	60%	No
Email	Exact	20%	No
Validar morada	Similar	20%	No

O funcionamento da regra deste quadro é descrito da seguinte forma:

1. “Validar morada” é um domínio composto contendo *Address*, *PostalCode*, *PostalCodeDescription* e *City*. Neste domínio é aplicada uma ponderação única ao conjunto dos domínios que o compõe.
2. Como pré-requisito: só comparar registos com *TaxNumber* iguais;
3. Dois *TaxNumbers* iguais não devem ser, por si só, uma garantia de que são uma mesma entidade pelos seguintes plausíveis dois motivos: a repetição do *TaxNumber* pode ocorrer por erro de introdução; a sequência de ‘11111111’ é o valor que é sempre colocado quando o *TaxNumber* é desconhecido;
4. Dois registos que tenham o mesmo *TaxNumber*, *Names* semelhantes em 60%, e que tenham *Emails* exactamente iguais e/ou “Validar morada” semelhantes em 20% são considerados como sendo uma mesma entidade. Isto é, dois registos são considerados serem a mesma entidade quando a soma do peso dos domínios for igual ou superior a 80% e tiverem o mesmo *TaxNumber*.

A ilustração seguinte mostra o *Cluster* “Ginásio Clube Português” com taxas de semelhanças superiores à 80%:

Rule	Record ID	Cluster	Score	Name	Taxnumber	Address	PostalCode	PostalCodeDescription	City	Email	UserID
1018261	1018259	88%	João Augusto da Conceição Caciones	241479177	R. Olival da Varinha, 7	2900-748	SETUBAL	Setúbal			34544
1013240	1013240	93%	Ginásio Clube Português	500127174	Praça Ginásio Clube Português, 1	1250-111	LISBOA	Lisboa			25917
1013241	1013240	93%	GINÁSIO CLUBE PORTUGUÊS	500127174	Praça Ginásio Clube Português, 1	1250-111	LISBOA	Lisboa			25918
1013242	1013240	93%	GINÁSIO CLUBE PORTUGUÊS	500127174	Praça Ginásio Clube Português, 1	1250-111	LISBOA	Lisboa			25919
1013244	1013240	92%	Ginásio Clube Português	500127174	Praça do Ginásio Clube Português						
1013245	1013240	91%	Ginásio Clube Português	500127174	Praça do Ginásio Clube Português						
1013243	1013240	91%	Ginásio Clube Português	500127174	Praça do Ginásio Clube Português						
1013241	1013241	87%	GINÁSIO CLUBE PORTUGUÊS	500127174	Praça Ginásio Clube Português, 1						
1013246	1013241	87%	Ginásio Clube Português	500127174	Praça do Ginásio Clube Português						
1013243	1013241	89%	Ginásio Clube Português	500127174	Praça do Ginásio Clube Português						
1013245	1013241	89%	Ginásio Clube Português	500127174	Praça do Ginásio Clube Português						
1013244	1013241	99%	Ginásio Clube Português	500127174	Praça do Ginásio Clube Português						
1013244	1013241	99%	Ginásio Clube Português	500127174	Praça do Ginásio Clube Português						

Matching Score Details

Field	Weight	Matched record terms	Pivot record terms	Score contribution
Name	0.6	GINÁSIO CLUBE PORTUGUÊS	Ginásio Clube Português	88%
Email	0.2			100%
Validar morada	0.2	Praça Ginásio Clube Português, 1 1250-111 LISBOA Lisboa	Praça Ginásio Clube Português, 1 1250-111 LISBOA Lisboa	100%

Matching Results Statistics

- Records: 274035
- Matched: 5240 (1.89%)
- Unmatched: 268795 (98.10%)
- Clusters: 2618
- Average cluster size: 2.25
- Min. cluster size: 2
- Max. cluster size: 14

Ilustração 24 - Matching Policy.

Observando os detalhes do registo escolhido na imagem anterior, verificamos que a semelhança entre os “Names” é de 88%. O *Email* e o *Address* têm uma taxa de confiança de 100% cada.

MATCHING

Nesta etapa aplicou-se a Política de *Matching* definida no passo anterior para eliminar os registos duplicados. O resultado da *query* apresentada abaixo mostra a lista de entidades já sem repetições.

```
SELECT Name, TaxNumber, Address, PostalCode, PostalCodeDescription, City, Telephone, Fax, Email
FROM ResultingMatchingResultsFileDoc1
WHERE (Name LIKE 'gin%club%')
ORDER BY Name
```

	Name	TaxNumber	Address	PostalCode	PostalCodeDes...	City
	Ginasio Club De Santo Tirso	501110810	RUA DA MISERI...	4780-501	SANTO TIRSO	Santo Tirso
	Ginásio Clube Figueirense	500127166	Av. 1º de Maio	3080-011	Figueira da Foz	Figueira da Foz
	Ginásio Clube Naval de Faro	501600892	Doca de Faro - ...	8001-901	Faro	NULL
▶	Ginásio Clube Português	500127174	Praça Ginásio Cl...	1250-111	LISBOA	Lisboa
	GINÁSIO CLUBE DE CORREIOS	111111111	Exmº SenhorPre...	2855-150	CORROIOS	Seixal
*	NULL	NULL	NULL	NULL	NULL	NULL

Ilustração 25 - Lista das entidades depois do Matching.

Este processo manual de Limpeza e Eliminação de duplicados, assistido por computador, serviu para produzir a primeira lista limpa e consistente das entidades.

3.3.6. PARTILHA DOS DADOS

Até aqui as tarefas foram: limpar e consolidar os *Master Data*. A tarefa seguinte foi de disponibilizar os dados à toda a organização.

O que diferencia o MDM das outras formas de integração é, como já dissemos, a sua capacidade de partilhar os dados toda a organização.

Neste projecto a partilha dos *Master Data* do FileDoc fez-se utilizando uma *Subscription view*, conforme ilustração abaixo.

As *Subscription Views* têm como principal objectivo permitir que as aplicações acedam aos *Master Data* que estão sob garantia de qualidade do MDS. Para este efeito a lista limpa criada nos passos anteriores foi exportada na fase final do processo anterior (*Matching*).

Script for SelectTopRows command from SSIS *****

```

SELECT
    [VersionName]
    , [Name]
    , [Address]
    , [PostalCode]
    , [PostalCodeDescription]
    , [EnterDateTime]
    , [EnterVersionNumber]
    , [LastChgDateTime]
    , [ValidationStatus]
FROM [MDM].[mdm].[EntidadesFileDoc]

```

VersionName	Name	Address	PostalCode	PostalCodeDescription	EnterDateTime	EnterVersionNumber	LastChgDateTime	ValidationStatus	
1	VERSION_1	Maria Alice Martins Amaro	CARVALHAL	2230-862	CARVALHAL ABT	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
2	VERSION_1	Dulce Nazaré Valério Conceição Vales	Rua José Gomes Ferreira nº 4 - 3ª. Enq.	2035-463	Santo António da Charneca	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
3	VERSION_1	SIAG- Sistema Integrado de Apoio à Gestão para a Administração.	Av. General Norton de Matos, 71. 4ª.	1495-148	ALGES	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
4	VERSION_1	A Rede Natura 2000	Rua da Madalena, 191, 3ª C	1100-319	LISBOA	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
5	VERSION_1	Alice Maria Dos Santos	VAU	2510-665	VAU	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
6	VERSION_1	António Libério da Silva Felício	Monte Idalécio da Silva, Btudes	NULL	NULL	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
7	VERSION_1	António Manuel De Jesus Jorge E Edgard Pontes De Jesus Jorge	Quinta da Mougeta, EN 116 - Ságados	2640-425	MAFFRA	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
8	VERSION_1	Carlos Do Carmo Pinto Roque	Monte dos Carregais	7050-012	Montemor-o-Novo	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
9	VERSION_1	Helena Flora Ramos Gonçalves Pereira Vigeant Gomes	Rua Afonso Albuquerque, 6	2765-184	ESTORIL	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
10	VERSION_1	Henrique M F Sobreiro. Unipessoal, Lda	Rua 11 de Agosto, Nº 3 - Landal	2500-541	LANDAL	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
11	VERSION_1	IGFSS - Instituto de Gestão Financeira da Segurança Social	Alameda D. Afonso Henriques, 82. 1ª e 2ª	1049-076	LISBOA	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
12	VERSION_1	Joaquim José Dias Antunes	Estrada do Codeçal	2640-602	SOBRALDA ABELHEIRA	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
13	VERSION_1	Lidia Maria Gonçalves Casimiro Robalo	Calçada do Espírito Santo nº 3	6300-115	GONCALD	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded
14	VERSION_1	Manuel Antunes da Silva	Almeque	6100-318	CERINACHE DO BONARDIM	2015-02-08 21:30:18.633	1	2015-02-08 21:30:20.443	Validation Succeeded

Ilustração 26 - A Subscription View das entidades

Microsoft SQL Server 2012 Master Data Services Explorer > Entidades > EntidadesAPA | NovaTabelaDeEntidadesFileDoc-v2 | VERSION_1(=1) | (Aberta)

Entidade NovaTabelaDeEntidadesFileDoc-v2

Adicionar Membro | Eliminar Membro | Aplicar Regras | Ver Transações | Filtro | Definições

	Nome	Código	TaxNumber	DefaultAddressID	Address	Posta
✓	Maria Alice Martins Amaro	990	111148073	132203	CARVALHAL	2230-86
✓	Dulce Nazaré Valério Conceição Vales	991	111111111	6	Rua José Gomes Fer	2835-4
✓	SIAG- Sistema Integrado de Apoio à Gestão para a Administração Pública	993	111111111	8	Av. General Norton	1495-14
✓	A Rede Natura 2000	994	111111111	9	Rua da Madalena, 1	1100-31
✓	Alice Maria Dos Santos	995	158294874	248288	VAU	2510-66
✓	António Libério da Silva Felício	997	111111111	12	Monte Idalécio da Si	
✓	António Manuel De Jesus Jorge E Edgard Pontes De Jesus Jorge	998	106423444	100643	Quinta da Mougeta	2640-42
✓	Carlos Do Carmo Pinto Roque	1003	188935436	272371	Monte dos Carregai	7050-01
✓	Helena Flora Ramos Gonçalves Pereira Vigeant Gomes	1017	108346684	28	Rua Afonso Albuque	2765-18
✓	Henrique M F Sobreiro. Unipessoal, Lda	1018	506622169	29	Rua 11 de Agosto, A	2500-54
✓	IGFSS - Instituto de Gestão Financeira da Segurança Social	1020	111111111	31	Alameda D. Afonso	1049-07
✓	Joaquim José Dias Antunes	1021	111111111	32	Estrada do Codeçal	2640-60
✓	Lidia Maria Gonçalves Casimiro Robalo	1025	106872079	36	Calçada do Espírito	6300-11
✓	Manuel Antunes da Silva	1026	111111111	37	Almeque	6100-31
✓	Maria Emília Leal	1028	114211922	39	Sítio dos Semos, E	6250-18
✓	Maria Helena Mendes Faria	1029	170697720	40	Largo Francisco da	2350-51
✓	Maria Manuela Contreiras Seguro Serra Correia Nunes	1030	127417575	105567	Rua Gomes Leal, Nº	1000-16
✓	MOTA-ENGL, Engenharia e Construção, SA	1031	500197814	369175	Rua Mário Dionísio,	2799-55
✓	Pavizés . Lda	1032	503241660	180682	ODIVELAS	2675-67
✓	Prescor - Produção De Escórias Moidas, Lda.	1033	504205420	243879	Parque Industrial da	2840-07
✓	Recicampo, Lda.	1034	508478812	207445	Rua Dr. Botelho de	3240-13
✓	RUPAUTO Automóveis de Aluguer	1035	500236925	46	Rua da Beneficênci	1600-01

1-50 de 271660 membros | Page 1 of 5434

Detalhes

Nome: IGFSS - Instituto de Gestã
Código: 1020
TaxNumber: 111111111
DefaultAddressID: 31
Address: Alameda D. Afonso Henriques, 82 1º e 2º
PostalCode: 1049-076
PostalCodeDescription: LISBOA
City: Lisboa
Telephone: 218445650
Fax: 218445660
Email:
Anotações: Ver Anotações
Erros de Validação (0)
OK Cancelar

Ilustração 27 - Lista de entidades no MDS Datastore.

Usando as funcionalidades “Combine”, “Match” e “Publish” do MDS Addin-Excel os administradores de dados “Data Stewards” deste projecto podem facilmente gerir os Master Data. Sempre que necessário os novos dados são carregados e combinados

(“Combine”) com os existentes no MDS Datastore. A eliminação de registos duplicados é feita através das funções “Match” e “Publish”.

	D	E	F	H	I	J	K
1	Ligação MDS: mdm(http://mdm/mdm/) Modelo: EntidadesAPA Versão: VERSION 1 Entidade: NovaTabela de EntidadesFileDoc-v2 Obtido: 08-02-2015 21:29:56						
2	Name	Code	TaxNumber	PostalCode	PostalCodeDescription	City	Telephone
3	Associação de Escoteiros de Portugal	10000	111111111	2745-086	QUELUZ	Queruz	
4	Sérgio Santos Unipessoal Lda.	100000	508875617	3700-742	MILHEIRÓS DE POIARE	Santa Maria da Feira	
5	Amarelo E Incolor, Lda	100001	510145248	2440-334	BATALHA	Batalha	
6	Jpmt-Auto, Lda.	100002	510143369	1685-791	FAMÕES	Odivelas	
7	Hortimilk - Sociedade Agrícola, Lda.	100003	510142818	4710-571	ADAÚFE	Braga	
8	Azimuth-Gabinete De Projectos	100004	510142516	4610-594	Borba de Godim	Borba de Godim	
9	Canteiro Da Carvalhosa, Lda.	100005	510141943	3750-818	VALONGO DO VOUGA	Águeda	234646259/917092383
10	Rui L. Gomes Unipessoal Lda.	100006	510141358	3150-287	SEBAL	Condeixa-a-Nova	
11	António Quaresma & Filhos, Lda.	100007	510140360	3300-113	ARGANIL	Arganil	
12	Japblue - Comércio Automóveis, Lda.	100008	510140106	4560-162	GUILHufe	Penafiel	
13	Joaquim Bonifácio - Resíduos E Sucata	100009	510139701	2985-213	PEGÕES	Montijo	
14	Carlos Carvalho Vieira, Unipessoal	100010	510139515	2420-479	ARRABAL	Leiria	
15	Supermonchique - Supermercados,	100011	510138802	8550-328	MONCHIQUE	Monchique	
16	Aura Light	100012	510138535	2705-869	TERRUGEM SNT	Sintra	210999344/915825934
17	Ecosize - Reciclagem Lda	100013	510137008	4705-136	BRAGA	Braga	
18	Prioridynamic	100014	510136605	4475-310	MAIA	Maia	
19	Agofransil-Unipessoal, Lda.	100015	510135811	4455-111	LAVRA	Matosinhos	
20	Autostoi Unip. Lda	100016	510135730	8005-455	FARO	Faro	
21	Auto Paulo Neto Reparação De Autoc	100017	510135668	4615-139	FIGUEIRÓ (SANTA CRIS)	Amarante	
22	R2P - Reciclagem E Peças, S.A.	100018	510135048	3360-287	SAZES DO LORVÃO	Penacova	
23	Pinto & Cruz - Serviços De Manuten	100019	510134980	4100-247	PORTO	Porto	
24	Associação De Estudos E Defesa Do	10002	501779990	2000-096	SANTARÉM		
25	Decifacil Lda	100020	510132553	2680-111	CAMARATE	Loures	
26	Respostalider - Unipessoal, Lda	100021	510131794	3750-069	AGUADA DE CIMA	AGUADA DE CIMA	
27	Verdes Quimeras, Lda	100022	510130771	4400-676	VILA NOVA DE GAJA	Vila Nova de Gaia	
28	Sal E Sabor, Lda	100023	510129870	4705-086	BRAGA	BRAGA	
29	Guita - Transformação E Exportação	100024	510129161	4970-737	ARCOS DE VALDEVEZ	Arcos de Valdevez	
30	Quinta De Gandarém, Lda.	100025	510129048	5160-035	Foz do Sabor	Foz do Sabor	

Ilustração 28 - Interface Master Data com MDS Addin-Excel.

4. CONCLUSÃO

4.1. CONCLUSÕES

O presente trabalho, com a aplicação dos conceitos de MDM, permitiu à APA implementar um sistema centralizado de *Master Data* consistentes (*consistency*) e precisos (*accuracy*).

A APA passou a ter uma arquitectura global de **Integração de Aplicações e Sistemas** ao nível de dados. Sistemas como o DW, BI, ERP e iniciativas de Interoperabilidade aplicacional passam a dispor de uma plataforma de **Party Data** comuns com uma economia de escala maior.

De salientar que com a implementação do MDM foi possível identificar e eliminar mais de 100.000 falsas entidades resultantes essencialmente de processos de integração mal sucedidos e de introdução manual.

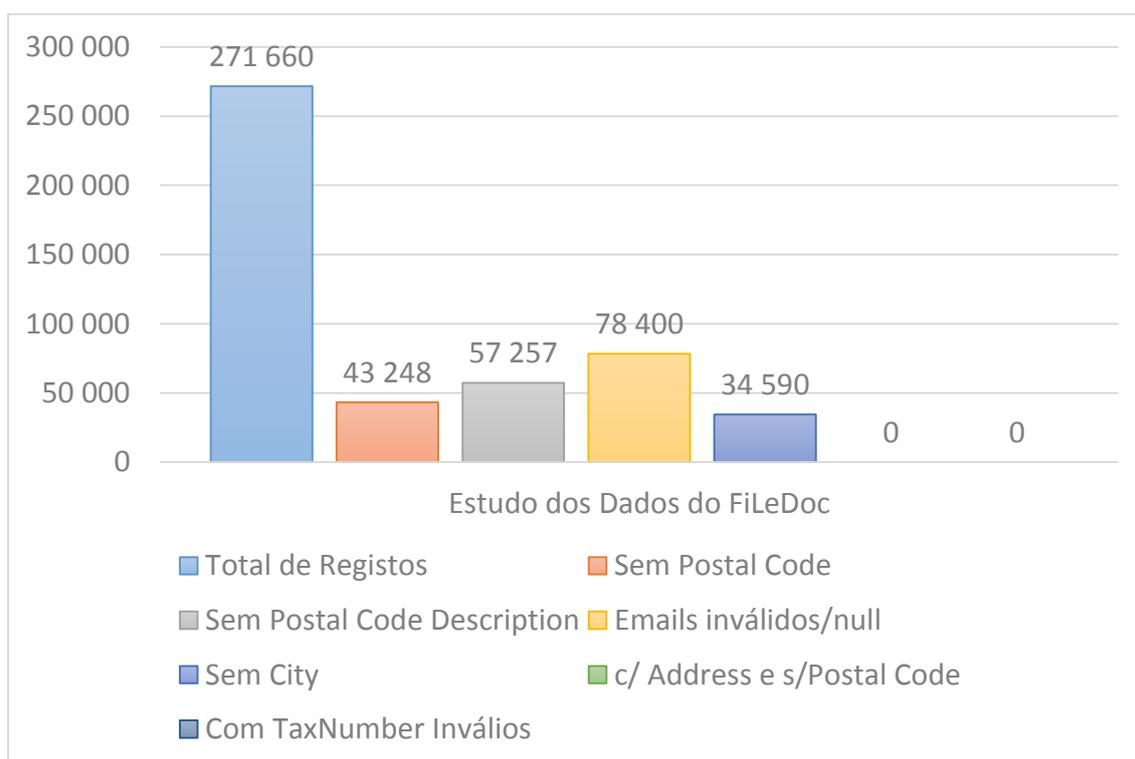


Ilustração 29 - Estado dos dados depois do Cleansing e Matching.

As modificações introduzidas na aplicação FileDoc com o contributo deste trabalho, e que poderão replicar-se na integração de *Master Data* de outras aplicações, criaram novas funcionalidades de validação de dados. Por exemplo o campo código postal

passou a ser introduzido por “*view lookup*”, o formato dos novos emails passou a ser validado na aplicação, etc.

4.2. LIMITAÇÕES

Corrigir moradas é uma tarefa árdua e complexa quase impossível de ser realizada manualmente em tempo útil quando a volumetria de dados é muito grande. A inexistência actualmente em Portugal de serviços de *Reference Data* de moradas foi o factor mais limitativo deste trabalho. Com recurso a um serviço de *Reference Data* teriam sido corrigidas mais moradas e portanto teria sido possível detectar mais registos duplicados em menos tempo.

Em Portugal os CTTs e a Autoridade Tributária (AT) são dois organismos que poderiam validar os dados de moradas. Em resposta aos pedidos destes serviços, os CTT informaram que provavelmente em 2016 disponibilizariam *webservices* de *Reference Data*. A AT por sua vez declinou o pedido de disponibilizar *webservices* de validação de dados invocando restrições legais.

Não obstante os autores mais citados ao longo deste trabalho serem os mais referenciados em todos os trabalhos consultados, o leque reduzido de opções bibliográficas disponíveis em Portugal, quer em português quer em outra língua, foi um factor delimitativo da lista de Referência Bibliográficas deste trabalho.

4.3. RECOMENDAÇÕES

São as seguintes as recomendações sugeridas por este trabalho:

1. É importante tirar proveito da tecnologia que já possa estar disponível na organização. Na APA a opção pelo Sql Sever 2012 Enterprise Edition evitou custos de aquisição de novos licenciamentos.
2. Se o objectivo é integrar variados domínio de dados (pessoas, produtos, contractos, etc.) recomenda-se a utilização de um *software* que suporte a implementação de sistemas MDM Multi-Form
3. No estilo de implementação de MDM tipo Coexistência os *Master Data* podem ser alterados no MDM e nas aplicações fonte. Para aumentar a eficiência da solução podem ser desenvolvidos algoritmos de automatização dos processos de

actualização e consolidação dos *Master Data*. Por exemplo, do lado das aplicações o procedimento poderá ser:

- 1) Quando é necessário criar uma nova entidade, a mesma deve ser pesquisada a ver se já existe na base de dados da aplicação;
- 2) Caso já exista, deverá ser usado esse registo;
- 3) Caso não exista na aplicação, a pesquisa deverá estender-se ao MDM. Caso exista no MDM, o registo deverá ser copiado para a aplicação.
- 4) Caso não exista nem na aplicação nem no MDM, então a aplicação deverá criar um novo registo.

Do lado do MDM o procedimento pode ser o seguinte:

- 1) Obter periodicamente, em modo batch, os novos registos entretanto criados nas aplicações fonte;
- 2) Limpar, consolidar;
- 3) Devolver os dados a cada aplicação invocando os respectivos webservices para serem actualizados os registos nas respectivas bases de dados.

4.4. INVESTIGAÇÃO FUTURA

Para investigação futura, aos estudos que envolvam a construção de um sistema de MDM como continuação deste estudo, propõe-se o estudo de *Total Cost of Ownership* (TCO) e *Return On Investment* (ROI).

Conforme dissemos anteriormente, para a implementação da solução de MDM na APA optamos pela tecnologia que já dispúnhamos e, por isso, o presente estudo não fez referências a outras tecnologias. Propõe-se, portanto, para a investigação futura o estudo de TCO e ROI comparativo de diferentes tecnologias.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ALEX, Berson; LARRY, Dubov (2011)

- Master Data Management and Data Governance, Second Edition. New York: Mc Graw Hill. 9780071744591.

ALVES, Teresa

- CTT: Pesquisa de Códigos Postais [Mensagem em linha] para Nascimento Joaquim Augusto. 12 Nov. 2014. [Consult. 12 Nov. 2014]. Comunicação pessoal.

CERVO, Dalton; ALLEN, Mark (2011)

- Master Data Management in Practice Achieving True Customer MDM. New Jersey: Wiley Corporate F & A. 9781118085684.

BATINI, Carlo; SCANNAPIECO, Mónica (2006)

- Data Quality Concepts, Methodologies and Techniques. Berlin: Springer. 3540331727.

BUTLER, David (2011)

- Master Data Management: an oracle overview paper [Em Linha]. ORACLE. [Consult. 12 Maio. 2014]. Disponível em WWW: <URL:<http://www.oracle.com/us/products/applications/master-data-management/018876.pdf>>.

CERVO, Dalton; ALLEN, Mark (2011)

- Master Data Management in Practice Achieving True Customer MDM. New Jersey: Wiley Corporate F & A. 9781118085684.

DREIBELBIS, Allen; HECHLER, Eberhard; MILMAN, Ivan; OBERHOFER, Martin; RUN Paul; WOLFSON, Dan (2008)

- Enterprise Master Data Management. Boston: IBM Press. 9780132366250.

DREIBELBIS, Allen; OBERHOFER, Martin (2008)

- An introduction to the Master Data Management Reference Architecture. [Em Linha]. IBM. [Consult. 12 Jun. 2014]. Disponível em WWW: <URL:<http://www.ibm.com/developerworks/data/library/techarticle/dm-0804oberhofer/>>.

ENDESA, S.A. (20 Nov. 2013)

- Informação rectificativa [Mensagem em Papel] para Nascimento Augusto. [Consult. 20 de Nov. 2013] Comunicação pessoal.

ENDESA, S.A. (24 de Out. 2013)

- Aviso de Rescisão do Contracto de Fornecimento de Energia [Mensagem em Papel] para Nascimento Augusto. [Consult. 24 de Out. 2013] Comunicação pessoal.

ERL, Thomas (2006)

- Service-Oriented Architecture Concepts, Technology, and Design. New Jersey: Prentice Hall. 0131858580.

FERNANDES, António (2005)

- A Qualidade dos Dados No Apoio À Tomada de Decisão Em Ambientes Complexos. Lisboa: Instituto Superior de Economia e Gestão.

FERREIRA, João; MIRANDA, Miguel; ABELHA, António; MACHADO, José (2010)

- O Processo ETL em Sistemas Data Warehouse. [Em Linha]. Lisboa. [Consult. 15 Ago. 2014]. Disponível em WWW:

<URL:<http://inforum.org.pt/INForum2010/papers/sistemas-nteligentes/Paper080.pdf>>.

FRED, Cummins. (2002)

- Enterprise Integration. New York: John Wiley & Sons. 0471400106.

GALHARDAS, Helena; TORRES, Luís; DAMÁSIO, João (2010)

- Master Data Management: a proof of concept. [Em Linha]. Lisboa. [Consult. 18 Ago. 2014]. Disponível em WWW:

<URL:http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202010/Papers/2B3_MDMPProofOfConcept.pdf>.

GARTNER (2013)

- Magic Quadrant for Data Quality Tools [Em Linha]. [Consult. 20 Jun. 2014]. Disponível em WWW: <URL: <http://www.informatica.com/pt/data-quality-magic-quadrant/>>.

IGFEJ (2013)

- Esclarecimento do Instituto de Gestão Financeira e Equipamentos da Justiça, I.P. [Em Linha]. [Consult. 15 Set. 2014]. Disponível em WWW: <URL: <http://www.google.pt/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CB8QFjAA&url=http%3A%2F%2Fwww.citius.mj.pt%2FportalDnn%2FLinkClick.aspx%3Ffileticket%3DaHfi2oFh9uY%253D%26tabid%3D59&ei=kvPqVMHvOYf1UryagbAJ&usg=AFQjCN66RU3FTASlVImuO4kDEjYy0cnq6w&sig2=6lsRFBYPLCDDFFhal3yN6Gw&bvm=bv.86475890,d.d24>>

IMHOFF, Cláudia; COLIN, White (2006)

- Master Data Management. [Em Linha]. [Consult. 3 Jul. 2014]. Disponível em WWW: <URL:ftp://ftp.software.ibm.com/software/emea/de/db2/WP_B-Eye-creating-a-single-view-of-the-business.pdf>.

KAREL, Rob. (2006)

- Introducing Master Data Management: achieving a single version of the truth. [Em Linha]. [Consult. 18 Jun. 2014]. Disponível em WWW: <URL:<http://i.zdnet.com/whitepapers/Forr051104010700.pdf>>.

LEVY, Dyché (2006)

- Customer Data Integration: Reaching a Single Version of the Truth. New Jersey: John Wiley & Sons . 9780471916970

LINTHICUM, David (2000)

- Enterprise Application Integration. Boston: Addison-Wesley.0201615835.

LOSHIN, David (2009)

- Master Data Management Knowledge Integrity, Inc. USA: Morgan Kaufmann. 978-0123742254.

OSLON, Jack, (2003)

- Data Quality: the accuracy dimension. São Francisco: Morgan Kaufman.1558608915.

OTTO, Boris; EBNER, Verena (2010)

- Measuring Master Data Quality. [Em Linha]. Institute of Information Management University of St. Gallen. [Consult. 20 Jun. 2014]. Disponível em WWW:

<URL:
http://www.google.pt/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&ved=0CDUQFjAC&url=http%3A%2F%2Fwebdoc.sub.gwdg.de%2Funiverlag%2F2010%2Fmkwi%2F03_anwendungen%2Fbusiness_intelligence%2F04_measuring_master_data_quality.pdf&ei=OMvEVPSFOIW2Uen4gPAF&usg=AFQjCNECgbx5CFxOrfOzDfl02-h0cVbxiA&sig2=wmcCGhPrMuAmfiE6jvKM6g>.

QUADRADO, Felisbina

- Agência Portuguesa do Ambiente: Projecto de Gestão de Entidades da APA, L.P. (Master Data Management) [Mensagem em linha] para Nascimento Joaquim Augusto. 26 Mar. 2015. [Consult. 26 Mar. 2014]. Comunicação pessoal.

Rádio Notícias TSF (2014) [Em Linha]. Lisboa. [Consult. 20 Set. 2014]. Disponível em WWW:

<URL:http://www.tsf.pt/PaginalInicial/Portugal/Interior.aspx?content_id=4121797>.

Rádio Notícias TSF (2014) [Em Linha]. Lisboa. [Consult. 20 Set. 2014]. Disponível em WWW:

<URL:http://www.tsf.pt/PaginalInicial/Portugal/Interior.aspx?content_id=4131775&page=-1>.

SaS Institute Software, Lda. (2014)

- Criar Uma Visão Única do Cliente. [Em Linha]. Lisboa: SaS Portugal. [Consult. 18 Maio. 2014]. Disponível em WWW:

[URL:http://www.sas.com/offices/europe/portugal/campanhas/ks_visao_cliente.pdf](http://www.sas.com/offices/europe/portugal/campanhas/ks_visao_cliente.pdf).

SH (2014)

- Autoridade Tributária: Validação de moradas e NIF [Mensagem em linha] para Nascimento Joaquim Augusto. 5 Nov. 2014. [Consult. 5 Nov. 2014]. Comunicação pessoal.

SILVA, Miguel (2003)

- Integração de Sistemas de Informação. Lisboa: FCA.9727223915.

TOMI, Dahlberg; JUKKA, Heikkilä; MARIKKA, Heikkilä (2011)

- Framework and Research Agenda for Master Data Management in Distributed Environments. [Em Linha]. Finlândia. [Consult. 6 Jan. 2014]. Disponível em WWW:

<URL:"https://jyx.jyu.fi/dspace/bitstream/handle/123456789/37203/2011mdm21.06.2011delivered.pdf?sequence=1">.

WAYNE, Eckerson (2002)

- Data Warehousing Special Report Data Quality and the bottom line. [Em Linha]. [Consult. 4 Ago. 2014]. Disponível em WWW:

<URL:

http://www.estgv.ipv.pt/PaginasPessoais/jloureiro/ESI_AID2007_2008/fichas/TP06_anexo1.pdf>.

WHITE, Colin. (2007)

- Using Master Data in Business Intelligence. [Em Linha]. [Consult. 12 Jan. 2014]. Disponível em WWW:

<URL: http://fm.sap.com/pdf/mar09/Using Master Data in Business Intelligence.pdf>.

WILLIAM, Ruh; FRANCIS, Maginnis; WILLIAM, Brown (2001)

- Enterprise Application Integration. New York: John Wiley & Sons. 0471376418.

WOLTER, Roger, HASELDEN, Kirk (2006)

- The What, Why, and How of Master Data Management. [Em Linha]. [Consult. 18 Dez. 2014]. Disponível em WWW: <URL

http://67.207.137.15/paper/filename/34/Microsoft_-_The_What_Why_and_How_of_Master_Data_Management.pdf >.