

Integração de Informação entre Sistemas Geracionais Distintos

Uma Experiência e Sistematização na ITV

2008 / 2009

1010003 – João Pedro Barbosa da Silva



isep

instituto
superior de
engenharia do
porto



Instituto Superior de Engenharia do Porto

Integração de Informação entre Sistemas Geracionais Distintos

Uma Experiência e Sistematização na ITV

João Pedro Barbosa da Silva

Dissertação para a obtenção de Grau de Mestre em

Engenharia Informática

Área de especialização em

Tecnologias do Conhecimento e Decisão



Orientador: **Dr. Nuno Alexandre Pinto Silva**

Porto, Outubro de 2009

«Aos meus pais Carlos e Amélia
à minha esposa Zé»

Agradecimentos

Ciente de que a ausência nunca será compensada, não posso deixar de agradecer à minha família toda a compreensão, suporte e apoio que sempre demonstraram ao longo da minha recente formação, assim como aos meus amigos mais próximos todo o apoio que me proporcionaram para que pudesse realizar este trabalho.

Quero deixar também uma palavra de apreço às organizações que se disponibilizaram para uma experiência documentada de migração e integração num ambiente real.

Finalmente um agradecimento ao meu orientador, Dr. Nuno Alexandre Pinto Silva, pela forma entusiasta com que me incentivou e orientou, pela total disponibilidade e pelo muito que me deu a aprender.

Resumo

A natural evolução dos sistemas de informação nas organizações envolve por um lado a instalação de equipamentos actualizados, e por outro a adopção de novas aplicações de suporte ao negócio, acompanhando o desenvolvimento dos mercados, as mudanças no modelo de negócio e a maturação da organização num novo contexto.

Muitas vezes esta evolução implica a preservação dos dados existentes e de funcionalidades não cobertas pelas novas aplicações. Este facto leva ao desenvolvimento e execução de processos de migração de dados, de aplicações, e de integração de sistemas legados.

Estes processos estão condicionados ao meio tecnológico disponível e ao conhecimento existente sobre os sistemas legados, sendo sensíveis ao contexto em que se desenrolam.

Esta dissertação apresenta um estado da arte das abordagens à migração e integração, descreve as diversas alternativas, e ilustra de uma forma sistematizada e comparativa os exercícios realizados usando diferentes abordagens, num ambiente real de migração e integração em mudança.

Palavras-chave

Integração de informação, Migração de dados, Migração Continuada, Sistemas Legados, Ontologia

Abstract

The natural evolution of information systems on organizations involves the installation of upgraded equipment and the adoption of new applications to support the business, following the market development, the changes of the business model and the maturation of the organization in a new context.

Often this implies the preservation of existing data and functionalities that are not included in new applications. This leads to the development and implementation of new procedures regarding the data and applications migration, and the integration of legacy systems.

These integration processes are constrained by the available technological environment and also by the existing knowledge about legacy systems, being sensitive to the context in which they take place.

This paper presents a state of the art on approaches to migration and integration, describes various alternatives, and illustrates, in a systematic and comparative way, the performed exercises using different approaches, in a changing real world migration and integration scenario.

Keywords

Information Integration, Data Migration, Continuing Migration, Legacy Systems, Ontology

Índice

Agradecimentos.....	6
Resumo.....	7
Abstract	8
Índice.....	9
Tabelas	12
Figuras.....	14
Abreviaturas	16
1 Introdução	17
1.1 Caso de Estudo.....	18
1.1.1 Sistema de Informação Legado	18
1.1.2 Sistema ERP/SCM.....	18
1.2 Organização do documento.....	19
2 Contexto tecnológico	20
2.1 Objectivos.....	20
2.2 Abordagens técnico-organizacionais.....	20
2.3 Integração de informação.....	23
2.3.1 Descrição dos dados	25
2.3.2 Tipos de sistemas.....	26
2.3.3 Integração de esquema (migração de dados).....	26
2.3.4 Integração de catálogos	27
2.3.5 Integração de dados	29

2.3.6	Outras abordagens	30
2.3.7	Comparação de abordagens	32
2.4	Estado da Arte	34
2.4.1	TOSEM.....	34
2.4.2	OMIS	35
2.4.3	IKF-IES	37
2.4.4	TAO.....	38
2.4.5	Attachmate.....	39
2.5	Sumário	39
3	Migração ad hoc.....	41
3.1	Introdução.....	41
3.2	Terceiros.....	43
3.2.1	Análise semântica comparativa	43
3.2.2	Mapeamento	46
3.2.3	Abordagem tecnológica.....	49
3.3	Conclusão.....	56
4	Migração declarativa.....	58
4.1	Migração de dados baseado em Ontologias	58
4.2	Semantic Bridging Ontology.....	60
4.2.1	Ponte de Conceito (<i>Concept Bridge</i>).....	61
4.2.2	Ponte de Propriedades (<i>Property Bridge</i>).....	62
4.2.3	Relação entre PonteConceito e PontePropriedades	62
4.3	Pós-processamento.....	63
4.4	Solução desenvolvida.....	64
4.4.1	De esquema para ontologia.....	65
4.4.2	Mapeamento	66
4.4.3	Pós-processamento	73
4.5	Conclusão.....	74
5	Migração continuada.....	76
5.1	Revisão.....	76
5.2	Novo contexto	77
5.2.1	Migração de outras áreas	77
5.2.2	Sobreposição de funções entre sistemas	79
5.2.3	Mudança de layout organizacional	80
5.3	Vista global	81
5.4	Ferramenta de mapeamento e migração.....	82
5.5	Conclusões	86

6	Conclusão.....	87
6.1	Migração ad hoc.....	88
6.2	Migração declarativa.....	89
6.3	Migração continuada.....	90
6.4	Notas finais.....	90
7	Trabalho Futuro.....	92
7.1	Domínio de conhecimento	92
7.2	Conhecimento dos dados.....	93
7.3	Limpeza dos dados e correcções	93
7.4	Conhecimento dos processos	93
7.5	Outras áreas de integração no caso estudado	95
	Referências bibliográficas	96
	Anexo 1	98
	Plano de Contas.....	98
	Anexo 2	102
	Casos atípicos nos documentos com reflexos na identificação de terceiros.....	102
	Anexo 3	105
	Novo layout organizacional.....	105
	Anexo 4	108
	Experiência de Migração Continuada – utilização do tradutor do sistema legado.....	108
	Testes e Experiências.....	111
	Refinamentos	116

Tabelas

Tabela 1 - Modelos de crescimento de sistemas de informação.....	22
Tabela 2 - Processos de migração e integração	23
Tabela 3 - Comparação entre as abordagens à integração.....	33
Tabela 4 - Caracterização de abordagens de integração de informação.....	34
Tabela 5 - Conceptualização nas contas POC de terceiros nos dois sistemas	45
Tabela 6 - Comparação de modelação de Terceiros entre os sistemas Legado e SCM.....	45
Tabela 7 - Código Agrupador para um cliente no Sistema Legado.....	46
Tabela 8 - Mapeamento de terceiros do Sistema Legado para o SCM.....	49
Tabela 9 - Registos de um terceiro migrados para o SCM.....	51
Tabela 10 - Registos do mesmo Terceiro no Sistema Legado com grafia diferente	52
Tabela 11 - Processo manual de mapeamento de Terceiros.....	55
Tabela 12 - Extracto da Ontologia do sistema SCM em RDFS	65
Tabela 13 - Instâncias da Ontologia de Terceiros origem.....	68
Tabela 14 - Excerto da transformação de dados da ontologia origem para instâncias de destino.....	68
Tabela 15 - Terceiros do Sistema Legado pós-processamento.....	73
Tabela 16 - Tabela de contas	73
Tabela 17 - Tabela de Terceiros actualizada	74
Tabela 18 - Especificação RDF do processo pós-processamento	74
Tabela 19 - Utilização das condições em Synon	94
Tabela 20 - Classes de contas do POC	98
Tabela 21 - Contas de Terceiros por mercado e produto.....	101

Tabela 22 - Contas de Clientes por mercados e produtos	101
Tabela 23 - Contas de Fornecedores por mercado e natureza	101
Tabela 24 - Código de cores no mapeamento de Terceiros.....	117
Tabela 25 - Associação de Terceiros e Código Agrupador.....	118
Tabela 26 - Verificação da duplicação de NIF após mapeamento assistido.....	119
Tabela 27 - Entidades diferentes com o mesmo NIF, após primeira verificação.....	120
Tabela 28 - Número de registos de terceiro mapeados o sistema SCM	121

Figuras

Figura 1 - Um cenário genérico de integração de informação	24
Figura 2 - Integração – Abordagens à identificação de correspondências sobre vistas.....	25
Figura 3 - Integração de esquema.....	27
Figura 4 - Cenário de Integração de Catálogo com correspondência.....	28
Figura 5 - Cenário com Integração de Dados.....	29
Figura 6 - Mediação - Vista integrada só de leitura	30
Figura 7 - Mediação com leitura e escrita	30
Figura 8 - Federação de Base de dados	31
Figura 9 - Coordenação de múltiplas bases de dados utilizando <i>workflow</i>	32
Figura 10 - <i>Wrappers</i> na interface entre sistemas legados e novos.....	35
Figura 11 - Resumo de Memória Organizacional	36
Figura 12 - Esquemas e instâncias equivalentes de Terceiros nos sistemas Legado e SCM.....	47
Figura 13 - Associação ou Mapeamento de Terceiros do Sistema Legado para o SCM.....	48
Figura 14 - Cronograma para migração/integração.....	49
Figura 15 - Representação informal de Migração Baseada em Ontologias	58
Figura 16 - MAFRA - MApping FRAMework	60
Figura 17 - Representação UML de relação conceptual entre Ponte Semântica e Serviço.....	61
Figura 18 - Representação UML da taxonomia SBO de pontes semânticas	61
Figura 19 - Ontologias do Sistema Legado e SCM representadas no MAFRA Toolkit	66
Figura 20 - ConceptBridges e especificação extensional	67
Figura 21 - Representação das transformações dos identificadores do Sistema Legado para o SCM ..	69

Figura 22 - Pontes de propriedades de Terceiro.....	70
Figura 23 - PontePropriedades com Serviço CopyRelation.....	71
Figura 24 - Extracto da migração de dados com mapeamento.....	72
Figura 25 - Resultado da transformação sobre os dados legados.....	72
Figura 26 - Cronograma das diferentes fases e tentativas de migração.....	78
Figura 27 - Novo modelo do processo no grupo.....	81
Figura 28 - Cronograma da Integração/Migração Continuada.....	82
Figura 29 - Novo processo de integração de dados do negócio “em produção”.....	83
Figura 30 - Extensão a Mapeamento de terceiros do Sistema Legado para o SCM.....	85
Figura 31 - Condições em Synon.....	94
Figura 32 - Expedição e Facturação para destinos diferentes.....	103
Figura 33 - Novo modelo do processo no grupo.....	106
Figura 34 - Ferramenta para Integração do Sistema Legado com SCM.....	109
Figura 35 - Interface para a tabela de mapeamento de terceiros entre sistema Legado e SCM.....	110
Figura 36 - Identificação de mapeamentos de terceiros.....	112
Figura 37 - Número de contribuinte igual para clientes diferentes (ITO e O C M).....	113
Figura 38 - Duplicação de um NIF em entidade diferentes.....	113
Figura 39 - Cliente O C M só com movimentos por Factoring.....	114
Figura 40 - Registos do terceiro ITO com números de contribuinte diferentes.....	115
Figura 41 - Extensão a Mapeamento de terceiros do Sistema Legado para o SCM.....	116
Figura 42 - Mapeamento de Terceiros ordenados por nome.....	117
Figura 43 - Mapeamento de Terceiros ordenado por Código SCM.....	118
Figura 44 - Script SQL que detecta Entidades diferentes com mesmo NIF.....	119
Figura 45 - Lista de Terceiros com o mesmo NIF.....	120

Abreviaturas

API	Application Programming Interface – Interface de Programação de Aplicações
BD	Base de Dados
BDR	Base de Dados Relacional
CRM	Customer relationship management – Gestão do relacionamento com clientes
ERP	Enterprise Resource Planning – Planeamento de Recursos da Empresa
FTM	Ficha Técnica da Malha
ID	Identificação
ITV	Indústria Têxtil e do Vestuário
JVM	Java Virtual Machine
KB	Knowledge Base – Base de Conhecimento
PME	Pequenas e Médias Empresas
POC	Plano Oficial de Contas
SAFT-PT	Standard Audit File for Tax purposes - Portugal – Ficheiro padrão de auditoria fiscal
SBO	Semantic Bridging Ontology – Ontologia de Pontes Semânticas
SCM	Supply Chain Management – Gestão da Cadeia de Fornecimento – neste documento representa também o novo sistema implementado, por oposição ao sistema legado, no caso prático
SGBD	Sistema de Gestão de Base de Dados
SGBDR	Sistema de Gestão de Base de Dados Relacional
SOA	Service-Oriented Architecture – Arquitectura Orientada ao Serviço

1 Introdução

Os sistemas legados representam um importante capital técnico, funcional, cognitivo e histórico das organizações. Estes sistemas são frequentemente menosprezados e mesmo ignorados com respeito ao seu papel no negócio central. A constante evolução operada nos sistemas de informação na sequência das alterações do negócio envolve por um lado a manutenção e actualização de equipamentos, e por outro a evolução de aplicações em subsistemas, no conjunto dos sistemas que vão compondo ao longo do tempo o parque tecnológico dos sistemas de informação. A não evolução dos subsistemas legados pode representar uma perda ou um custo de reengenharia de funcionalidades anteriormente implementadas.

As dificuldades na manutenção e desenvolvimento contínuos compreende o desconhecimento das plataformas computacionais implementadas, das linguagens de programação, dos sistemas de persistência de dados em geral e bases de dados em particular, e das tecnologias de comunicação empregues.

Em geral estes sistemas executam programas já descontinuados, fora do circuito comercial, e que por vezes já não existe contacto com os fabricantes, representantes ou programadores. Frequentemente os profissionais que conceberam ou programaram os sistemas nas organizações com desenvolvimento interno já não se encontram disponíveis. A documentação está desactualizada ou é de difícil compreensão pela falta de especificação ou pela falta de conhecimentos de quem consulta esses documentos.

Vulgarmente, as plataformas de hardware legado já não são suportadas pelos fabricantes e a substituição por equipamentos mais recentes pode estar impossibilitada por incompatibilidade do software.

Contudo, os sistemas são muitas vezes de enorme importância para o negócio, o que é percebido na hora de fazer evoluir o negócio ou o sistema de informação. Contabilizados os custos de substituição, percebe-se que estes são muitas das vezes inoportáveis com o orçamento disponível, daí que sejam equacionadas outras alternativas.

Os sistemas de informação legados dispõem de dois importantes componentes:

1. Dados, armazenados em SGBD, formatos proprietários ou já não suportados;
2. Lógica de negócio, em linguagens de programação antiquadas, de difícil ou impossível migração.

Além disso, estes dois componentes estão, enúmeras vezes, entrançados entre si no sistema, de tal forma que a sua separação e migração independente se torna tecnologicamente difícil e financeiramente irrealista.

A integração destes sistemas e dos seus dados com outros mais recentes tende a ser tão mais desejável quanto maior o número de sistemas com funcionalidades específicas num domínio, e quanto maior a necessidade de interacção entre organizações que colaboram em cadeia de fornecimento, fabricação ou projecto.

O desenvolvimento de sistemas e a utilização de práticas que facilitem e promovam a integração apresenta-se com importância crescente e como oportunidade de negócio, integrando as prioridades de consórcios de fabricantes e produtores de software.

Esta dissertação tem como objectivo analisar e descrever diversas alternativas que se colocam nestes cenários, através da aplicação a um caso e da comparação entre elas.

1.1 Caso de Estudo

Neste documento será abordado um caso de estudo específico, retirado dum cenário real de evolução dum sistema de informação numa grande empresa do ramo da Indústria Têxtil e do Vestuário (ITV), com mais de 50 anos de existência e com sistemas de informação desde 1980.

Em termos de operação, a empresa recebe encomendas de roupa em malha, produz ou subcontrata a produção total ou parcial dos artigos encomendados e remete-os ao cliente.

1.1.1 Sistema de Informação Legado

A empresa desenvolveu internamente e ao longo das três últimas décadas, um ERP em sistemas IBM AS/400 e arquitecturas Intel x86 em ambientes DOS e Windows da Microsoft.

Este sistema de informação opera num processador RISC IBM AS/400 modelo 600 com o sistema operativo OS/400 V4R4M4, e aplicações em plataforma x86 Cliente Servidor.

Os programas AS/400 estão desenvolvidos em RPG/400 e CL/400, e na ferramenta CASE Synon 2E, que gera programas RPG/400. É utilizada a base de dados relacional nativa AS/400, nos formatos de ficheiros físicos e lógicos multi-membro.

As aplicações x86 são desenvolvidas em WinDev 4.5 e Java, e interligam-se com o servidor AS/400 através de interfaces ODBC e API EasyCom para WinDev.

O sistema de informação da empresa, e em particular o ERP, evoluíram ao longo das últimas décadas, adaptando-se às necessidades e especificidades da organização ao longo do tempo.

1.1.2 Sistema ERP/SCM

Recentemente a organização operou um processo de desverticalização, tornando as tradicionais secções produtivas em unidades autónomas na forma de organizações independentes mas cooperantes, criando assim oportunidades de diversificação no mercado.

Este processo, muito comum neste ramo da indústria do norte do país, infligiu mudanças no *layout* da organização passando de uma estrutura orientada para o modelo produtivo com planeamento integral das diferentes fases, para uma estrutura de cadeia de fornecimento.

Tendo em vista o suporte ao novo paradigma organizacional, foi adquirida uma nova solução de gestão empresarial que responde a mais necessidades do actual paradigma da organização que o sistema ERP legado, apresenta um melhor desempenho na execução das tarefas correntes, e permite uma evolução para a gestão integrada da cadeia de fornecimento.

O novo sistema de informação opera num servidor multi-processador Intel Xeon com o sistema operativo Windows 2003 Server, utilizando um sistema de gestão de base de dados SQL Server 2005 da Microsoft.

A solução ERP/SCM seleccionada é a GM-IC da Macwin Sistemas¹. O software é desenvolvido no paradigma de objectos em Omnis e o código fonte é disponibilizado à organização. Está em desenvolvimento uma interface Web em PHP para servir de portal aos fornecedores na cadeia de distribuição.

1.2 Organização do documento

O restante documento está organizado da seguinte forma. O capítulo 1 apresentou o contexto de negócio e tecnológico do caso prático que serve de base ao documento. O capítulo 2 apresenta diversas alternativas conceptuais à evolução/integração de sistemas de informação legados, bem como alternativas tecnológicas específicas disponíveis na literatura e no mercado (e.g. TAO, TOSEM, Attachmate). O capítulo 3 descreve detalhadamente vários problemas sentidos no processo de migração de parte dos dados do sistema legado para o novo sistema, bem como abordagem seguida na sua resolução. O capítulo 4 descreve o processo de migração dos mesmos dados usando uma abordagem baseada em ontologias e mapeamento declarativo. Este capítulo adquire fundamental importância porquanto permite a comparação entre a abordagem prática informal, com uma abordagem mais erudita mas pouco empregue e testada em ambientes industriais. O capítulo 5 apresenta o esforço que se seguiu para obviar limitações observadas, sugerindo uma abordagem de uso complementar e ortogonal dos dois sistemas. No capítulo 6 apresentam-se as conclusões sobre o trabalho realizado. Finalmente no capítulo 7 perspectiva-se trabalho futuro neste âmbito.

¹ Um sistema fortemente orientado para a ITV, desenvolvido por uma pequena organização com contacto muito estreito com os clientes neste ramo da indústria têxtil. <http://www.macwin.pt>

2 Contexto tecnológico

2.1 Objectivos

O sistema de informação desta organização, inicialmente orientado para o planeamento vertical, seguimento e controlo de ordens de fabrico para satisfazer encomendas, torna-se um sistema demasiado pesado em detalhes e cálculo que não se ajusta ao modelo de cadeia de fornecimento pretendido, sendo constantemente readaptado para simplificar a operação e responder a novas questões. Este refinamento da organização foi razão para a evolução do sistema implementado, considerando-se a implementação de um novo sistema ERP, que irá garantir as operações no novo paradigma organizacional.

Apesar das diferenças de arquitectura de informação e representação da informação dos dois sistemas, existem elementos comuns passíveis de equivalência que importa manter. De facto, o sistema em operação é rico em elementos da cultura industrial no ramo, que se poderão perder já que não são representados totalmente nos sistemas mais recentes.

Neste processo de mudança/crescimento pretende-se transferir a grande maioria das funções para o novo sistema, mantendo algumas funcionalidades específicas no sistema legado.

Um conjunto de dados irá migrar para o novo sistema sendo abandonada a sua exploração no sistema legado, e um mais restrito manter-se-á ainda no sistema legado, já que não se vêm representados todos os seus atributos no novo sistema SCM.

A integração sustenta-se na comunicação e partilha entre os sistemas legados e os novos sistemas implementados.

Deverão então ser implementados mecanismos que, de forma automática ou semi-automática, reduzam a intervenção humana usual na migração e integração de dados, e suportem:

- conectividade;
- mapeamentos;
- transformações de dados;
- sincronismo entre os sistemas.

O domínio de conhecimento comum e as diferenças conceptuais, representam uma oportunidade para estudar as metodologias adoptadas para a migração de dados, integração, mapeamento e equivalência de entidades entre o sistema legado e o novo sistema.

2.2 Abordagens técnico-organizacionais

Num processo de evolução do sistema de informação, a primeira questão que se coloca diz respeito à migração ou substituição do sistema.

Na evolução dos sistemas de informação de uma organização, podem encontrar-se três cenários quanto à utilização dos dados e informação residentes nos sistemas:

- Substituição do sistema computacional – Trata-se da substituição da totalidade ou de parte dos sistemas (hardware e software) por um equipamento e aplicações novas, sem haver lugar à partilha, migração ou qualquer reutilização de dados ou de informação. Nesta abordagem ocorre uma definição e configuração do sistema novo e abandono puro e simples do sistema anteriormente usado;
- Migração de dados – Os equipamentos e software são substituídos mas existe um processo que migra, ou replica, para o novo sistema os dados dos sistemas que se irão abandonar. Este processo poderá abranger total ou parcialmente os dados existentes no sistema de partida. Esta opção depende:
 - da existência de possibilidade física de transporte¹ de dados entre os sistemas envolvidos;
 - da aplicabilidade dos dados em novos paradigmas organizacionais que contextualizam as mudanças de sistemas de informação;
 - da natureza e adequação dos dados às aplicações existentes nos novos sistemas;
 - da complexidade da migração – pode implicar custos incontroláveis;
 - do orçamento para o processo de migração.
- Integração – Os equipamentos e software mantêm-se, no todo ou em parte, e são adicionados ao conjunto novos elementos de hardware e software.

Os sistemas complementam-se, sendo que o sistema de informação é a soma dos subsistemas integrados. Neste modelo são construídas ferramentas para a partilha e sincronismo de dados entre os sistemas, que operam de forma continuada. A integração pode contemplar a migração de sistemas, i.e. áreas do sistema de informação que migrarão de subsistema.

A arquitectura de integração pode apresentar diferentes desenhos de acordo com a natureza e dispersão das fontes de dados, temporização e disponibilidades dos dados e cadeia de utilização dos dados e informação.

Esta abordagem depende igualmente das condições referidas para a Migração de dados, embora, em geral, com complexidade e volume de dados numa escala superior.

No contexto actual em que o desenho dos modelos de negócio sofrem constantemente mutações e os objectivos são realinhados com as mudanças da organização e do meio, as organizações ponderam bastante a abordagem a adoptar, considerando os resultados esperados e a sua adequação ao modelo de negócio no futuro próximo. A Tabela 1 descreve e compara de forma resumida as três abordagens.

¹ A designação de “Transporte de dados” neste contexto pretende ilustrar a imagem de “deslocamento” dos dados entre locais (sistemas), os dados atravessam canais de um sistema origem para um outro de destino.

	Substituição	Migração	Integração
Continuidade do sistema	Não	Não	Sim
Persistência dos dados	Perdidos	São transferidos	Sim
Processo	uma fase	uma ou várias fases (iterativo)	Contínuo
Intervenção humana sobre os dados	Inserção nos novos sistemas dos dados necessários	Só para complemento ou correcção dos dados	Só para complemento ou correcção dos dados
Preservação de valores estatísticos ou informação histórica	Não ou Inserida nos novos sistemas	Sim desde que representável no novo sistema	Sim
Duração do processo	Rápido	Médio	Longo

Tabela 1 - Modelos de crescimento de sistemas de informação

A Substituição é opção quando os recursos disponíveis são demasiado escassos; os processos são demasiado complexos nas transformações de dados do sistema legado para o novo sistema; os processos falham tecnicamente ou no tempo consumido no seu desenvolvimento, ou não existe capacidade técnica nos agentes que operam a transição. É o modelo aparentemente mais económico, mas com custos que dependem das perdas de dados, muitas vezes intangível, ou nas tarefas de digitação dos dados considerados indispensáveis.

A opção pelo modelo de Migração faz-se tentando uma forma económica de manter os dados no novo sistema, mas tendo em vista o abandono do sistema legado. Com um investimento reduzido poupam-se horas de digitação de dados já existentes em sistemas, preservam-se dados considerados fundamentais e reduzindo os custos de manutenção de sistemas.

A Integração é adoptada quando existe uma forte motivação para a preservação da cultura da organização, mantendo dados e funcionalidades dos sistemas legados em cooperação com os novos sistemas implementados. Contudo, é corrente que os projectos de integração de sistemas são mais demorados, envolvem mais recursos humanos e técnicos, e financiamentos mais avultados, sendo por isso pouco usuais em organizações com mais limitações orçamentais e de recursos.

As dificuldades em integrar e as vantagens de integrar não são facilmente contabilizadas, mas podem ser sistematizadas nas seguintes dimensões:

- Os sistemas legados mantêm em operação funcionalidades desejáveis, pelo que o seu abandono não é recomendável;
- A evolução dos sistemas legados é difícil, pois os colaboradores que conceberam ou programaram os sistemas nas organizações com desenvolvimento interno já não integram a organização e a documentação, quando existe, está desactualizada ou é de difícil compreensão, quer por falta de especificação quer pela falta de conhecimento de conceitos de base ou cultura no domínio de conhecimento;

- Paradigmas de programação e sistemas de bases de dados nos sistemas legados que não contemplam funcionalidades autónomas, hoje vulgarizadas como a integridade relacional, são uma dificuldade para a integração dos SGBD legados em novos projectos;
- Evolução estratégica e social das organizações incentiva a implementação de novas soluções e a renovação do parque tecnológico de forma continuada;
- O hardware legado foi descontinuado e opera sem manutenção preventiva, pelo que se torna um factor de risco no funcionamento da organização, e como tal promove o seu desmantelamento.

Na conceptualização de Migração e Integração há que distinguir sobre cinco processos, a seguir descritos na Tabela 2, reflectindo na acção (migração e integração), e na natureza do objecto (dados e aplicações):

Designação	Conceptualização
Migração de dados	Refere-se à operação de cópia de dados, com ou sem transformação, de um repositório de dados para outro.
Migração de utilização de aplicações	Refere-se ao abandono da utilização de uma aplicação para dar lugar à utilização de uma outra aplicação, semelhante à primeira, e globalmente com objectivos produtivos semelhantes. Pode combinar-se com Migração de dados.
Migração de aplicações	Refere-se à reescrita de uma aplicação, normalmente para um sistema diferente do original. Pode ser combinada com Migração de dados.
Integração de aplicações	Refere-se à utilização de aplicações e sistemas diferentes que partilham ou complementam dados e funções entre si através de peças de software produzidas para esse efeito, não havendo abandono de nenhum dos sistemas integrados.
Integração de dados	Refere-se ao acesso e utilização de dados por uma aplicação, que residem num sistema diferente do da aplicação ou que são originalmente gerados ou mantidos por uma outra aplicação.

Tabela 2 - Processos de migração e integração

Há portanto que distinguir claramente entre migração/integração de dados e migração/integração de aplicações.

Nas secções seguintes exploram-se algumas abordagens à Migração e Integração de dados e de informação, assim como alguns exemplos de projectos em curso nesta temática.

2.3 Integração de informação

A integração de informação pretende possibilitar a partilha de informação entre sistemas diferentes, tornando simples e tão unificado quanto possível o acesso à informação dispersa, encapsulando a complexidade dos mecanismos que possibilitam esta partilha e visão consolidada.

Na Figura 1, baseada em (Jérôme Euzenat e Pavel Shvaiko 2007), esquematiza-se um cenário genérico de integração de informação.

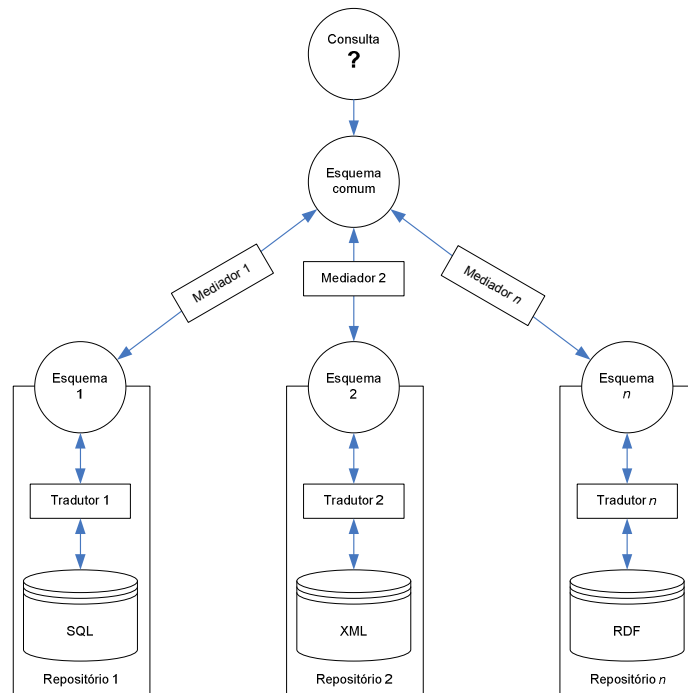


Figura 1 - Um cenário genérico de integração de informação

As fontes de dados, baseadas em repositórios de dados tecnologicamente heterogéneos, nomeadamente em bases de dados que fornecem dados em formatos distintos, e.g. *SQL*, *XML*, *RDF*, etc., são confrontadas com um repositório comum. Os mediadores transformam as consultas geradas de acordo com o esquema do repositório comum em consultas formatadas para cada repositório local das fontes, e traduzem também o resultado no sentido inverso.

O objectivo é possibilitar que consultas a um repositório de dados de domínio sejam traduzidas para repositórios de dados locais, com as devidas adaptações decorrentes de estruturas ou semântica distintas.

O processo de integração de informação implica, em geral:

- Identificar as correspondências entre as entidades dos esquemas locais e comuns, semanticamente relacionadas. A este respeito existem várias abordagens possíveis, nomeadamente as baseadas em vistas sobre os esquemas locais ou sobre os esquemas globais, como representado na Figura 2 a título ilustrativo.

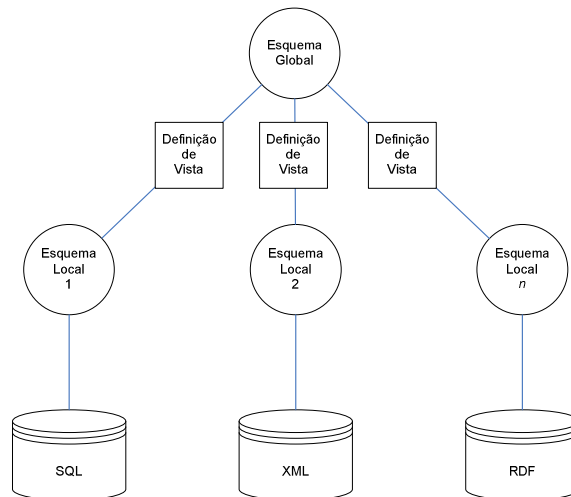


Figura 2 - Integração – Abordagens à identificação de correspondências sobre vistas

- Interpretar a consulta (*query*) elaborada consoante repositório comum e reescrevê-la de acordo com cada um dos repositórios e esquemas locais;
- Traduzir as instâncias relevantes das fontes para o formalismo da informação no sistema integrado;
- Reconciliar os resultados obtidos nas diferentes fontes a fim de não ter repetições, nomeadamente detecção de duplicados.

2.3.1 Descrição dos dados

Para o planeamento do processo de integração referido atrás, conta-se com descrições mais ou menos formais do significado dos dados, tal como os clássicos dicionários de dados e os esquemas das bases de dados, assim como a documentação em geral que descreve os sistemas aplicativos.

Estes documentos, nomeadamente os esquemas de bases de dados, apesar de se mostrarem um suporte imprescindível nos trabalhos de migração e integração, revelam-se insuficientes para uma descrição clara dos dados e completa dos sistemas em geral. Não captam informação suficiente que permita uma boa proximidade entre a conceptualização do sistema descrito e os conceitos do domínio inerentes à lógica de negócio, e por isso não garantem uma migração ou integração correcta.

De outra forma, esquemas de base de dados não são suficientes para descrever a lógica dos sistemas, nem as transformações e equivalências necessárias à migração ou integração de dados de um repositório para outro.

No sentido de estender a descrição de sistemas tendo em vista a integração, constata-se que as ontologias revelam-se uma ferramenta importante que, descrevendo formalmente os sistemas, “publicam” os conceitos que os sistemas representam numa linguagem mais próxima da natural com rigor, já que “traduzem” as suas representações proprietárias em representações do domínio do conhecimento “público”. Deste modo, os sistemas tornam-se partilháveis, uma vez que se encontra uma linguagem de interface, conceptual, que pode interligar sistemas de natureza diferentes num plano

superior aos dos esquemas de base de dados. É pois previsível que esta tecnologia na área do conhecimento se torne um facilitador na migração e integração de sistemas num domínio de conhecimento.

2.3.2 Tipos de sistemas

Na descrição anterior usaram-se três conceitos que actuam em transformação, mediação e distribuição:

- Wrapper (Hull 1997:53; Philippe Thiran et al. 2006:331) é um componente do sistema, capaz de traduzir (sem alteração de estrutura ou semântica) entre linguagens de representação (e.g. tabular para hierárquico) ou modelos de dados (e.g. modelo relacional para modelo XML), focado ao nível da heterogeneidade de plataformas;
- Mediador (Hull 1997:53). Um mediador é um sistema que suporta uma visão integrada sobre múltiplas fontes de informação. O Mediador (*mediator*) conhece um esquema usado para as consultas assim como os esquemas dos repositórios de dados que consulta. Um mediador também pode ser um fornecedor de informação.
- Broker (Jérôme Euzenat e Pavel Shvaiko 2007:22). Um Broker é uma peça de software que concentra pedidos e os distribui por fornecedores de conteúdos conhecidos. Recebe consultas, entrega-as a mediadores que as traduzem para os sistemas locais, e tem a incumbência de agregar os dados resultantes das consultas aos repositórios locais, garantindo, entre outros, a unicidade das instâncias que irão compor o resultado devolvido.

Nas secções seguintes apresenta-se uma descrição das três abordagens à integração de informação proposta por (Jérôme Euzenat e Pavel Shvaiko 2007).

2.3.3 Integração de esquema (migração de dados)

É o cenário usual de integração (*schema integration*), habitualmente designado por migração de dados (cf. Tabela 2), bastante utilizado quando o objectivo é produzir uma base de dados final a partir de duas ou mais originais. É frequente encontrar-se este cenário no contexto de uma junção de organizações, e.g. alienação na banca, numa actualização de sistemas e tecnologias de informação ou na integração de novos subsistemas que irão alimentar áreas de um sistema já em produção.

Em geral este processo é motivado pelas mudanças da organização e promovido com entusiasmo por novos colaboradores por não dominarem os sistemas instalados nem reconhecerem a imagem da organização nesses sistemas, e porque o sistema legado não contempla a totalidade dos modelos a implementar por estes.

Normalmente uma das bases de dados originais é a integradora, mantendo grande parte do seu esquema original. As restantes contribuem normalmente com dados instanciados, embora por vezes também com estruturas que irão enriquecer o esquema da base de dados final.

O processo passa por transportar (existe realmente a alteração de local de armazenamento dos dados) os dados úteis do sistema legado para os novos sistemas. Este transporte exige um estudo de

mapeamento/equivalência de tabelas e campos, tradução de codificações e conversão de tipos. São encontradas dificuldades no mapeamento de atributos entre os esquemas, e.g. atributos de tipo e dimensão diferente para um tipo de documento contabilístico; e de tradução de valores instanciados, e.g. codificação de tipos de famílias de produtos diferentes nos repositórios dos dois sistema. Ainda que no mesmo domínio, os esquemas apresentam diferenças porque as origens dos sistemas produzidos são distintas, foram desenvolvidos por equipas diferentes e os objectivos das organizações que os produziram serão também díspares.

É um processo executado para o arranque do novo sistema, pelo que o processo de migração dos dados ocorre uma só vez.

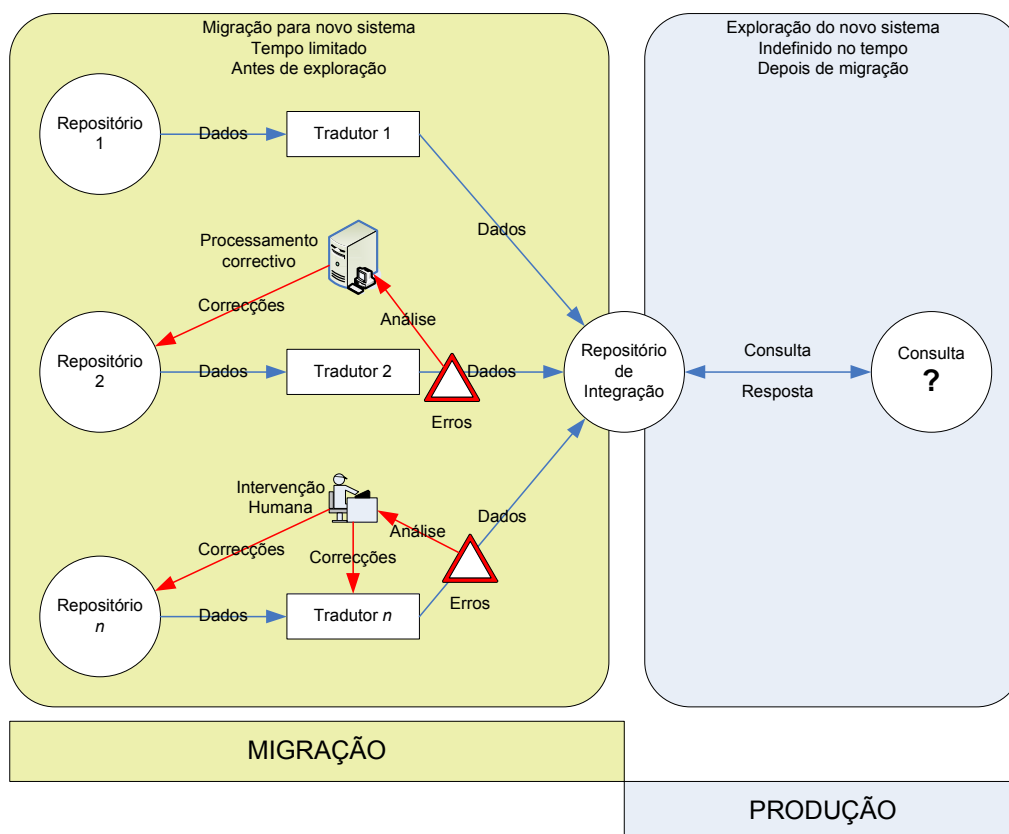


Figura 3 - Integração de esquema

Globalmente trata-se de um único processo de migração, por lotes (*batch*), composto de inúmeros processos por lotes e interactivos, executados iterativamente até à obtenção de um resultado aceitável. Concluída a migração, o novo sistema entra em produção por tempo indeterminado.

2.3.4 Integração de catálogos

Esta abordagem (*catalogue integration*) é apresentada no problema da manutenção de catálogos de fornecedores de bens que contribuem para um catálogo geral de uma organização que os comercializa. Trata-se dum modelo actual e comum no mercado electrónico na Web, bem como no campo das Datawarehouses.

Esta integração assenta no conceito de correspondências entre catálogos, ou esquemas de dados numa perspectiva de fornecedores de informação e serviços para um espaço comum, partilhado ou explorado por uma comunidade.

Conforme explícito na Figura 4, cada fornecedor confronta o seu catálogo (*Repositório₁* a *Repositório_n*) com o do mercado (*Repositório comum*), e dessa verificação é gerado um tradutor (*tradutor_i*) do seu catálogo para o mercado.

Com o tradutor, os fornecedores actualizam o catálogo do mercado no sentido da Base de Dados local a integrar para a que tem o mercado integrado.

Os consumidores, ou utilizadores, consultam o mercado e obtêm respostas deste, baseadas nos dados agrupados no repositório central.

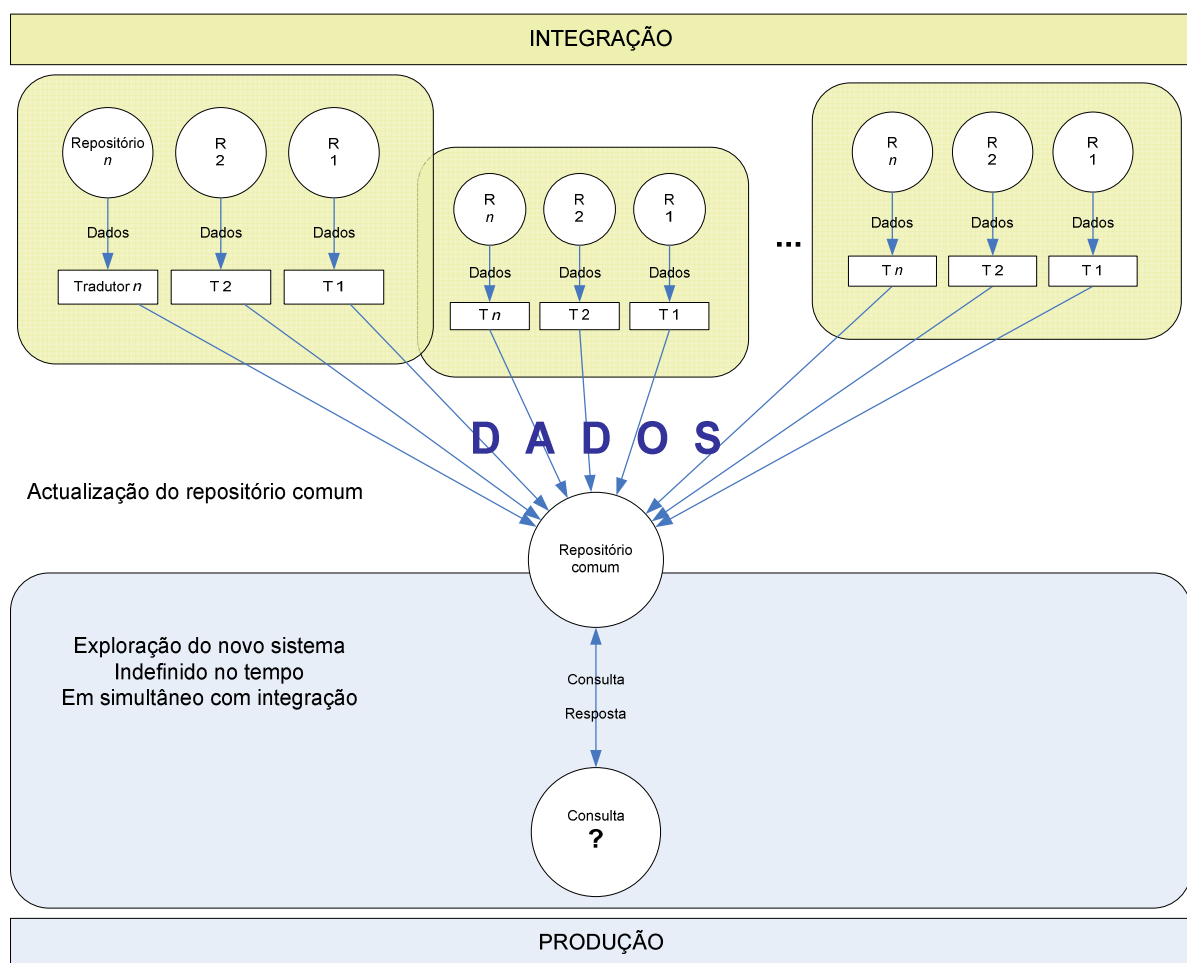


Figura 4 - Cenário de Integração de Catálogo com correspondência

A principal característica deste tipo de abordagem por contraponto à integração de esquemas respeita à periodicidade de actualização do repositório central com os dados dos repositórios locais. De facto, as actualizações são realizadas periodicamente, não havendo portanto garantias de que os dados no repositório global estão actualizados. Sendo assim, os fornecedores deverão implementar procedimentos de sincronismo do seu repositório com o da comunidade que reflectam de forma rápida

e amiúde as actualizações dos seus produtos no repositório comum, sob pena de não concorrerem no mercado ou de participarem com produtos e/ou valores desactualizados.

2.3.5 Integração de dados

Este cenário (*data integration*) (cf. Tabela 2) pressupõe que a informação é acedida através do repositório comum, vinda de múltiplas fontes, sem primeiro os fornecedores carregarem os dados num armazém central, i.e. o acesso aos dados é realizado “a pedido” (*on-demand*).

Os fornecedores de informação estão identificados e possuem os seus esquemas locais, assim como as bases de dados privadas (cf Figura 5). Estas não alimentam previamente o sistema de integração. Isto permite interoperabilidade entre várias fontes tendo a comunidade acesso a dados actualizados.

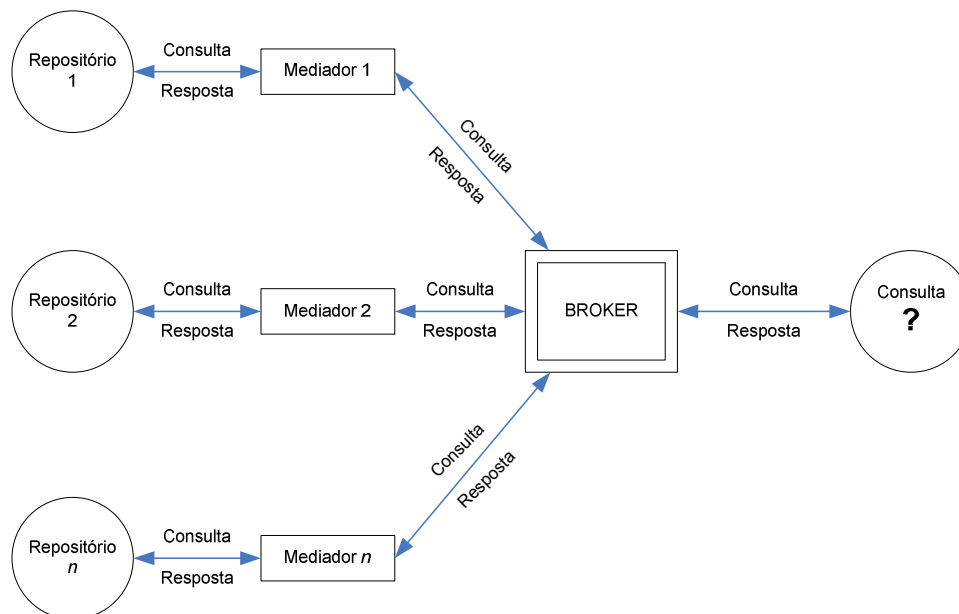


Figura 5 - Cenário com Integração de Dados

As consultas são colocadas a um *broker* que as despacha para os mediadores específicos de cada repositório local. Os mediadores são responsáveis por reescreverem a consulta (*query*) original de acordo com o esquema local e as relações semânticas (correspondência) previamente estipuladas entre repositórios.

A correspondência entre entidades semanticamente equivalentes é estabelecida pelo processo de *matching*. Normalmente a fase de *matching* é processada em *off-line*, gerando os mediadores usados no momento das consultas e respostas.

O processo de integração de informação é dinâmico e depende em larga medida da capacidade do Mediador em transformar a *query* original ao esquema local, otimizando a pesquisa. O Broker tem a incumbência de agregar os dados resultantes das consultas aos repositórios locais, garantido (entre outros) a unicidade das instâncias.

2.3.6 Outras abordagens

Segundo a literatura, existem outras abordagens de integração de informação, nem sempre consensuais entre si. De facto, apesar das categorias identificadas por diversos autores terem características em comum, existem outras categorizações disponíveis que enfatizam outras dimensões do problema. Por exemplo (Hull 1997), enfatiza a categorização dos cenários segundo a possibilidade de consulta e actualização dos repositórios centrais:

- Vistas integradas de leitura: Mediação; (*Integrated read-only views: Mediation*). Nesta arquitectura (Figura 6) visa-se suportar uma vista integrada e unificada, só de leitura, de informação dispersa em múltiplas bases de dados. Esta abordagem é conceptualmente equivalente à integração de dados descrita anteriormente;

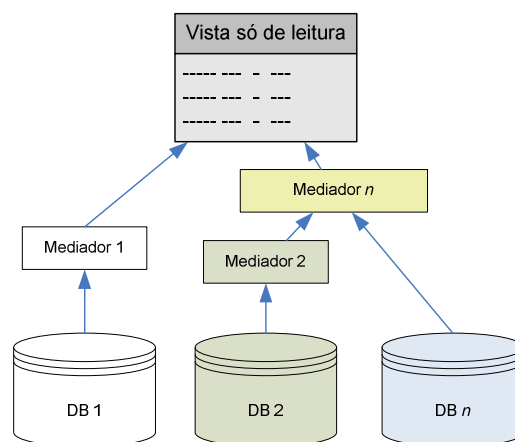


Figura 6 - Mediação - Vista integrada só de leitura

- Vistas integradas de leitura e escrita: Mediação com alteração (*Integrated read-write views: Mediation with update*). Será uma natural extensão da arquitectura de mediação de leitura, contudo com limitações próprias a alteração sobre as vistas. As limitações na escrita estarão relacionadas com as habituais limitações nos SGBD de actualização de dados sobre vistas com ligações e consolidação de tabelas, com a propagação de alterações a conjuntos de registos, etc. (Figura 7);

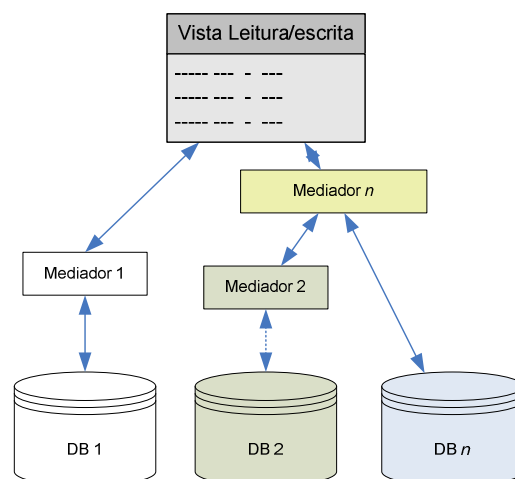


Figura 7 - Mediação com leitura e escrita

- Múltiplas bases de dados partilhando informação: Federação. (*Multiple databases sharing information: Federation*). Proposta por (Sheth e Larsen 1990), em contraste com a integração, a arquitectura federativa (Figura 8) oferece um ambiente onde várias bases de dados heterogéneas se juntam numa federação. Cada base de dados, como membro da federação, estende o seu esquema para incorporar subconjuntos de dados mantidos nas outras bases de dados membro. Nesta abordagem é possível a actualização de dados nas bases de dados que participam na federação, o que deve ser previsto nos elementos de software que suportam a implementação.

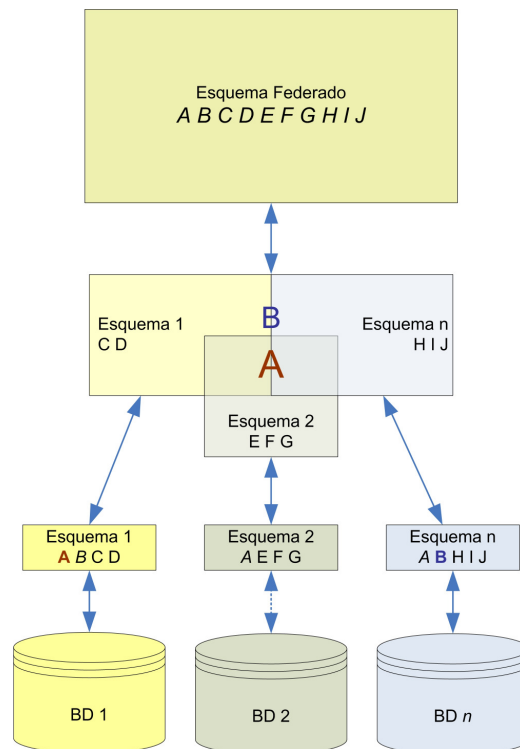


Figura 8 - Federação de Base de dados

A Figura 8 representa um cenário federativo, na perspectiva da disponibilização de dados. Na representação existem três bases de dados (DB 1; DB 2; DB n). O esquema local de cada base de dados apresenta um subconjunto de elementos, ou atributos, de um conjunto (A; B; C; D; E; F; G; H; I; J) que se pretende disponibilizado no esquema da federação. As bases de dados que participam na federação tanto fornecem como colhem conteúdos da federação, e portanto das demais bases de dados federadas.

O atributo A é original da Base de dados BD 1 (a cor vermelha) e o atributo B é da BD n (a cor azul). As bases de dados BD 2 e BD n estendem os seus esquemas representando A de BD 1 (em itálico), e a base de dados BD 1 estende o seu esquema representando B de BD n (em itálico). As bases de dados deverão individualmente proporcionar a partilha do seu esquema, e absorver dos esquemas federados as estruturas que permitam o relacionamento entre os dados que representam e os dados disponibilizados pelas restantes bases de dados. Para isso existirão

componentes nas bases de dados que produzem transformações das representações dos seus esquemas locais em representações partilhadas que serão combinadas ao nível da federação.

- Coordinating multiple databases: Workflow (Georgakopoulos , Hornick, e Sheth 2008) Muitas organizações têm vários repositórios de dados integrados no funcionamento organizacional. De uma perspectiva semântica a interacção entre essas bases de dados pode ser classificada como usando o paradigma *workflow*. Na Figura 9 apresenta-se uma metáfora em que os processos de fluxo coordenam a actualização das diferentes bases de dados e exploram-nas no sentido do controlo e da continuidade de fluxo.

Os processos de fluxo poderão ser totalmente electrónicos ou incluir intervenção humana, mas geram sempre eventos no sistema que desencadeiam as mudanças de estado e os avanços no circuito de fluxo.

O modelo apresentado na Figura 9 assenta sobre bases de dados, mas o conceito pode ser alargado a repositórios de informação e sistemas que encapsulam estruturas mais complexas que, no seu conjunto, representam os objectos manipulados no negócio (e.g. encomendas, artigos, procedimentos), o seu estado, os eventos que contribuem para as mudanças de estado (e.g. recepção de mercadorias, mensagens de aprovação, avisos de situações irregulares, pagamentos) e os seus actores (e.g. fornecedores, clientes, colaboradores, processos, etc.). Os repositórios de dados passam a estar integrados no contexto dos processos de fluxo passando estes a constituir o agente integrador entre os repositórios de dados, ou bases de dados.

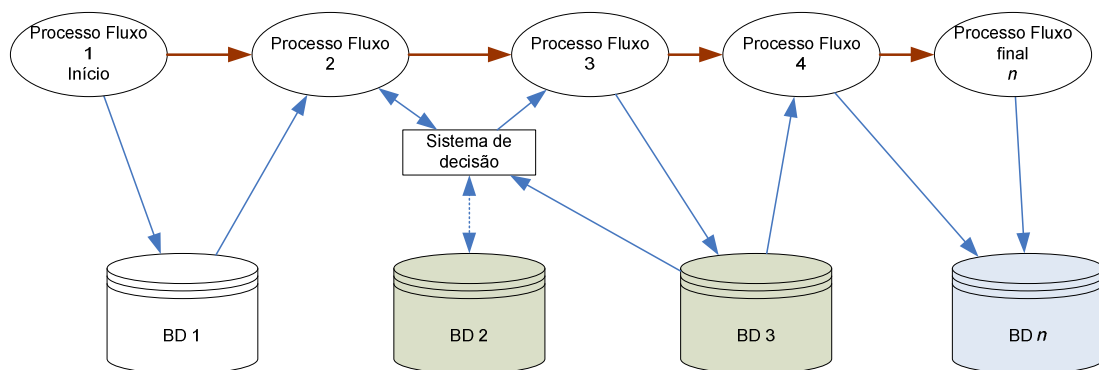


Figura 9 - Coordenação de múltiplas bases de dados utilizando *workflow*

2.3.7 Comparação de abordagens

Nesta secção apresenta-se uma comparação entre as três abordagens de integração de informação apresentadas nas secções anteriores. Para isso, cada cenário foi analisado segundo quatro dimensões:

- Sentido, diz respeito à iniciativa de integração a cada momento, i.e. qual o sistema que inicia o processo integrador, numa perspectiva de “entrega de informação/actualização” ou de “solicitação de informação/actualização”;
- Momento, diz respeito à frequência com que o processo de integração ocorre.

- Processamento, diz respeito ao modo de execução do processo de integração, numa perspectiva de autonomia, intervenção e iniciativa de execução:

Lotes (*batch*) – quando o processo é executado sem interacção com um utilizador, uma ou mais vezes, sendo iniciado automaticamente ou por acção de um utilizador, processando um conjunto de dados, também designado por unidade de processamento;

Interactivo – quando o processo é interrompido para obter intervenções continuando de seguida a partir do ponto de paragem. Normalmente o processo é iniciado por um utilizador;

On-line – tempo-real, quando um processo é invocado sempre que necessário para processar uma unidade de processamento;

- Integrador, diz respeito à entidade, site ou peça de software que toma a iniciativa de executar o processo de integração, em qualquer uma das abordagens apresentadas.

A Tabela 3 resume as diferentes características das três abordagens em função das dimensões anteriores.

Abordagem	Sentido	Momento	Processamento	Integrador
Integração de Esquema (<i>Schema integration</i>)	De qualquer um dos sistemas envolvidos para o outro	Uma só vez Poderão existir várias iterações por correcções ou refinamentos. Um batch no todo (numa observação macro)	Lotes (<i>batch</i>) Se existirem várias iterações, cada execução será em batch (muitos batches)	Processo de migração, ou agente integrador
Integração de catálogo (<i>Catalogue integration</i>)	Fornecedores de dados actualizam BD de mercado ou comunidade	Periodicamente	Lotes (<i>batch</i>)	Cada fornecedor junto da comunidade
Integração de dados (<i>Data integration</i>)	Mercado procura dados nos fornecedores no momento da consulta	A pedido (<i>on-demand</i>)	On-line	Broker, recolhe informação dos fornecedores e despacha as consultas para mediadores.

Tabela 3 - Comparação entre as abordagens à integração

A Tabela 4 categoriza e resume as diferentes abordagens à integração de informação nas dimensões:

- Processo, diz respeito à forma de acesso e tratamento dos dados;
- Ocorrências, diz respeito à frequência de utilização do processo;
- Actualização dos dados, é um indicador da qualidade e actualidade dos dados no sistema na perspectiva do momento em que são explorados;
- Materializável, diz respeito à existência física dos dados no sistema integrador, no sistema que disponibiliza os dados ao cliente final;
- Tipos de interacção, diz respeito às acções possíveis sobre os dados integrados.

Abordagem	Processo	Ocorrências	Actualização dos dados	Materializável	Tipos de interacção
Integração de esquemas	Transformação	1	Garantida após a ocorrência	Sim	Leitura
Integração de vistas de leitura	Consulta	Variável/ Configurável			
Integração de dados		Consulta e Alteração	Segundo ocorrência de consultas/ pedidos	Garantido	Não
Integração de vistas de leitura e escrita	Transformação			Garantida após a ocorrência	Sim

Tabela 4 - Caracterização de abordagens de integração de informação

2.4 Estado da Arte

Esta secção descreve projectos e organizações cuja preocupação é a integração de sistemas legados ou a resolução de problemas que se encontram nesta tarefa. Estes projectos ou organizações tentam sistematizar a problemática da integração de sistemas e informação, apontando ou implementando soluções para facilitar e industrializar o processo.

2.4.1 TOSEM

O projecto TOSEM (Philippe Thiran et al. 2006) tem por objectivo criar uma metodologia de criação de componentes designados por *wrapper (adapter)* que terão a responsabilidade de interagir com o programa cliente de tecnologia recente, transmitindo-lhe a ilusão de que determinada base de dados legada garante a integridade relacional.

De facto, alguns sistemas legados incluem lógica de dados no código da aplicação. Cada programa onde fossem manipulados dados para os quais importava manter a integridade relacional, tinha essa lógica codificada repetidamente. Hoje em dia, esta lógica está tipicamente embutida no software de base dos servidores (e.g. a autenticação) e sistemas de gestão de base de dados (e.g. mecanismos de manutenção da integridade relacional). No que respeita à integridade relacional, as aplicações actuais abstraem-se dessas tarefas delegando-as nos SGBD, o que se mostra um problema quando integram com sistemas legados que não dispõem dessa facilidade.

Como ilustrado na Figura 10, a abordagem preconizada pelo projecto TOSEM sugere que a nova aplicação (usando tecnologia recente) comunique com o *wrapper* e não com o SGBD ou sistema de ficheiros legado. Neste contexto é tido que o *wrapper* é do tipo *W/R wrapper* (permite operações de leitura e escrita), numa materialização da abordagem de Integração de vistas de leitura e escrita apresentada na secção 2.3.6. No entanto, como se observa, são descritos dois tipos de *wrapper* classificados quanto ao sentido em que são colocados no modelo de integração;

- interface entre uma nova aplicação e uma BD legada;
- interface entre uma aplicação legada e um novo SGBD.

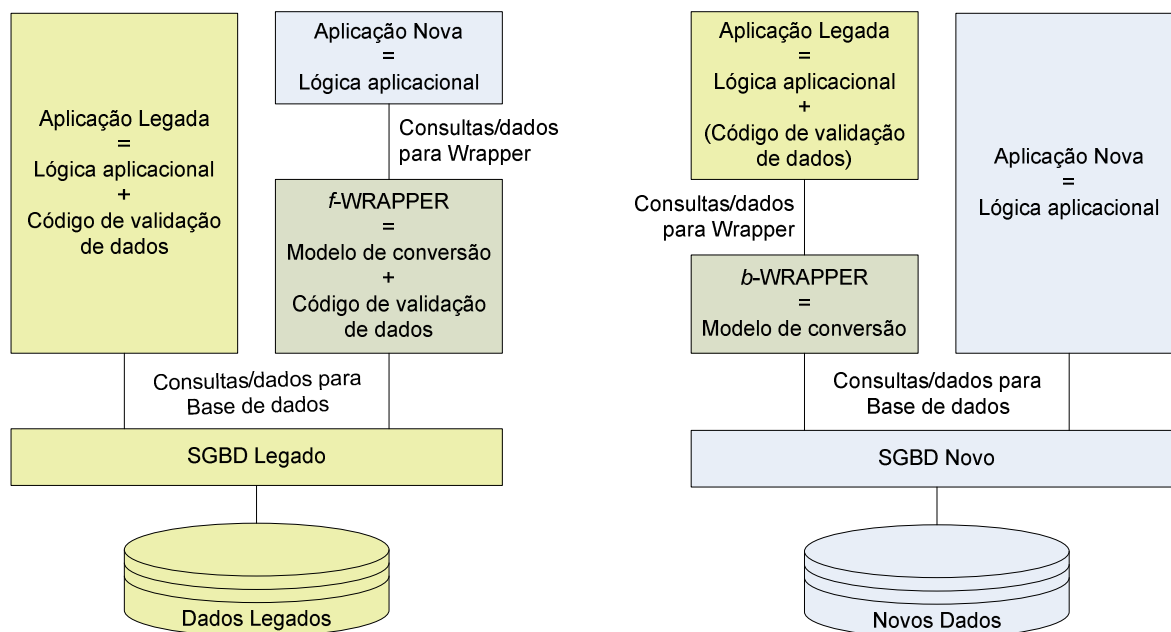


Figura 10 - *Wrappers* na interface entre sistemas legados e novos

A Figura 10 ilustra o papel dos *wrappers* na coexistência de sistemas legados e novos componentes, numa evolução de sistemas (esquerda) e migração de sistemas (direita).

Ou seja, considera dois tipos conceptuais de *wrappers*:

- *f-wrapper* (*forward-wrapper*), quando uma nova aplicação integra uma aplicação e BD legados. O *wrapper* é usado para ligar uma aplicação nova a um sistema de dados legado, encapsulando o código de validação, tendo em consideração que o SGBD legado não implementa validações como a integridade relacional. Os pedidos de leitura/escrita da nova aplicação são encaminhados para o *f-wrapper* que os interpreta e traduz em solicitações ao sistema legado com o objectivo de validar o pedido e de o executar;
- *b-wrapper* (*backward-wrapper*) quando uma aplicação legada é integrada numa aplicação e BD novas. O *wrapper* é usado para ligar a aplicação legada a um novo sistema de dados, dando-lhe a ilusão de que opera com o sistema de dados legado. O código de validação embutido na aplicação legada será redundante, já que o SGBD moderno deverá ele próprio implementar essa validação. Por essa razão, o *b-wrapper* será potencialmente mais simples que o *f-wrapper*.

2.4.2 OMIS

O projecto OMIS - An Organisational Memory Information System using Ontologies (Vasconcelos, Gouveia, e Kimble 2002) tem por objectivo a criação de uma *framework* dinâmica para a modelação e integração empresarial baseada em ontologias, na expectativa de definir um sistema num domínio, com a preocupação de gerir o conhecimento empresarial.

O conhecimento nas organizações é uma colecção de especializações, experiências e informações, retida por indivíduos e grupos de trabalho, de importância comparável ao capital, capacidade produtiva e recursos. Por outro lado muito do conhecimento está descrito de forma directa ou indirecta

nos processos, manuais ou aplicações existentes, assim como nas pessoas. A rotatividade de colaboradores nas organizações representa uma ameaça à retenção do conhecimento no seio da organização.

Assume-se que a conceptualização da organização transcende os sistemas de informação, e poderá ser descrita conceptualmente recorrendo a ontologia(s).

Conforme a Figura 11, a memória organizacional deve compreender o conhecimento gerado ao longo do tempo numa organização, e é apresentada como a base que suporta as tarefas com uso intensivo desse conhecimento organizacional.

Descrições formais de conhecimento representam todas as fontes de informação existentes na organização. Essas fontes incluem informação estruturada e não estruturada, descrições de processos, bases de conhecimento, multimédia, etc., que contemplam informação factual, processos de negócio, informação de negócio tradicionalmente em bases de dados, comunicações, conhecimentos e heurísticas detidas pelos empregados, conjunto de experiências e conhecimento pericial.

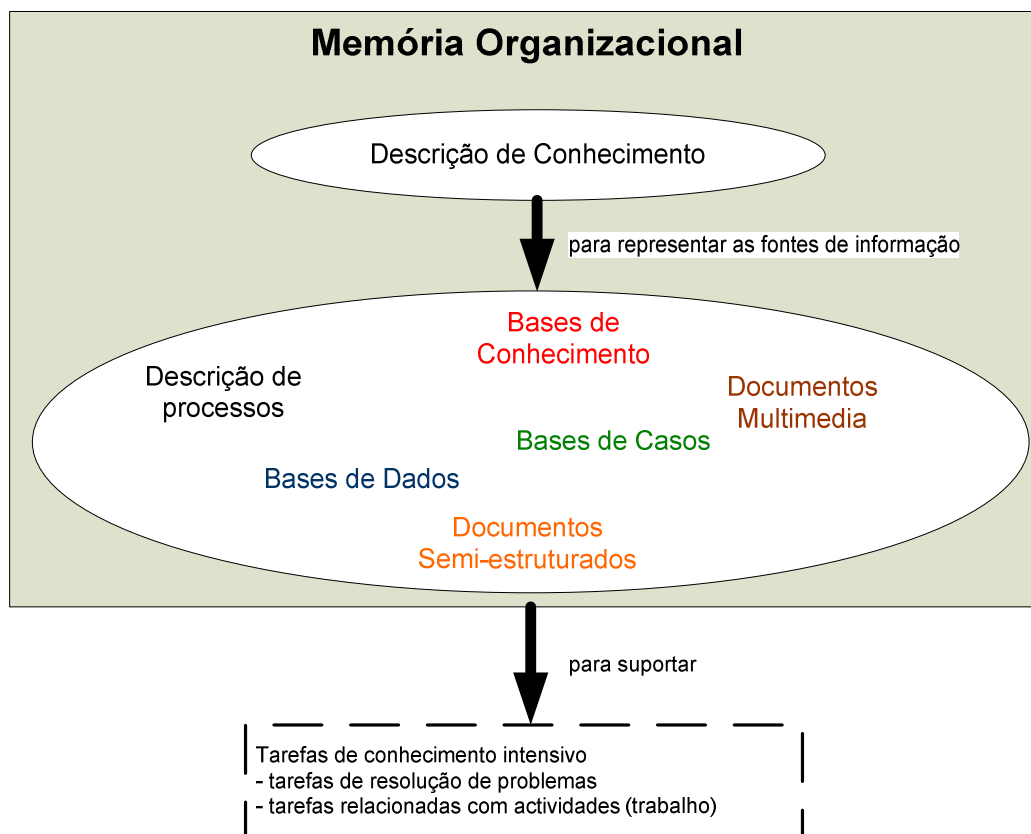


Figura 11 - Resumo de Memória Organizacional

É tido que este conhecimento é a base dos processos de negócio, do comportamento da organização e determinante para o crescimento da própria memória organizacional.

A integração deste conhecimento organizacional disperso, com as tarefas de negócio é designada de Memória Organizacional.

A Memória Organizacional apresenta-se neste projecto como uma evolução natural dos sistemas de informação organizacionais, enquanto informação tangível (e.g. informação administrativa, de

processo empresarial em base de dados) se integra com informação intangível (e.g. conhecimento pericial, competências). Formaliza, descreve e relaciona a experiência da organização no tempo com a informação tradicionalmente suportada nos sistemas de informação, o que se pode tornar num facilitador para integração entre sistemas num domínio.

Esta abordagem ao desenvolvimento de sistemas empresariais baseada em ontologias sustenta-se antes na modelação do modelo em vez da necessidade de aplicação. É uma vantagem uma vez que o modelo é mais estável no tempo e existe um foco mais concentrado no negócio e não na aplicação, e é independente dos pedidos dos utilizadores, em geral válidos no período em que o colaborador se mantém na organização.

2.4.3 IKF-IES

Ontology-driven Natural Language access to Legacy and Web services in the Insurance Domain (Mikhail Simonov, Aldo Gangemi, e Massimo Soroldoni 2004).

“As ontologias podem desempenhar um papel importante nos sistemas industriais permitindo o acesso a recursos legados de uma forma transparente e distribuída.”

Este projecto explora a utilização de ontologias num caso de estudo no domínio dos seguros, no sentido de disponibilizar informação, até então só acessível pelas aplicações do sistema legado, a sistemas CRM Web em linguagem natural.

Com base no projecto de investigação IKF¹, foi desenvolvido um sistema adaptado ao domínio dos seguros com a geração da ontologia no domínio designada de IES, que actualiza um novo modelo de interacção de negócio a partir de um sistema legado robusto e que responde, por si, satisfatoriamente às necessidades das operações do negócio, como é tradicional no domínio das seguradoras.

No sentido de integrar sistemas distintos, as ontologias descrevem os elementos de negócio, especificando restrições estruturais e semânticas, mas não o processamento (cálculo) e regras.

Portanto, as ontologias por si não descrevem a plenitude do sistema.

Uma estratégia para combinar processamento e regras com ontologias poderá basear-se na utilização de peças de software no sistema legado que, implementando as funcionalidades necessárias, disponibilizem os resultados do cálculo, realimentando portanto as bases de conhecimento associadas às ontologias. Para isso, os resultados são representados e tornados públicos em estruturas descritíveis em linguagens formais, como aquelas usadas na descrição de ontologias.

Nesta abordagem as ontologias são então utilizadas para descrever os serviços sobre o sistema legado (*legacy services*) que serão usados para responder às consultas colocadas no CRM do tipo “Qual o valor de rendimento do meu seguro de vida?” em cada sistema.

¹ IKF é um projecto internacional co-financiado EUREKA, com vista ao desenvolvimento de *frameworks* que permitam a gestão de conhecimento em várias áreas de negócio. Referenciado no site <http://www.ikfproject.com> actualmente indisponível.

O projecto aponta duas abordagens para o problema do cálculo e regras:

1 – pré-cálculo – O sistema já calculou e registou na BD o que poderá ser pesquisado, e responde com os valores calculados. O cálculo pode voltar a ser executado periodicamente.

2 – existência de serviços sobre o sistema legado que efectuem o cálculo on-line.

Nesta perspectiva também admite a concepção de um *wrapper* para a integração de sistemas legados. Assim, este projecto mostra-se como um passo importante na integração de sistemas que individualmente representam bem cada unidade de negócio, mas que até então estavam limitados na partilha de conhecimento do domínio entre eles. Esta é uma área em desenvolvimento com um valor económico significativo num domínio de negócio em que a interdependência de informações entre organizações revela um peso crescente nas operações das seguradoras.

2.4.4 TAO

O projecto europeu TAO - Transitioning Applications to Ontologies (Bontcheva 2007) visa a sistematização de processos de baixo custo para a transição de aplicações legadas para aplicações em ambientes distribuídos/partilhados.

Este projecto estuda um caso prático na indústria da manutenção aeronáutica, rico em elementos de conhecimento, como por exemplo caracterização hierárquica de peças e componentes; equivalências de materiais; procedimentos e planos de manutenção, e especificações dos fabricantes da cadeia aeronáutica.

Trata-se de uma indústria com forte tradição em sistemas de informação legados heterogéneos e com grande empenho na integração, uma vez que a necessidade de partilha de informação actualizada é fundamental para a continuidade do negócio da manutenção, assim como para a eficiência e eficácia das intervenções em manutenção.

Os custos envolvidos em projectos de integração são muitas vezes optimistas numa fase primária mas transformam-se em somas avultadas depois, o que em alguns casos são motivo suficiente para o abandono da integração. Outras vezes são proibitivos logo na primeira análise (Bontcheva 2008).

As recentes tecnologias e ambientes computacionais, como SOA e web semântica, abrem terreno ao desenvolvimento de plataformas de integração de sistemas num ambiente orçamental sustentado.

A transição de sistemas legados para ontologias é tomada como um desafio à engenharia de software semanticamente assistida.

No conjunto dos sistemas a integrar e no âmbito dos objectivos do projecto, foram identificados vários problemas típicos, de onde se destacam:

- linguagens de programação ou modelos em desuso;
- programas mal estruturados e difíceis de manter;
- sistemas mal documentados e incompreensíveis;
- difícil de integrar entre sistemas legados e novos;
- necessária migração para aplicações web 2.0 e serviços.

O projecto revela uma orientação para migração a baixo custo¹ tendo em vista (Bontcheva 2008):

- mudança rápida e efectiva para a utilização e aplicação de ontologias;
- construir um processo de transição reutilizável;
- minimizar o tempo de consulta e integração;
- reduzir atrasos na integração e limitar o risco.

O processo de transição é faseado, de onde se distinguem as fases de:

- aprendizagem semi-automática de ontologias no domínio a partir de software e conteúdos;
- acréscimo semântico dos conteúdos legados e definições de Serviços Web;
- repositórios semânticos heterogéneos e distribuídos.

2.4.5 Attachmate

A Attachmate Corporation² é uma empresa de software de origem inglesa que faz da integração de sistemas, nomeadamente de sistemas legados, o seu objecto de negócio. Contempla no seu site uma lista internacional de mais de quarenta casos de sucesso de integração de sistemas legados nas mais diversas áreas, e.g. financeira e banca, comunicações, distribuição e logística, aeronáutica e transportes³.

Consciente da solidez das soluções em produção nos sistemas dos seus clientes e da necessidade da adaptação a novos paradigmas organizacionais e à distribuição de aplicações e informação, faz integração de sistemas de informação baseados em mainframes, para a Web e novas gerações de aplicações (e.g. CRM, SCM), oferecendo soluções de emulação de terminal, comunicações seguras, e integração de sistemas legados incluindo *web enablement* e integração em novos paradigmas, e.g. componentes⁴.

Esta empresa preconiza portanto uma abordagem bastante diferente dos projectos descritos anteriormente, muito mais orientada para a integração de sistemas do que integração de informação.

2.5 Sumário

Este capítulo apresentou metodologias e abordagens sistematizadas à integração de informação e de sistemas sob diferentes perspectivas, enfatizando as características mais relevantes que as tornam preferíveis para os diferentes cenários de integração.

Foram referidos projectos de relevo no estudo e sistematização do processo de integração de sistemas legados, e.g. TAO e OMIS, e outros que se ocupam com a resolução de problemas localizados que

¹ <http://www.tao-project.eu/>

² <http://www.attachmate.co.uk>

³ <http://www.attachmate.co.uk/CustomerStories/BCP+Bank.htm>

⁴ <http://www.attachmate.co.uk/Products/Host+Connectivity/Host+Integration/Host+Integration.htm>

ocorrem quando se integram sistemas legados e se enfrentam dificuldades que as ontologias por si não resolvem, e.g. TOSEM e IKF-IES.

Foi descrita uma abordagem mais comercial da integração pela Attachmate, que realça a importância dos sistemas legados dos seus clientes.

3 Migração ad hoc

Neste capítulo descreve-se o processo de integração de informação no caso de estudo apresentado sumariamente no Capítulo 1. Trata-se dum processo ad hoc¹ de integração de parte da informação do sistema legado para o novo sistema SCM.

3.1 Introdução

No caso que aqui se descreve ficou desde muito cedo decidido que se adoptaria simultaneamente uma abordagem de migração de dados, migração de aplicações e também de integração dos dados e integração de aplicações, pois haveria necessidade de:

- manter em funcionamento diversas partes do sistema legado (Descrição técnica de malhas, Recursos humanos, Contabilidade);
- transferir funcionalidades até então em uso para o novo sistema SCM (Definição de artigos, Encomendas);
- utilizar novas potencialidades do novo SCM (Gestão orientada à encomenda, gestão da subcontratação e acompanhamento das requisições, seguimento de materiais).

Do ponto de vista do sistema legado, o processo de integração focou-se na migração de três tipos de dados:

- Terceiros;
- Malhas;
- Acessórios.

Este documento descreverá o processo de migração de Terceiros, não descrevendo os dois últimos na medida em que não trazem à discussão elementos conceptualmente novos à análise do problema.

Outras “áreas” de negócio não serão consideradas na migração de dados, uma vez que (i) esses dados não são tidos como necessários ou críticos na utilização do novo sistema SCM e (ii) não foi revelado interesse da organização no acesso a dados históricos no novo sistema. Serão ao invés alvo de migração da utilização de aplicações (Tabela 2):

- Artigos de vestuário;
- Encomendas;
- Compras;
- Facturação;

¹ A expressão ad hoc aplica-se para enfatizar a característica de se aplicar ao fim específico em causa, não tendo a presunção de ser aplicada a outros contextos ou genericamente. Pode-se inclusive compreendê-lo no contexto de engenharia de software em que é usado para “designar ciclos completos de construção de softwares que não foram devidamente projectados em razão da necessidade de atender a uma solicitação específica do utilizador, ligada a prazos, qualidade ou custo” [Wikipedia - <http://pt.wikipedia.org/wiki/Software>].

- Contabilidade;
- Stocks;
- Recursos humanos abrangendo salários e prémios de produção (não migra dados nem utilização de aplicação).

Foram várias as razões para limitar o processo às áreas anteriores, nomeadamente:

- Constrangimento de recursos humanos disponíveis, quer em termos de programadores, quer em termos de orçamento para a subcontratação dos serviços. De facto, estes constrangimentos não permitiriam escalonar a migração de todas áreas de negócio num prazo aceitável para ter o novo sistema SCM em operação, garantindo a operação sem quebras em nenhuma área;
- Artigos de Vestuário – Da sazonalidade dos artigos de vestuário criados por colecções de cliente, resulta que todos os artigos serão novos para o sistema SCM, pelo que a sua migração não é necessária para o sistema SCM. Além do mais pretende-se que o sistema legado continue disponível, permitindo o eventual acesso à informação não migrada;
- Encomendas, Compras e Stocks – No sistema SCM tratar-se-ão as novas encomendas, esgotando-se as transacções das encomendas em curso no sistema legado. Os registos de Compras e Stocks associados às encomendas são então gerados no sistema onde a encomenda seja criada/processada;
- Contabilidade, Facturação e Recursos Humanos – São áreas organizacionais estáveis que não motivam alterações, não estando previstos investimentos no curto prazo para a sua evolução;
- O sistema de contabilidade legado dá resposta às necessidades da organização, não se projectando investimentos nesta área no curto prazo, pelo que os movimentos de expedição gerados no SCM irão integrar em facturação e contabilidade no sistema legado;
- Mudança de paradigma organizacional – Alterações operacionais motivam a migração de um modelo assente no planeamento de produção para um de gestão de cadeia de fornecimento representados, respectivamente, nos sistemas legado e SCM, uma vez que:
 - o sistema legado é orientado às necessidades de materiais e produção de matérias intermédias dentro da organização suportando-se nas ordens de fabrico, planeamento de cargas e de recursos que já não existem na organização e apresenta uma relação frágil entre a compra de matérias-primas, produção de matérias intermédias e as encomendas de cliente;
 - o sistema SCM é orientado à especificação de produtos com base em componentes e operações numa perspectiva de subcontratação, mantém uma relação forte entre encomendas de cliente, requisição de materiais e de serviços, até à expedição da mercadoria, e permite o acompanhamento documentado da subcontratação, movimentos de fornecimento de materiais e prestação de serviços;

- este cenário atribui prioridade à migração das áreas Artigos de Vestuário, Encomendas e gestão da cadeia (incluindo requisições e movimentações de stocks de produtos intermédios) sobre as restantes áreas de aplicação disponibilizadas pelo novo sistema.

Nas secções seguintes descreve-se o processo conduzido na identificação, análise e sistematização da informação existente no sistema legado associada a Terceiros, e que deverá ser migrada para o sistema SCM. Ao mesmo tempo, é descrito o mapeamento entre a estrutura e semântica originais e a estrutura e semântica do novo sistema.

3.2 Terceiros

Um terceiro é uma entidade externa à organização com quem se estabelecem transacções. Essas transacções são materializadas em produtos ou serviços, dão origem a contrapartidas financeiras (pagamentos e recebimentos) e são suportadas por documentos de facturação e contabilísticos.

O conjunto dos terceiros inclui fornecedores e clientes.

Pretende-se migrar os dados de terceiros que tenham documentos associados (e.g. facturação) de data posterior ao ano de 2006.

3.2.1 Análise semântica comparativa

Concentrando-nos nos fornecedores e clientes, em alguns casos típicos nesta indústria de transformação, um cliente é também fornecedor de uma matéria-prima ou intermédia, e.g. tecidos, botões, etiquetas. O contexto da transacção ditará o papel desempenhado pelo terceiro nessa transacção. Por outro lado, um mesmo terceiro pode estar localizado em espaços geográficos distintos, e.g. localidades ou países, o que se reflectirá em pormenores de documentação de expedição e fiscal, como tratamento de impostos e declarações fiscais.

Também a natureza dos materiais transaccionados tem influência nas contas de contabilidade usadas nas transacções, e.g. artigos de vestuário, malhas, serviços, matérias-primas e equipamentos para imobilizado.

No sistema legado, a identificação de terceiros usa a semântica da estrutura hierárquica usual do plano de contas (POC¹) descrita no Anexo 1.

Desta sistematização, de acordo com a legislação e semântica aplicada na organização, resulta que um terceiro pode ter várias contas no plano de contas devido a uma das seguintes razões:

- Está implantado em vários países;
- Tem endereços de entrega ou de facturação distintos em mercados ou países distintos, e.g. mercado nacional, externo comunitário, externo não comunitário;

¹ Plano Oficial de Contas

- Tem endereços de entrega ou de facturação distintos no mesmo país e com o mesmo número fiscal;
- Tem endereços de destino noutras entidades, portanto com número fiscal distinto;
- É fornecedor de serviços e de materiais, obrigando por isso a ter contas de fornecedor no POC diferentes;
- É simultaneamente cliente e fornecedor, portanto terá contas diferentes na classe 2.

Deste facto resulta que são potencialmente encontrados no sistema legado vários registos de terceiros para a mesma entidade real, levando à ocorrência de redundância de alguns dados, e.g. endereços e número de contribuinte¹, uma vez que neste sistema o terceiro é identificado pela sua conta do plano de contas.

Portanto, o modelo implementado no sistema legado apresenta uma visão contabilística das organizações, para contabilistas, onde o enfoque se centra nas contas.

Por contraponto, no sistema SCM, um terceiro é uma entidade identificada de forma independente das contas ou endereços que lhe estão associadas. Um terceiro adquire uma identificação única com um prefixo que o classifica: CL/FN, Cliente e Fornecedor, respectivamente. Os endereços e contas do POC (juntamente com outras informações contabilísticas) residem em tabelas distintas, relacionadas com a entidade. Este modelo ajusta-se melhor à realidade observada na sociedade comercial actual, reflectindo a conceptualização de “empresa”, destinos de entrega de mercadorias e contas contabilísticas, bem como está focado nas transacções orientadas às empresas.

A Tabela 5 exemplifica algumas das diferenças existentes entre as duas conceptualizações no que respeita aos números das contas, numa perspectiva de migração automática de terceiro do sistema legado para o SCM.

	Entidade real	Conta do Plano de contas		
		Legado	SCM	
1 2 3	A	2111. 32 (nacional) 2112. 32 (internacional comunitário) 2113.58 (internacional extra-comum.)	CL32	2111. 32 2112. 32 2113. 32
4 5 6	B	2112.50 (endereço 1) 2112.51 (endereço 2) 2162. 54 (endereço 1) (cliente malhas)	CL54	2112. 54 (implica criação de 2 endereços distintos associadas a terceiros e não a contas) n/a 2162. 54
7 8	C	2112. 58 (nº único em classes distintas) 2112. 54 (erro. Colide com linha 4 a 6 para entidade B)	CL58	2112. 58 (58 não colide com linha 3 porque nessa linha não passou para identificação no SCM) n/a
9	D	2112. 33 (cliente comunitário)	CL33	2112. 33
10	D	2212. 76 (fornecedor comunitário)	FN76	2212. 76
11	E	2212. 33 (fornecedor comunitário)	FN33	2212. 33

¹ Também designado de NIF – Número de Identificação Fiscal

12	F	2213.33 (fornecedor extra-comunitário) (erro. Colide com linha 11 para entidade E)	FN34	2213.34
----	---	--	------	---------

Tabela 5 - Conceptualização nas contas POC de terceiros nos dois sistemas

Nestes exemplos o código de terceiro adotado para o sistema SCM é o número de terceiro de uma das contas do terceiro no sistema legado (e.g. no caso das linhas 1,2 e 3, utilizou-se o número 32). No entanto o número de terceiro no sistema SCM pode não ter qualquer correspondência com o número de terceiro no sistema legado, que é o caso ilustrado na linha 12.

Sistematizando o ensaio anterior, a Tabela 6 apresenta uma comparação de algumas características de modelação de terceiros do sistema legado e do SCM.

Características	Sistema Legado	Sistema SCM
Número de Tabelas	1 (Terceiros)	3 (Entidades, Contas e Endereços)
Registos por entidade	Vários. Tantas quantas as contas no plano de contas (e.g. são criadas novas contas para endereços alternativos que o terceiro detém)	1 por entidade/país na tabela de entidades
		1 por endereço na tabela de endereços
		1 por conta POC na tabela de contas
Chaves únicas	1 constituída pela conta POC	1 constituída por uma identificação sequencial independente dos restantes atributos e relações
Identificação de cliente	Contas 21	Prefixo CL concatenado no nº de entidade
Identificação de fornecedor	Contas 22, 261 ou 268	Prefixo FN concatenado no nº de entidade
Unificador das contas POC da mesma entidade	Não existe.	O número de identificação na classe é sempre o mesmo, variando as classes, e.g. 2112 876 e 2113 876 para o cliente CL876. A mesma entidade não tem duas contas na mesma classe, e 2212 876 e 2213 876 para o fornecedor FN876.
Complexidade Semântica	Complexo. É difícil relacionar os vários registos da mesma entidade, várias identificações dificilmente relacionáveis.	Simple. Uma entidade com um registo único e uma identificação única partilhada pelas tabelas relacionadas.
Sequência/lógica na chave	Os nºs de conta são atribuídos por ordem de “pedido” e não agrupando as várias contas de um terceiro.	ID é único por entidade. É replicado para as tabelas relacionadas.
Integridade referencial	Não	Sim

Tabela 6 - Comparação de modelação de Terceiros entre os sistemas Legado e SCM

Esta tabela sugere já algumas dificuldades em conseguir uma equivalência de dados entre os dois sistemas uma vez que é complexo identificar registos da mesma entidade no sistema legado.

Ainda a respeito do conceito de terceiros, consideram-se relevantes as questões relacionadas com dois conceitos associados:

- **Agrupador de cliente.** O conceito de agrupador de cliente do sistema legado é usado para agrupar os custos e vendas por cliente, a um nível superior ao dos registos de terceiros, pois, como visto, um cliente real pode estar representado por vários registos de clientes no sistema, sendo os vários registos do terceiro agrupados por este código numérico.

Vista sobre Terceiros					
CodAgrup	Conta	Num	Nome	Marca	Gestor
3	2113	82	BANANA	BANANA	02
3	2113	487	GPS INC.	BANANA	02
3	2113	486	BANANA, INC	BANANA	02
3	2113	123	BANANA, INC.	BANANA	02
15	2113	84	H & M HENNES & MAURITZ GBC AB	H&M	01
15	2113	85	H & M HENNES & MAURITZ GBC AB	H&M	01
21	2112	903	H & M HENNES & MAURITZ GBC AB	COS	01

Tabela 7 - Código Agrupador para um cliente no Sistema Legado

No sistema SCM, esta associação é natural por via da relação de agrupamento dos registos das encomendas (do cliente) com requisições e pagamentos (a fornecedores). Além disso, o sistema SCM permite associar várias entidades entre si, e.g. para os casos em que a mesma entidade física é cliente e fornecedor ou tem representação em países distintos, permitindo uma visualização consolidada das transacções. Assim, a relação entre fornecimentos e vendas está naturalmente criada para novos movimentos. Contudo, como referido anteriormente, os dados do sistema legado referentes a estas transacções não são migrados, nomeadamente por razões de custos, mantendo-se disponíveis no sistema legado, onde serão integradas as facturas geradas no sistema SCM;

- **Gestor de mercado.** O Gestor de mercado é o colaborador do departamento comercial responsável pelo acompanhamento dos clientes de determinado mercado junto da empresa. Este campo irá gerar uma ambiguidade em alguns clientes, pois no sistema legado o mesmo cliente poderá ter dois gestores de mercado distintos, e.g. gestor 01 para artigos de vestuário e gestor 90 para negócio de malhas, tendo o cliente contas POC diferentes por mercado. Por outro lado, no sistema SCM, cada cliente é uma única entidade, para a qual está associado um único gestor de mercado. Esta diferente conceptualização, provoca portanto um problema de integração que será avaliado e tratado no futuro se assim a organização o entender.

3.2.2 Mapeamento

Na Figura 12 representa-se em UML o esquema dos dois sistemas relacionados com Terceiros e diversas instâncias de cada um dos sistemas.

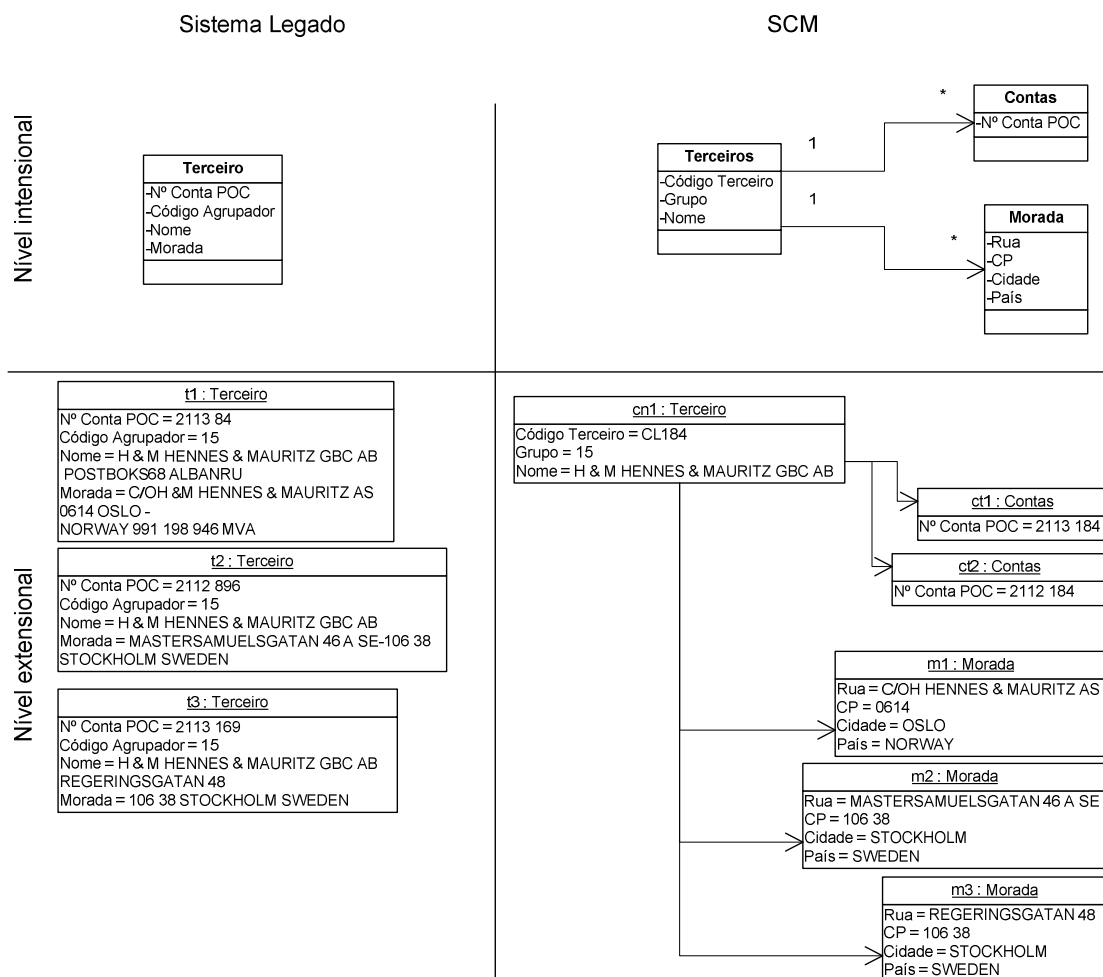


Figura 12 - Esquemas e instâncias equivalentes de Terceiros nos sistemas Legado e SCM

Como advém da análise semântica comparativa, no sistema SCM um terceiro apresenta um único Código de Terceiro, na tabela Entidades, que se relaciona com as tabelas de contas POC e endereços (moradas).

A migração destes dados implica:

- a identificação ou descoberta dos vários registos do sistema legado respeitantes à mesma entidade Terceiro;
- a selecção de um número para a identificação do terceiro – para um mesmo Terceiro, vários registos do sistema legado darão origem a um único registo de Entidade no sistema SCM;
- a selecção de um dos nomes encontrados no sistema legado para figurar na entidade – para um mesmo Terceiro, vários nomes podem ser usados;
- a concatenação de dois atributos que compõem o nome no sistema legado num só para o SCM;
- manipulação dos campos de endereço do sistema legado para o SCM;
- mapeamento da codificação de países, diferente nos dois sistemas;

- a criação dos registos no ficheiro de contas, ignorando as contas criadas por efeito de endereços de envio distintas;
- a criação dos registos das diferentes moradas;
- o registo das contas do plano do sistema legado em cada entidade inserida no sistema SCM para garantir rastreio com o sistema legado, já que a informação contabilística continua a ser gerada no sistema legado.

Desta enumeração ressaltam processos de manipulação de atributos, mapeamento e critérios na escolha de valores alternativos.

Na Figura 13 representa-se um conjunto de equivalências, nos níveis intensional e extensional, que ilustram as principais transformações e mapeamentos a operar nesta migração.

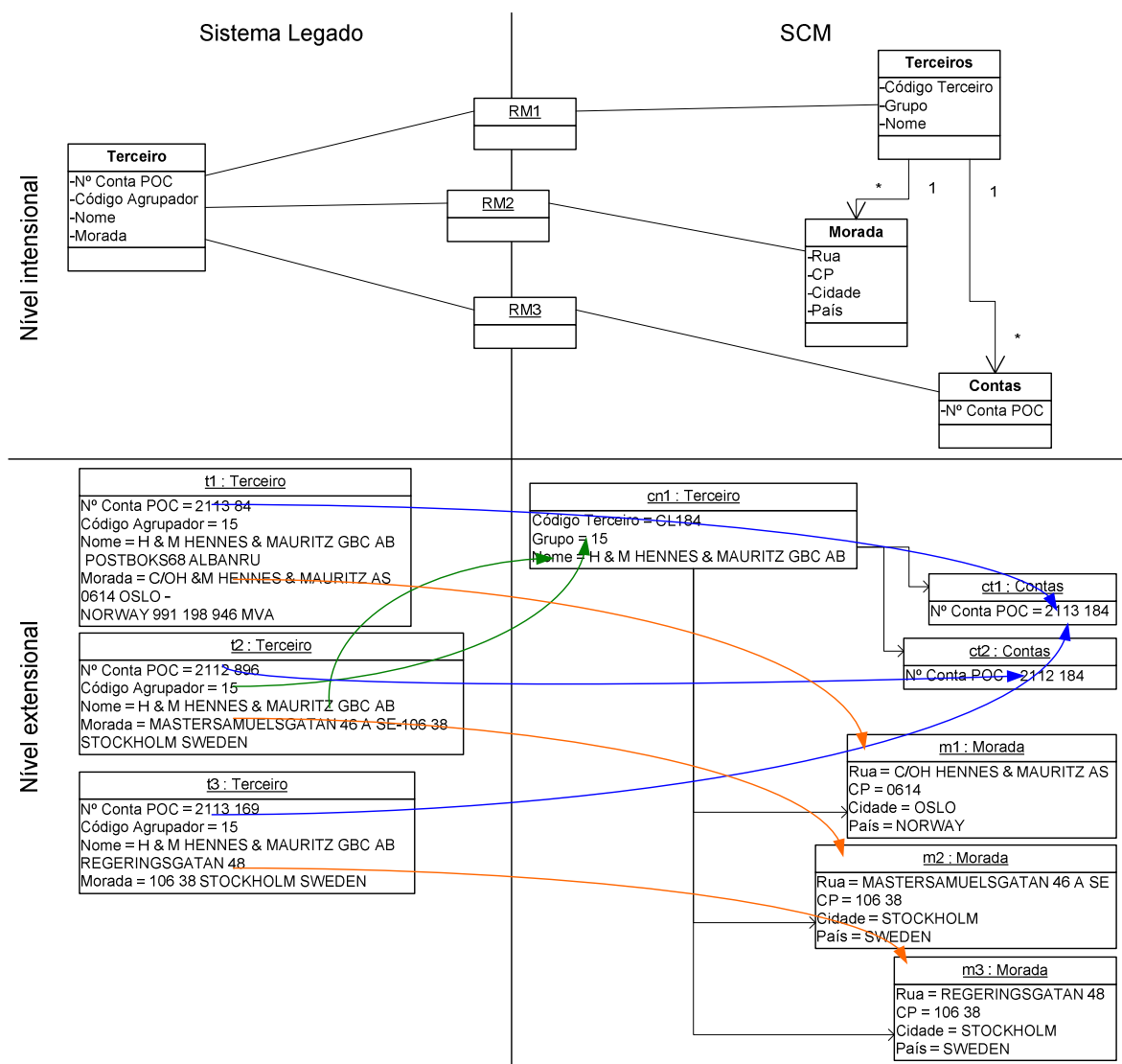


Figura 13 - Associação ou Mapeamento de Terceiros do Sistema Legado para o SCM

No nível intensional descreve-se o mapeamento composto por regras de mapeamento, e.g. RM1; RM2; RM3, que fazem corresponder registos e conjuntos de atributos da tabela de terceiros do sistema legado a registos, conjuntos de atributos e tabelas do sistema SCM.

No nível extensional, várias instâncias de terceiro do sistema legado serão mapeadas para uma instância de terceiro no sistema SCM. De facto, embora várias instâncias de terceiro dêem origem a uma única instância de terceiro no SCM, e.g. t2 em cn1, os atributos de t1 e t3 deverão também ser considerados e instanciados noutras tabelas (i.e. Moradas e Contas).

Na Tabela 8 evidenciam-se estes mapeamentos de instâncias de uma forma matricial, tendo como origem as instâncias do sistema legado, e como objecto as instâncias do sistema SCM

Instâncias Legado	Instâncias SCM		
	Terceiro	Conta	Morada
t2	cn1	ct2	m2
t1		ct1	m1
t3			m3

Tabela 8 - Mapeamento de terceiros do Sistema Legado para o SCM

3.2.3 Abordagem tecnológica

Sob um ponto de vista prático, o processo de migração de dados foi dividido em duas fases distintas:

- Automático, para acesso aos dados no sistema legado, envolvendo a criação de scripts SQL embutidas em folhas de cálculo Excel;
- Manual, constituído pela visualização e manipulação dos dados nas folhas de cálculo produzidas na fase automática.

A Figura 14 representa um cronograma global do caso em estudo a curto e médio prazo, que ilustra a migração de dados e a migração da utilização de aplicações.

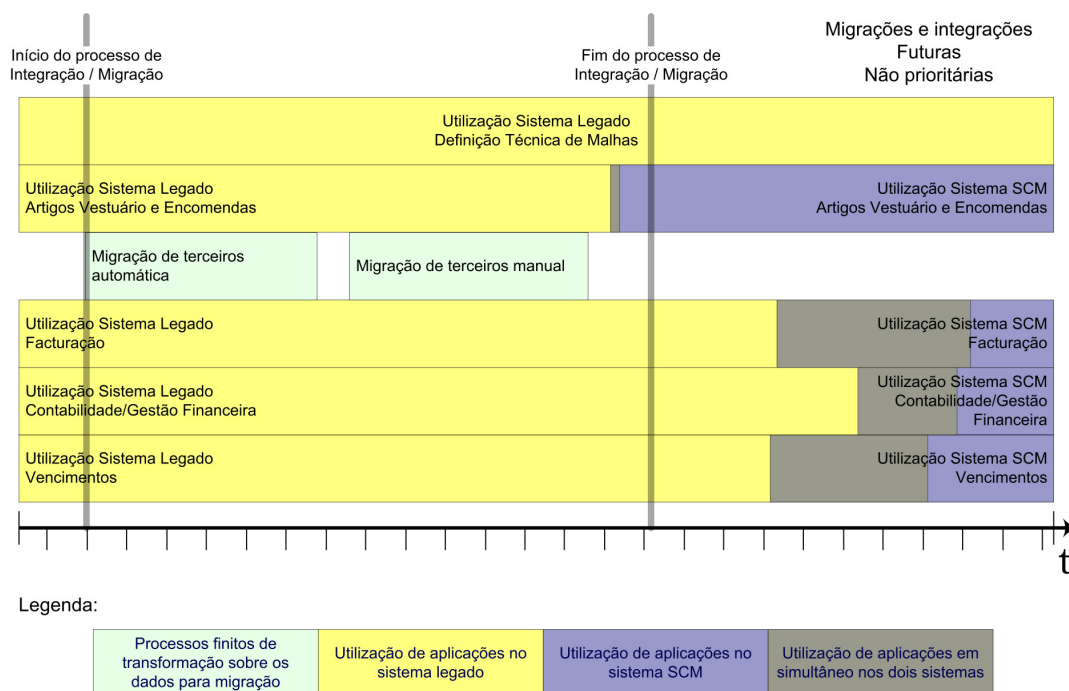


Figura 14 - Cronograma para migração/integração

As barras horizontais representam processos de migração ou utilização de áreas aplicacionais nos sistemas legado e SCM. Em particular as barras verdes representam a execução dos processos de migração de dados. As amarelas representam a utilização das aplicações do sistema legado, e as azuis a utilização das aplicações no sistema SCM. As azuis só surgem depois das amarelas, e têm início após a conclusão das verdes, portanto só se dá início à utilização das aplicações do sistema SCM depois dos dados terem sido migrados. As zonas a cinzento representam áreas aplicacionais em utilização nos dois sistemas simultaneamente.

Da figura realça-se que não está prevista a utilização da definição técnica de malhas no sistema SCM, porque não é tão detalhada quanto no sistema legado, pelo que se prevê a sua utilização no sistema legado e integração no SCM.

Os processos de migração de terceiros automático e manual determinam o início da utilização de encomendas de vestuário no sistema SCM, após a validação da migração.

3.2.3.1 Automático

Tendo por objectivo a migração de dados do sistema legado ERP para o novo sistema SCM, este processo foi conduzido em parceria entre a organização, que extraiu os dados do sistema legado, e a empresa fornecedora do novo SCM, que os transformou e inseriu na base de dados da nova instalação. A migração de dados do sistema ERP actual para o novo sistema ERP/SCM foi idealizada em dois passos:

1. Usando scripts de SQL para extrair dados do ERP para folha de cálculo tradicional (e.g. Excel). Neste processo são usados filtros que seleccionam a informação mais relevante, tipicamente por critérios de actualidade, e.g. clientes activos com movimentos recentes. São seleccionados todos os atributos das tabelas. Esta folha pode sofrer refinamentos como concatenação ou extracção de *strings*, codificações e outros processos de tradução ou complemento de dados. Este passo é executado pelo departamento de informática da organização;
2. Estes dados são posteriormente importados para o SGBD do sistema SCM e integrados na nova base de dados. Esta tarefa é executada pelo fornecedor do software SCM.

Os scripts foram embutidos em folhas de cálculo Excel através da ferramenta “importar dados externos” e acedem ao sistema legado via driver ODBC.

Teve-se a preocupação de criar dois scripts específicos, um para clientes e outro para fornecedores, e criar critérios de selecção de dados, relacionando a tabela de terceiros com as tabelas de movimentos, no sentido de:

- Não importar conjuntamente clientes e fornecedores, com vista a simplificar a manipulação manual destas duas classes de terceiros;
- Só importar terceiros que tiveram movimentos de 2006 até à actualidade, relacionando os terceiros ora com facturação, ora com registo de créditos na contabilidade, para clientes e para fornecedores respectivamente, evitando a importação de dados não relevantes.

No ficheiro de terceiros do sistema legado existem 11.566 registos de terceiros. Com a aplicação destes critérios obteve-se uma redução em 89%, para um subconjunto de 1.280 registos.

O resultado é uma tabela cujas linhas se referem a um Terceiro no sistema legado, e as colunas correspondem às propriedades do Terceiro.

Em função do mapeamento referido na secção anterior, codificaram-se os scripts que se descrevem de seguida, onde se relacionam os ficheiros de terceiros, com os de movimentos de facturação e registo de contabilidade para obter os dados relevantes.

Para obter Clientes:

```
select * from libcom21.FORCLIL009
where (substring(fc1480,1,2)='21'
or substring(fc1480,1,3)='268' )
and concat(fc1480,concat(char(fc1482),concat(char(fc1484),char(fc1486)))) in (
select concat(mfc020,concat(char(mfc030),concat(char(mfc031),char(mfc032)))) from
libcom21.MOVDIC09 where mfc170>'2005'
);
```

Para obter Fornecedores:

```
select * from libcom21.FORCLIL009
where (substring(fc1480,1,2)='22'
or substring(fc1480,1,3)='261' )
and concat(fc1480,concat(char(fc1482),concat(char(fc1484),char(fc1486)))) in (
select concat(rgf350,concat(char(rgf352),concat(char(rgf354),char(rgf356)))) from
libcom21.REGFAC09 where rgf020>2005
);
```

Os dados obtidos por estes scripts são posteriormente formatados pelo integrador usando folhas de cálculo, que após manipulação, irá importar para as tabelas de bases de dados do sistema SCM.

O resultado deste processo, para o caso do cliente “H & M”, é o conjunto de registos migrados do sistema legado para o sistema SCM representado na Tabela 9:

Código	Nome	Endereço	Localidade	País	NomeAbrev	Cidade	Telef	Ncontr
CL0163	H & M HENNES & MAURITZ GBC AB POSTBOKS68 ALBANRU	C/OH &M HENNES & MAURITZ AS 0614 OSLO – NORWAY		NOR	2113.84	991 198 946 MVA	MARC A: H&M	.
CL0164	H & M HENNES & MAURITZ GBC AB S – STOCKHOLM	C/H&M HENNES & MAURITZ SA INDUSTRIESTRASSE 16		CH	2113.85	CH-4623 NEUENDORF- SWITZERLAND	MARC A: H&M	.
CL0171	H & M HENNES & MAURITZ GBC AB REGERINGSGATAN 48	106 38 STOCKHOLM		SE	2.113.169	SWEDEN	MARC A:H&M	.
CL0180	HENNES & MAURITZ (SHANGHAI) COMMERCIAL LTD.	B1-3F, NO 645-659 MIDDLE HUAI HAI ROAD, LU WAN DISTRICT		CHIN	2.113.500	SHANGHAI - P.R. CHINA	MARC A H&M	.
CL0272	H & M HENNES & MAURITZ GBC AB	REGERINGSGATAN 48 106 38 STOCKHOLM		GB	2112.896/211 2.903	SWEDEN	MARC A: COS	894525582

Tabela 9 - Registos de um terceiro migrados para o SCM

No processo de transformação, o integrador optou por inserir no campo “NomeAbrev” o número de identificação do terceiro usado no sistema legado, como memória da transformação do sistema legado para o SCM.

Uma análise visual rápida às folhas de cálculo produzidas mostrou que existem bastantes problemas nos dados, e.g. escrita diferente de nomes para a mesma entidade (e.g. “JOSE ADELIO E OFELIA

INDUSTRIA DE CONFECÇÕES, LDA. “ e “JOSE, ADELIO & OFELIA INDUSTRIA DE CONFECÇÕES, LDA”; “POSTO BP SANTAREM NASCENTE GEST24 SOC.UNIP.LDA “ e ” A.S. BP MEALHADA-CANTANHEDE CARDOL LDA.”) e informações inseridas em campos incorrectos (e.g. Cidade = ”SWEDEN”).

Além disso, note-se no exemplo ilustrado na Tabela 9, que no atributo “Nome” é apresentada parte do endereço (e.g. Código CL0163 – Nome: “H & M HENNES & MAURITZ GBC AB” Endereço: “POSTBOKS68 ALBANRU”; Código CL0171 – Nome: “H & M HENNES & MAURITZ GBC AB” Endereço: “REGERINGSGATAN 48”), que advém de uma utilização abusiva dos campos destinados ao “Nome” para conter partes do endereço.

Do processo semi-automático baseado nos scripts anteriores e folhas de cálculo, ficam por resolver os seguintes problemas de mapeamento:

- a selecção de um número de identificação para a identificação do único terceiro a inserir no sistema SCM;
- a selecção de um dos nomes encontrados no sistema legado, pela ambiguidade resultante da escrita do mesmo nome em registos diferentes e pela não existência de um critério para a selecção;
- identificação clara do país – a descrição do país surge no campo cidade, e o código do país é o do país destino da mercadoria, e.g. CL0272 com código de país GB=Reino Unido, mas morada na Suécia. Esta ocorrência revela que foi facturado ao cliente com instalações na Suécia uma mercadoria que teve como destino o Reino Unido;
- a criação desnecessária de registos na tabela de contas, ignorando a origem das contas criadas por efeito de moradas de envio distintas decorrentes da dificuldade em identificar claramente os registos respeitantes à mesma entidade no mesmo país;
- a criação dos registos das diferentes moradas não foi eficaz, pela dificuldade em identificar as mesmas moradas, já que foram escritas com diferente ortografia, e.g. Tabela 10:

Conta	Nome	Morada	Localidade
2113 149	HI – TEX	QUARTIER INDUSTRIEL AL MASSAR	LOT N° 915 – MA
2213 377	HI-TEX S.A.R.L.	Q.I AL MASSAR N.° 915	MARRAKECH MAROC

Tabela 10 - Registos do mesmo Terceiro no Sistema Legado com grafia diferente

- ao longo do tempo os utilizadores foram preenchendo de forma inadequada os campos de alguns registos do sistema legado, em especial:
 - os que compõem o nome e a morada, e.g. nalguns casos colocando no atributo “Cidade “ o país, embora exista um atributo para país;
 - preencheram o “Nome” com parte da morada;
 - no “Telefone” colocaram a indicação de uma marca.

Estes casos só se tornaram visíveis por inspeção e manipulação das folhas de cálculo produzidas. Esta fraca qualidade dos dados impossibilitou uma migração mais automatizada, tornando necessária uma fase manual, não sistemática e baseada no utilizador, de transformação dos dados na folha de cálculo.

3.2.3.2 Manual

A fase manual baseou-se na inspeção visual e manipulação da folha de cálculo produzida, inserindo-se colunas novas, copiando conteúdos, apagando linhas, e criando outras folhas de cálculo auxiliares de acordo com o mapeamento referido anteriormente.

De forma a facilitar o processamento de transformação de dados, foram criadas novas colunas:

- “Numero” o número de identificação de terceiro no sistema legado.
- “TNumero” o “Numero” anterior foi concatenado com a sigla “CL” ou “FN” e inserido na coluna “TNumero”, conforme se trate de um terceiro Cliente ou Fornecedor. Este valor será o código de terceiro no sistema SCM, excepto se entrar em conflito com um outro terceiro, procedendo-se conforme ilustrado na Tabela 5;
- “Conta_Geral” a Conta do POC desse terceiro no sistema legado, que será mantida no sistema SCM;
- “Conta_Geral_1” uma segunda Conta do POC do mesmo terceiro, conforme casos observados na Tabela 5, portanto diferente de “Conta_Geral”. Será mais uma conta POC associada ao mesmo terceiro no sistema SCM;
- “Numero_1” segundo número de identificação do terceiro no sistema legado a figurar no sistema SCM para, em conjunto com “Conta_Geral_1”, possibilitar o mapeamento com os terceiros do sistema legado.

Este modelo prevê inicialmente a existência de até dois registos de terceiro no sistema legado para um registo de terceiro no sistema SCM (suportado por dois grupos de colunas “Numero...” e “Conta_Geral...”), mas serão acrescentados mais conjuntos de colunas (n) se surgirem casos de equivalência (n:1).

Ao preencher numa linha as colunas “Numero_1” e “Conta_Geral_1” indica que foi identificada uma segunda linha do sistema legado para o mesmo terceiro. O “Numero” e “Conta_Geral” dessa linha são copiados para “Numero_1” e “Conta_Geral_1” da primeira linha, e a segunda linha é eliminada. Foram inicialmente copiados todos os números existentes no sistema legado, e corrigidos para o número único nos casos em que a entidade no sistema SCM é mapeada para mais que um registo no sistema legado.

Neste processo manual mapearam-se alguns casos, como CL001 e CL0896 (Tabela 11), o que contemplou manipulações como as descritas a seguir:

- copiaram-se células dos endereços (Moradas), designações sociais (Nome) e outras, entre as várias linhas para o mesmo terceiro, seleccionando-se as mais correctamente escritas;

- eliminaram-se linhas, copiando-se previamente para as colunas “Numero_1” e “Conta_Geral_1” da linhas mantidas do mesmo terceiro, as colunas “Numero” e “Conta_Geral” da linha eliminada, portanto a conta POC no sistema legado do segundo registo do terceiro;
- procedeu-se à limpeza de caracteres especiais no conteúdo de alguns campos.

Na Tabela 11 mostra-se um excerto da folha de cálculo na fase de correcção.

Original

Tipo	TNumero	Numero	Conta_Geral	Conta_Geral_1	Numero_1	Obs	Nome	Morada	Cidade	CPOSTAL	N_CNT	Telefone	Cond_Venda	Pais
CL	CL0001	1	2162	2112	793	128	BENIND SPA	VILLA MINELLI	ITALIA		02142210265		CIF	IT
CL	CL0793	793	2112			128	BENIND SPA	VILLA MINELLI	ITALY		02142210265		CIF	IT
CL	CL0896	896	2112			21	H & M HENNES & MAURITZ GBC AB	REGERINGSGATAN 48	SWEDEN		894525582	MARCA: COS	FCA	GB
CL	CL0903	903	2112			387	H & M HENNES & MAURITZ GBC AB	REGERINGSGATAN 48	SWEDEN		N0302078A	MARCA: H&M	FCA	ES
CL	CL0084	84	2113			387	POSTBOKS68 ALBANRU	C/OH & M HENNES & MAURITZ AS	991 198 946 MVA			MARCA: H&M	FCA	NOR
CL	CL0169	169	2113			387	H & M HENNES & MAURITZ GBC AB	REGERINGSGATAN 48	106 38 STOCKHOLM			MARCA:H&M	FCA	SE
CL	CL0085	85	2113			387	H & M HENNES & MAURITZ GBC AB S	- STOCKHOLM	C/H&M HENNES & MAURITZ SA	CH-4623 NEUENDORF-SWITZERLAND		MARCA: H&M	FCA	CH

Novo

Tipo	TNumero	Numero	Conta_Geral	Conta_Geral_1	Numero_1	Obs	Nome	Morada	Cidade	CPOSTAL	N_CNT	Telefone	Cond_Venda	Pais
CL	CL0001	1	2162	2112	793	128	BENIND SPA	VILLA MINELLI	ITALIA		02142210265		CIF	IT
CL	CL0896	896	2112	2112	903	21	H & M HENNES & MAURITZ GBC AB	REGERINGSGATAN 48	SWEDEN		894525582	MARCA: COS	FCA	GB
CL	CL0084	84	2113			387	POSTBOKS68 ALBANRU	C/OH & M HENNES & MAURITZ AS	991 198 946 MVA			MARCA: H&M	FCA	NOR
CL	CL0169	169	2113			387	H & M HENNES & MAURITZ GBC AB	REGERINGSGATAN 48	106 38 STOCKHOLM			MARCA:H&M	FCA	SE
CL	CL0085	85	2113			387	H & M HENNES & MAURITZ GBC AB S	- STOCKHOLM	C/H&M HENNES & MAURITZ SA	CH-4623 NEUENDORF-SWITZERLAND		MARCA: H&M	FCA	CH

Tabela 11 - Processo manual de mapeamento de Terceiros

Esta fase não se mostrou eficaz, tendo sido abandonada com muito poucos registos mapeados correctamente (ca. De 1280 registo de terceiros foram migrados 45 para o sistema SCM com mapeamento de mais de uma conta no sistema legado, faltando identificar os restantes casos semelhantes). Os motivos para este insucesso prendem-se com:

- quantidade elevada de registos a manipular (1.280 terceiros);
- o tempo necessário para o tratamento de cada registo (entre 2 minutos nos casos mais óbvios, e 14 minutos nos casos mais complexos em que existem várias contas para o mesmo terceiro);
- a complexidade da identificação de terceiros do sistema legado, já que se pretendia usar um dos números de terceiro do sistema legado no sistema SCM e em alguns casos isso não foi possível;
- escassez de meios humanos para a execução das tarefas manuais.

Contudo, da lista de dificuldades enumeradas na fase automática descrita em 3.2.3.1 ficaram conceptualmente resolvidas as seguintes questões:

- a selecção de um número de identificação para o terceiro a inserir no sistema SCM, por um processo manual, para os casos mais óbvios. Casos de clientes mais complexos, como “H & M” com mais de duas contas no sistema legado, revelaram-se de tratamento difícil e bastante demorado;
- a selecção de um dos nomes encontrados no sistema legado;
- correcção manual da descrição do país e cidade;
- correcção do conteúdo inadequado de alguns campos, como os que compõem o nome, morada e telefone.

Não ficaram de todo resolvidas:

- a criação de registos de endereços para as contas criadas por efeito de moradas de envio distintas;
- a identificação de marcas e atribuição nos campos reservados à marca no sistema SCM.

3.3 Conclusão

Neste processo foi possível migrar correctamente um conjunto amplo de terceiros que não apresentam multiplicidade de registos no sistema legado, mas que por força dos casos irregulares, terão também de ser manipulados pelo operador.

A utilização de folhas de cálculo revela-se interessante para a manipulação humana dos registos, e.g. classificação, agrupamento, transformação, filtragem. Com alguma facilidade procuram-se os registos de terceiros conhecidos ou tidos como problemáticos.

Contudo, a experiência da migração de terceiros descrita, com mapeamento de baixa qualidade do sistema legado para o SCM, foi suficiente para demonstrar que os processos manuais de tratamento e

manipulação de dados sobre folhas de cálculo não produzem uma transformação com satisfação mínima nos casos em que existe alguma complexidade no mapeamento/transformação.

Torna-se necessário encontrar uma solução suportada por software que se mostre como solução económica do problema de mapeamento.

4 Migração declarativa

Neste capítulo descreve-se a solução desenvolvida para endereçar os problemas descritos no capítulo anterior, mas desta feita aplicando uma abordagem tecnológica da engenharia do conhecimento, declarativa, baseada numa plataforma de migração de dados por mapeamento de ontologias. Estas abordagens são genéricas, declarativas e com poucas opções de configuração.

4.1 Migração de dados baseado em Ontologias

O Mapeamento de Ontologias é o processo pelo qual são definidas as relações semânticas, no plano ontológico, entre entidades de uma ontologia origem (*source*) e de uma ontologia destino (*target*); servindo como descrição do processo a aplicar às instâncias da ontologia origem para as transformar em instâncias da ontologia destino. A Figura 15 expressa essa perspectiva (Silva 2004:5).

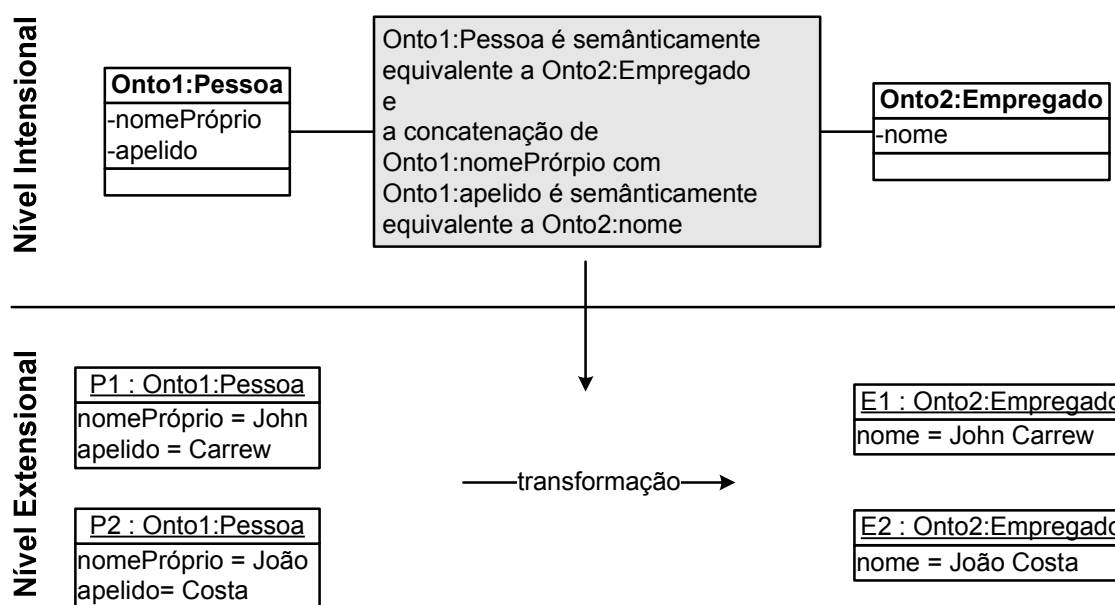


Figura 15 - Representação informal de Migração Baseada em Ontologias

Com a Migração Baseada em Ontologias não se pretende fundir ontologias nem os dados instanciados, mas antes transformar instâncias, de acordo com relações semânticas (relações de mapeamento) definidas no nível conceptual.

Assim, os repositórios mantêm-se autónomos e heterogéneos, mantendo a sua semântica e conteúdos inalterados.

Formalmente, a Migração Baseada em Ontologias é também descrita como um processo em duas fases (Silva 2004). A primeira, designada por fase de especificação de pontes semânticas (*semantic bridging*), é formalmente definida como uma relação entre as entidades das ontologias origem e destino:

$$M = E^o \times E^d$$

M é a especificação formal do mapeamento de ontologia, ou simplesmente mapeamento de ontologia, contendo a informação suficiente e requerida para transformar, aquando da fase de execução, as instâncias da ontologia origem nas instâncias da ontologia destino.

O objectivo desta fase é a especificação e representação de M de acordo com a semântica das ontologias origem e destino, complementada com conhecimento pericial do domínio.

A especificação de mapeamento de ontologias é um processo ao nível meta no sentido em que os objectos manipulados representam intensionalmente o domínio dos elementos do nível de instâncias, ou extensional.

A segunda fase, designada por fase de execução das pontes semânticas, é formalmente definida como a relação entre as bases de conhecimento origem e destino, parametrizada segundo o documento de mapeamento produzido ao nível intensional (M):

$$T(M) \subseteq I^o \times I^d$$

A fase de execução do mapeamento de ontologias ocorre ao nível das instâncias, respeitando a especificação de mapeamento (M).

Além disso, o mapeamento de ontologias é mais complexo e implica mais do que esses processos. MAFRA – MApping FRAMework é uma descrição conceptual que oferece uma visão geral do processo global de mapeamento de ontologias. A ferramenta separa claramente os requisitos operacionais dos requisitos complementares de operação. Está organizada segundo duas dimensões (Figura 16):

- A dimensão horizontal está relacionada com os requisitos operacionais, representando o processo de mapeamento de ontologia;
- A dimensão vertical está relacionada com os requisitos complementares de operação, representando e fornecendo funcionalidades complementares ao longo de todo o processo de mapeamento, mesmo no caso de não pertencer ao processo central ou nuclear.

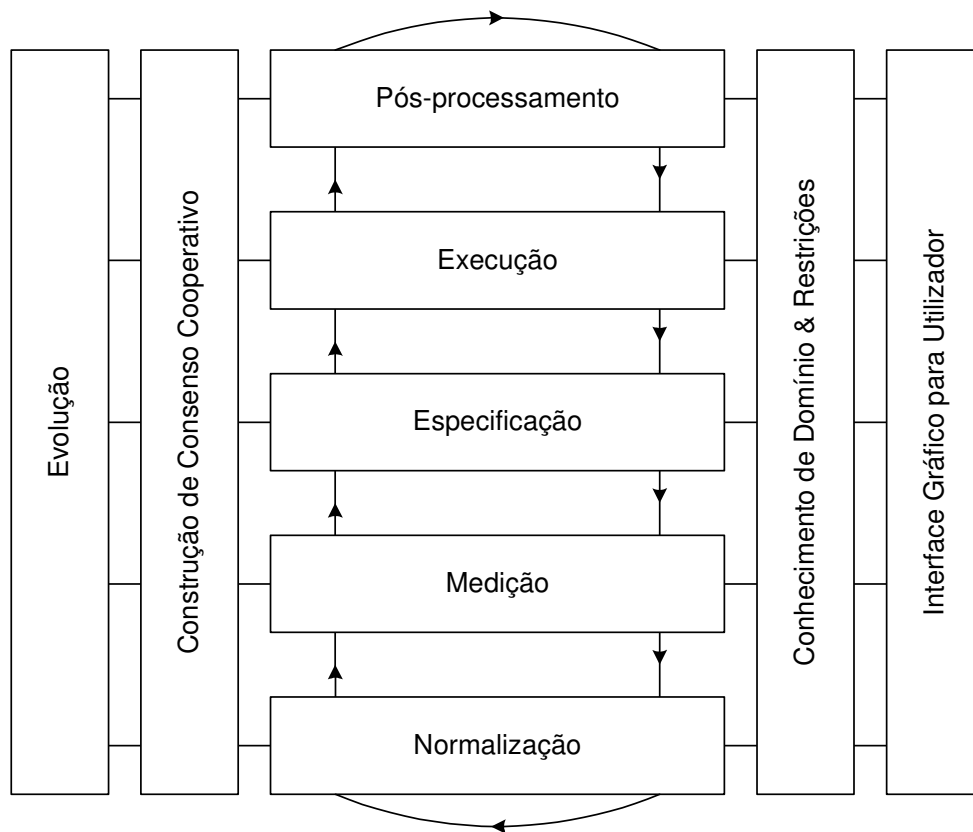


Figura 16 - MAFRA - MApping FRamework

Dos vários componentes que compõem o MAFRA, são de fundamental importância para este trabalho os seguintes:

- Especificação, que corresponde às competências que um sistema de migração de informação deve ter para representar o mapeamento entre as entidades do repositório origem e as entidades do repositório destino;
- Execução, correspondente à competência de transformação dos dados do repositório original em dados do repositório de destino, de acordo com o mapeamento definido;
- Pós-processamento, correspondente à competência do sistema em garantir que os dados migrados obedecem à semântica do repositório destino.

4.2 Semantic Bridging Ontology

A competência de Medição de Semelhanças, ou Similaridade, é uma fase nuclear do MAFRA, responsável pela geração de semelhanças entre as entidades das ontologias origem e destino. Estas semelhanças são, então, exploradas na fase de Especificação para construção das pontes semânticas. A ontologia Semantic Bridging Ontology (SBO) descreve os mapeamentos no domínio do conhecimento. Explorando as características conceituais e pragmáticas do conceito de ontologia, a SBO foi desenvolvida considerando dois propósitos:

- Capturar e descrever o conhecimento associado ao mapeamento de ontologias;
- Servir como um artefacto de representação de relações semânticas em cenários específicos.

Uma instanciação de SBO é um documento de mapeamento de ontologias.

A Ponte Semântica é o conceito fundamental sugerido pela SBO para relacionar entidades das duas ontologias. Este conceito não contém, por si próprio, capacidades de transformação nem é capaz de definir o processo necessário para transformar entidades de ontologias origem em entidades de ontologia destino. Como tal, é necessário associar a cada Ponte Semântica um serviço específico para estas funcionalidades (ou papéis) (Figura 17).

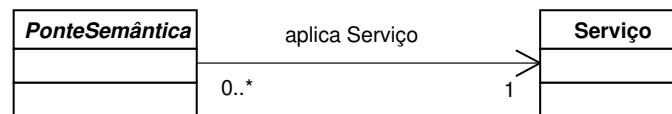


Figura 17 - Representação UML de relação conceptual entre Ponte Semântica e Serviço

A semântica e a assinatura de cada Serviço são definidas e descritas em consonância com os seus argumentos de entrada e de saída e respectivas características (e.g. conceito, atributo ou relação).

Uma PonteSemântica permite a especificação de condições. I.e. condições que devem ser respeitadas para que a PonteSemântica represente uma equivalência entre as entidades mapeadas. Só se as condições se verificarem é que a PonteSemântica é executada no nível extensional.

Os modelos típicos de representação de ontologias distinguem entidades de conceito e de propriedade (atributo ou relação). Seguindo a mesma abordagem, classe PonteSemântica é especializada em PonteConceito e PontePropriedades (Figura 18).

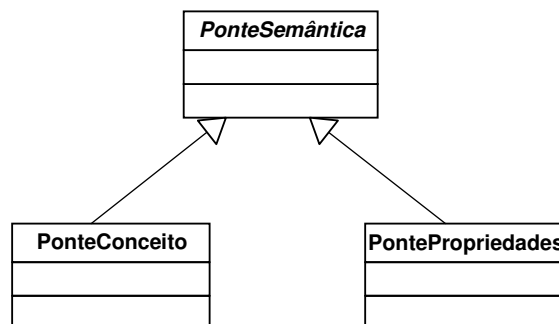


Figura 18 - Representação UML da taxonomia SBO de pontes semânticas

4.2.1 Ponte de Conceito (*Concept Bridge*)

A classe PonteConceito representa a relação semântica entre os conceitos das ontologias origem e destino. Semanticamente esta ponte significa que cada instância de um conceito origem dá origem a

uma nova instância do conceito destino. PonteConceito é uma especialização do conceito PonteSemântica em duas características:

- A cardinalidade de uma PonteConceito é sempre 1:1, o que significa uma relação semântica unívoca, i.e. um conceito origem está semanticamente relacionado com exactamente um conceito destino. No sentido de relacionar semanticamente um conceito origem com vários conceitos destino distintos, deverão ser descritas várias pontes semânticas;
- O processo de transformação executado na transformação de instâncias conceptuais origem em instâncias conceptuais destino é sempre o mesmo, e corresponde ao Serviço CopiaInstância (CopyInstance Service). Como consequência, a especificação do Serviço de transformação na PonteConceito é implícita.

4.2.2 Ponte de Propriedades (*Property Bridge*)

A classe PontePropriedades representa a relação semântica entre os conjuntos de propriedades das ontologias origem e destino. Semanticamente, esta ponte significa que o conjunto das instâncias de propriedades é transformado num conjunto de instâncias de propriedades destino. Ao contrário da PonteConceito, a transformação que ocorre entre as propriedades origem e destino apresenta uma variação acentuada. Como consequência, o serviço associado tem que ser mencionado de forma explícita na PonteSemântica. O Serviço associado é portanto implicitamente responsável por várias características da ponte, em particular:

- O tipo das Entidades relacionadas através da ponte. Um Serviço de concatenação, por exemplo, requer atributos do tipo String de ambas as ontologias origem e destino, por outro lado um Serviço que crie relações entre instâncias destino requer relações como entidades mapeadas;
- A cardinalidade. Retomando o exemplo, o Serviço de concatenação apresenta uma cardinalidade n:1 (várias strings serão concatenadas numa única), enquanto que o Serviço split apresenta uma cardinalidade 1:n.

Contrariamente ao que se verifica na PonteConceito, que aplica directamente conceitos das ontologias como seus argumentos, as PontePropriedades aplicam propriedades das ontologias no contexto de um conceito de domínio (*domain*) e considerando o seu contra-domínio (*range*).

4.2.3 Relação entre PonteConceito e PontePropriedades

A PonteSemântica é especializada nas classes PonteConceito e PontePropriedades. A diferença mais importante nestas duas classes reside no tipo de Serviço que cada uma delas utiliza. Enquanto o Serviço usado em PonteConceito é sempre responsável pela criação da instância de conceito destino, o Serviço associado a PontePropriedades varia de acordo com as relações semânticas entre as propriedades das ontologias.

Esta distinção é muito importante para a caracterização das suas inter-relações. De facto, note-se que as PonteConceito são responsáveis pela criação de instâncias dos conceitos da ontologia destino, que por seu turno funcionarão como contentores ou repositórios para instâncias de propriedade da ontologia destino, resultantes da execução de PontePropriedades. Isto significa que instâncias de conceito são compostas por instâncias de propriedades (“José” é o valor da propriedade “Nome” no conceito “Pessoa”).

Não faria muito sentido criar instâncias de conceitos se estes não tiverem valores de atributos ou relações com outras instâncias, nem seria possível transformar instâncias de propriedades sem existirem as instâncias de conceitos que desempenham o papel de contentores de valores de propriedades.

Como consequência, as PonteConceito e as PontePropriedades estão intimamente inter-relacionadas e interdependentes. A relação hasBridge representa essa inter-relação.

Durante a fase de Especificação, esta relação significa que as propriedades do conceito origem correspondem às propriedades do conceito destino de acordo com a PontePropriedades. Em tempo de execução a PontePropriedades relacionada com a função hasBridge será executada para todas as instâncias de propriedades definidas nas instâncias de conceito origem usado para a PonteConceito, dando origem a instâncias da propriedade para a instância do conceito de destino definido na PonteConceito.

4.3 Pós-processamento

A fase de Pós-processamento tem por objectivo verificar e incrementar a qualidade das instâncias destino resultantes da fase de execução. Esta competência é necessária porquanto o mapeamento não permite (ou é demasiado complicado de) especificar a semântica completa do mapeamento entre as duas ontologias.

A verificação é executada tendo em conta três questões:

- A ontologia destino, em especial porque a especificação de pontes semânticas pode não respeitar por completo a (semântica da) ontologia destino. Consideremos um cenário onde a cardinalidade de O1:Individual.name não é declarada, enquanto O2:Person.has_name está limitada a 1. Imaginemos que uma ponte semântica é declarada de forma que os valores de O1:Individual.name são copiados para O2:Person.has_name, que não especifica limitações ou restrições acerca da cardinalidade da transformação. Assim, as instâncias:

- O1: Individual(i_1), name(i_1 , “José”), name(i_1 , “Joaquim”)

serão transformadas em:

- O2:Person(i_1), has_name(i_1 , “José”), has_name(i_1 , “Joaquim”)

o que é claramente uma inconsistência.

De facto, as pontes semânticas podem ser especificadas de forma insuficiente ou inconsistente com a ontologia destino, originando instâncias de ontologia inválidas;

- A base de conhecimento destino. Um dos erros mais comuns que ocorre no nível extensional está relacionado com a identidade de objecto (*object-identity*). Object-identity diz respeito ao reconhecimento de que na base de conhecimento destino:
 - Duas (ou mais) instâncias distintas representam o mesmo objecto do mundo real. Esta situação ocorre por uma das seguintes razões:
 - A mesma instância existe em ambas as bases de conhecimento, origem e destino;
 - As pontes semânticas estão sub-especificadas permitindo a criação de objectos falsamente distintos;
 - Duas (ou mais) entidades similares representam dois objectos do mundo real diferentes. Esta situação ocorre porque:
 - Estas duas (ou mais) instâncias similares falsas já existiam na base de conhecimento origem;
 - As pontes semânticas estão sub-especificadas permitindo a criação de objectos falsamente similares.
 - As pontes semânticas. Pode acontecer que as pontes semânticas obtenham erros que são detectados mas não resolvidos na fase de execução. Essas excepções podem não evidenciar erros de identidade de objecto ou ontológicos, mas em contrapartida algumas instâncias poderão ser transformadas de modo diferente do esperado, i.e. não serem convenientemente transformadas.

O trabalho conceptual de Guarino e Welty (Nicola Guarino, Massimiliano Carrara, e Pierdaniele Giarretta 1994) em meta-modelação, na qual identidade, unidade, rigidez, e dependência de primitivas de meta-modelação são estudadas, fornece uma boa base conceptual para esta problemática.

Complementarmente, o extenso trabalho em Verificação e Validação (*Verification and Validation - V&V*) de bases de conhecimento e sistemas de bases de dados (Frans Coenen, Barry Eaglestone, e Mick J. Ridley 2001) é um bom ponto de partida para uma solução pragmática destes problemas.

4.4 Solução desenvolvida

A solução desenvolvida é baseada no MAFRA e suportada pela ferramenta MAFRA Toolkit, e compreende os seguintes passos:

1. Tradução dos esquemas das bases de dados originais para a linguagem de representação de ontologias RDFS, que serve de representação interna comum no MAFRA Toolkit;
2. Definição do mapeamento entre as duas ontologias, usando a SBO para o efeito;
3. Definição das regras de pós-processamento de acordo com a ontologia POST (Martins 2007);
4. Tradução dos dados do sistema legado para RDF (representação interna comum de instâncias);
5. Transformação dos dados do sistema legado em dados do sistema SCM;

6. Execução das regras de pós-processamento sobre os dados gerados, garantido assim que obedecem à semântica do sistema SCM;
7. Tradução dos dados anteriores para a representação interna no sistema SCM.

As próximas secções descreverão apenas os passos 1, 2 e 3 com detalhe, pois são estes os passos que requerem intervenção e competência de perito. As restantes são executadas pela ferramenta mediante os documentos especificados nos passos 1, 2 e 3.

4.4.1 De esquema para ontologia

As ontologias representativas dos esquemas dos dois sistemas foram desenvolvidas de acordo com as regras sugeridas em (Silva 2004:Anexo 1). As ontologias desenvolvidas correspondem basicamente aos diagramas UML apresentados na secção 3.2.2, mas agora representadas em linguagem RDFS. Apresenta-se de seguida um pequeno extracto da ontologia representativa de Terceiros no sistema SCM:

```
<rdfs:Class rdf:ID="Terceiro"/>
<rdfs:Class rdf:ID="Morada"/>
<rdfs:Class rdf:ID="Conta"/>

<rdf:Property rdf:ID="CodTerceiro">
  <rdfs:domain rdf:resource="#Terceiro"/>
  <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>

<rdf:Property rdf:ID="Grupo">
  <rdfs:domain rdf:resource="#Terceiro"/>
  <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>

<rdf:Property rdf:ID="Nome">
  <rdfs:domain rdf:resource="#Terceiro"/>
  <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>

<rdf:Property rdf:ID="temConta">
  <rdfs:domain rdf:resource="#Terceiro"/>
  <rdfs:range rdf:resource="#Conta"/>
</rdf:Property>

<rdf:Property rdf:ID="temMorada">
  <rdfs:domain rdf:resource="#Terceiro"/>
  <rdfs:range rdf:resource="#Morada"/>
</rdf:Property>
```

Tabela 12 - Extracto da Ontologia do sistema SCM em RDFS

Para simplificação da explicação e foco nos aspectos importantes, a descrição que se segue considera apenas parte das ontologias desenvolvidas. Essas ontologias encontram-se representadas e anotadas na Figura 19, extraída da representação que o MAFRA Toolkit faz das ontologias para serem manipuladas pelo perito.

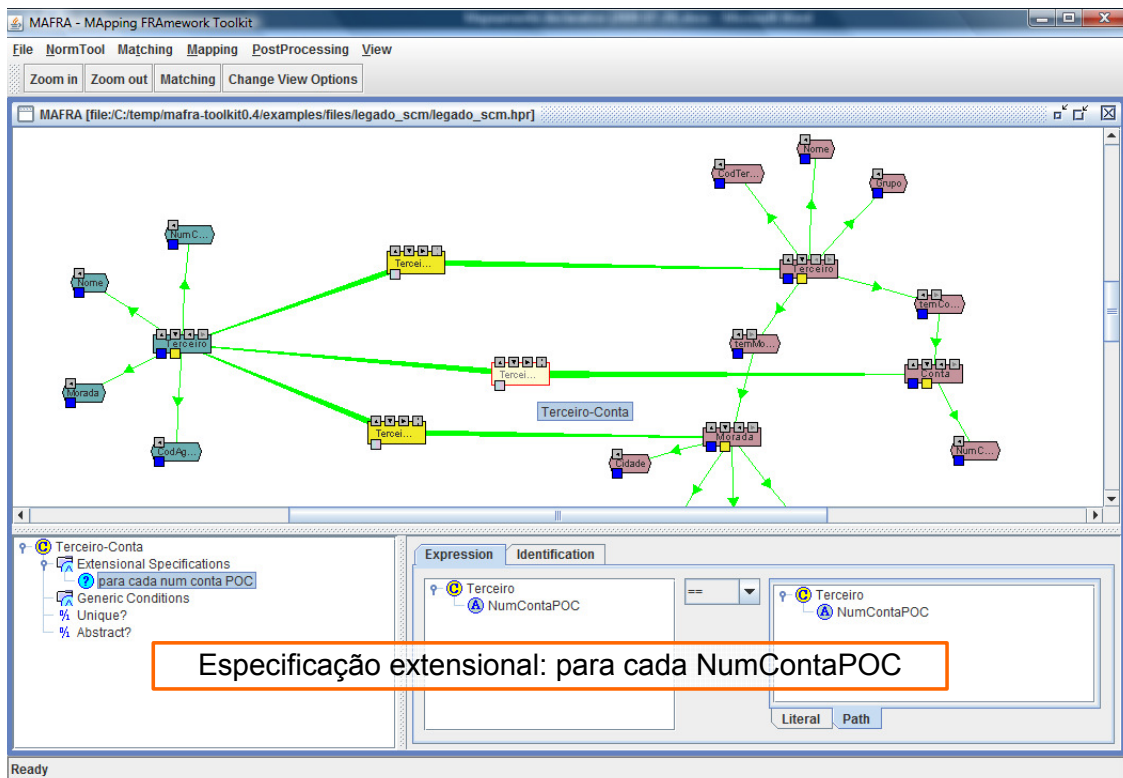


Figura 20 - ConceptBridges e especificação extensional

Com estas pontes semânticas é possível transformar as linhas da tabela Terceiros em linhas das tabelas Terceiro, Contas e Morada. Um exemplo dessa transformação apresenta-se de seguida, representando na Tabela 13 as instâncias de partida representadas na Ontologia origem, e na Tabela 14 as instâncias geradas na Ontologia destino:

```

<a:Terceiro rdf:ID="t1">
  <a:NumContaPOC>2113 184</a:NumContaPOC>
  <a:CodAgrupador>15</a:CodAgrupador>
  <a:Nome>HM Albanru</a:Nome>
  <a:Morada>Oslo Norway</a:Morada>
</a:Terceiro>
<a:Terceiro rdf:ID="t2">
  <a:NumContaPOC>2112 896</a:NumContaPOC>
  <a:CodAgrupador>15</a:CodAgrupador>
  <a:Nome>HM Mauritz</a:Nome>
  <a:Morada>Stockholm Sweden</a:Morada>
</a:Terceiro>
<a:Terceiro rdf:ID="t3">
  <a:NumContaPOC>2113 169</a:NumContaPOC>
  <a:CodAgrupador>15</a:CodAgrupador>
  <a:Nome>HM Mauritz</a:Nome>
  <a:Morada>106 38 Stockholm Sweden</a:Morada>

```

```

</a:Terceiro>
<a:Terceiro rdf:ID="t4">
  <a:NumContaPOC>2113 145</a:NumContaPOC>
  <a:CodAgrupador>10</a:CodAgrupador>
  <a:Nome>GAP</a:Nome>
  <a:Morada>34 Rodeo Drv Miami (FL), USA</a:Morada>
</a:Terceiro>

```

Tabela 13 - Instâncias da Ontologia de Terceiros origem

A Tabela 14 apresenta os resultados da execução da transformação mediante o mapeamento definido.

```

<a:Terceiro rdf:ID="i-1244498656069-1982437671"/>
<a:Terceiro rdf:ID="i-1244498656070-1161468662"/>
<a:Terceiro rdf:ID="i-1244498656070-193488316"/>
<a:Terceiro rdf:ID="i-1249296751472-1423191485"/>

<a:Morada rdf:ID="i-1244498656070-1958211802"/>
<a:Morada rdf:ID="i-1244498656066-392993274"/>
<a:Morada rdf:ID="i-1244498656069-1251661446"/>
<a:Morada rdf:ID="i-1249296751472-729275891"/>

<a:Conta rdf:ID="i-1244498656069-557097771"/>
<a:Conta rdf:ID="i-1244498656070-677547342"/>
<a:Conta rdf:ID="i-1244498656069-1930764581"/>
<a:Conta rdf:ID="i-1249296751472-103536258"/>

```

Tabela 14 - Excerto da transformação de dados da ontologia origem para instâncias de destino

Como se constata, o mapeamento definido tem ainda muitas falhas, nomeadamente porque nenhum atributo ou relação foi transformado.

É portanto necessário definir as pontes semânticas de propriedades. Este processo é extremamente dependente das relações semânticas e das funções de transformação (serviços). Das experiências descritas em (Silva 2004), é habitual que cada cenário de mapeamento necessite de novos serviços. Este cenário não fugiu à regra, tendo sido necessário desenvolver um novo serviço capaz de lidar com as diferenças semânticas entre os dois repositórios.

Para isso foi desenvolvido um novo serviço denominado POC2Terceiro, capaz de lidar com as relações do conjunto composto por NumContaPOC e CodAgrupador do sistema legado, e o conjunto constituído pelos atributos CodTerceiro e NumContaPOC do sistema SCM.

Para isso faz uso duma tabela de transformação que mapeia durante o processo de transformação, os dados do sistema legado para o sistema SCM. A Figura 21 representa o processo que o serviço executa e o papel da tabela auxiliar no processo.

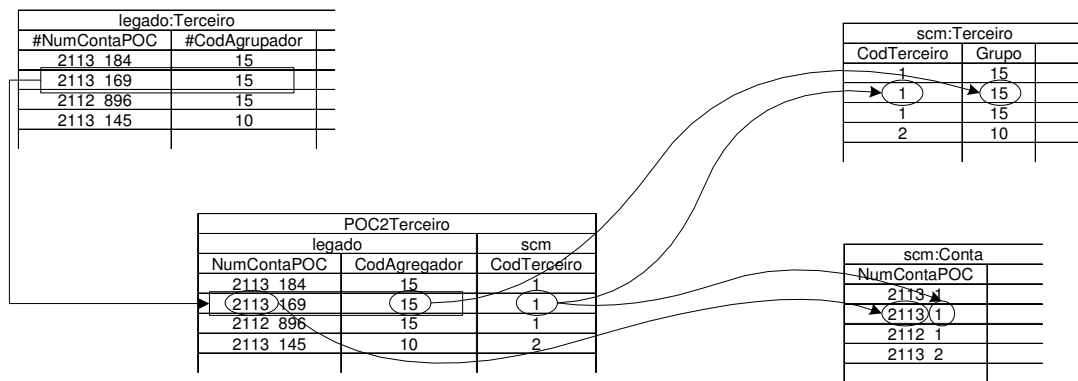


Figura 21 - Representação das transformações dos identificadores do Sistema Legado para o SCM

O valor de CodTerceiro é gerado sequencialmente por ordem de transformação. Em termos práticos, e uma vez sistematizadas as relações semânticas, a implementação é relativamente simples e rápida, pois o MAFRA Toolkit disponibiliza uma API específica para o efeito.

Quer a tabela (i.e. POC2Terceiro) quer o valor de CodTerceiro são únicos, pelo que a sua implementação é baseada nos padrões GoF Singleton (única instância) e Factory (criação e gestão de uma única instância duma classe).

Em termos de especificação das relações semânticas, trata-se de especificar pontes semânticas entre propriedades aplicando o serviço POC2Terceiro. Como descrito na Figura 21, os atributos NumContaPOC e CodAgrupador são os atributos do repositório de origem a serem processados e transformados nos atributos do repositório SCM CodTerceiro, Grupo e NumContaPOC. Como estes atributos são de dois conceitos diferentes, é necessário aplicar duas pontes semânticas distintas:

- POC+CodAgrupador-CodTerceiro, cujo serviço é POC2Terceiro, definida como PontePropriedades de Terceiro-Terceiro. Esta PontePropriedades transforma o valor de NumContaPOC e CodAgrupador da tabela Terceiro do sistema legado em CodTerceiro da tabela Terceiro do sistema SCM;
- POC+CodAgrupador-NumContaPOC, cujo serviço é também POC2Terceiro, definida como PontePropriedades de Terceiro-Conta. Esta PontePropriedades transforma o valor de NumContaPOC e CodAgrupador da tabela Terceiro do sistema legado em NumContaPOC da tabela Conta do sistema SCM.

Estas PontePropriedades são definidas com o MAFRA Toolkit como se apresenta na Figura 22.

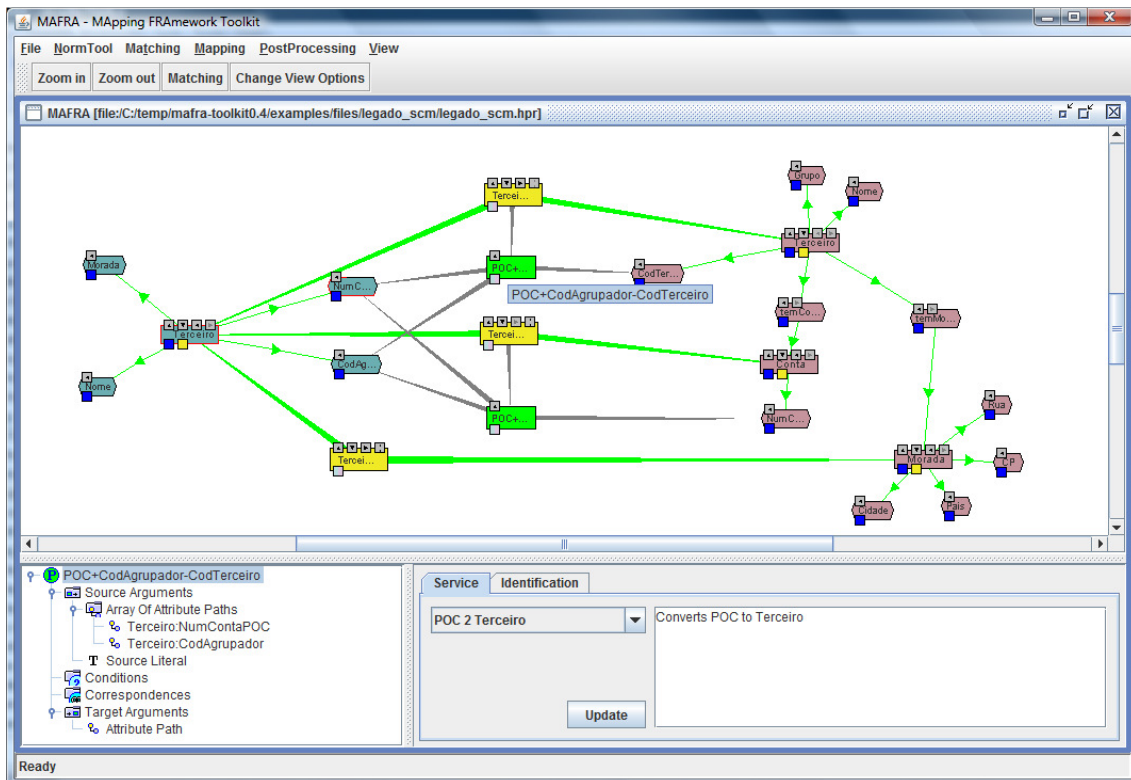


Figura 22 - Pontes de propriedades de Terceiro

Finalmente, é necessário criar as relações (chave estrangeira na terminologia do modelo relacional) entre as instâncias de Terceiro e Conta, bem como entre as instâncias de Terceiro e Morada.

Para isso usa-se o serviço de criação de relações existente de base no MAFRA Toolkit (CopyRelation). Para além do uso de especificação extensional (detalhadamente descrita em (Silva 2004), o seu uso não levanta qualquer tipo de problema neste cenário. O resultado em termos de especificação no MAFRA Toolkit é representado na Figura 23 (as restantes PontePropriedades foram escondidas para simplificação da figura).

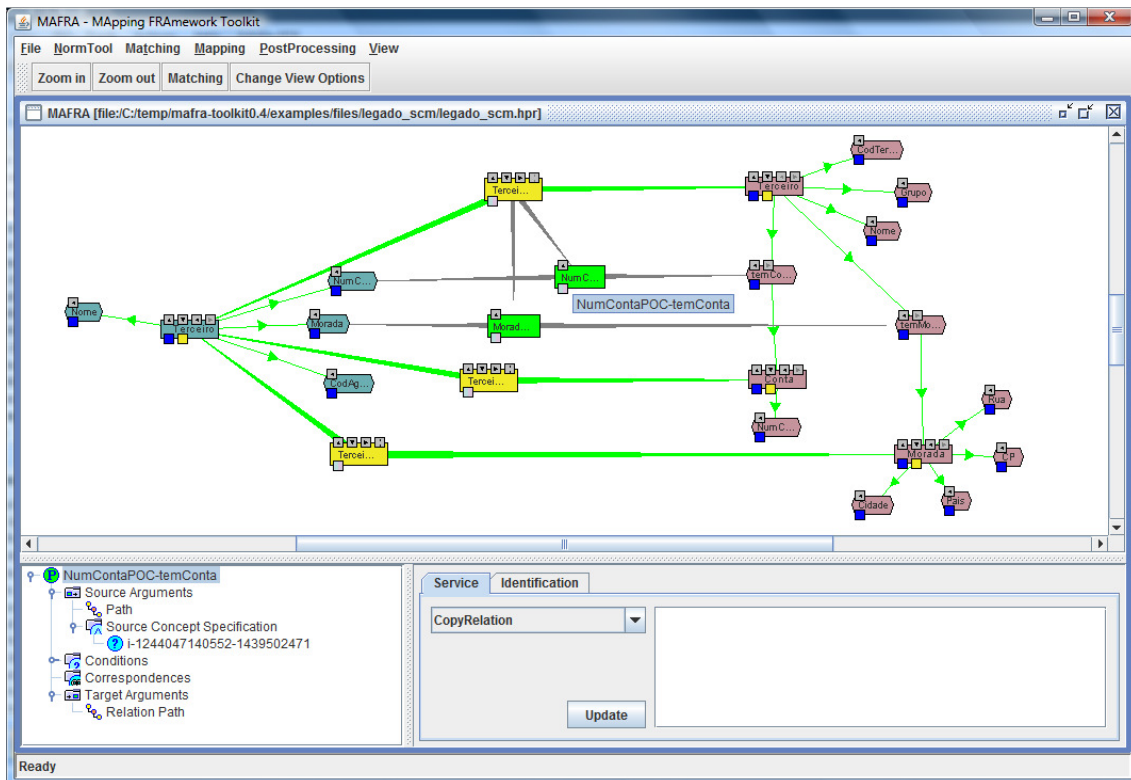


Figura 23 - PontePropriedades com Serviço CopyRelation

Algumas outras PontePropriedades foram especificadas para outros atributos, mas essas transformações são evidentes sob o ponto de vista semântico, e algo complexas sob o ponto de vista sintático. De facto, os dados introduzidos pelos operadores em Moradas varia de Terceiro para Terceiro, pelo que se torna bastante difícil (mas não impossível) separar a Morada em Rua, Cidade, CP e País. Contudo, trata-se apenas duma questão sintáctica, relacionada mais com limpeza de dados (Oliveira 2008) do que com mapeamento baseado em ontologias.

Apresenta-se de seguida um excerto dos dados migrados de acordo com mapeamento apresentado:

```

<a:Terceiro rdf:ID="i-1249296751470-419352991"
  a:CodTerceiro="1"
  a:Grupo="15"
  a:Nome="HM Mauritz">
  <a:temConta rdf:resource="#i-1249296751470-70919791"/>
  <a:temMorada rdf:resource="#i-1249296751470-1524121177"/>
</a:Terceiro>
<a:Terceiro rdf:ID="i-1249296751470-538309871"
  a:CodTerceiro="1"
  a:Grupo="15"
  a:Nome="HM Mauritz">
  <a:temConta rdf:resource="#i-1249296751470-1481906237"/>
  <a:temMorada rdf:resource="#i-1249296751467-1792237266"/>
</a:Terceiro>
<a:Terceiro rdf:ID="i-1249296751475-2057478880"
  a:CodTerceiro="1"
  a:Grupo="15"
  a:Nome="HM Albanru">
  <a:temConta rdf:resource="#i-1249296751475-76641701"/>
  <a:temMorada rdf:resource="#i-1249296751472-729275891"/>
</a:Terceiro>

```

```

<a:Terceiro rdf:ID="i-1249296751472-1423191485"
  a:CodTerceiro="2"
  a:Grupo="10"
  a:Nome="GAP">
  <a:temConta rdf:resource="#i-1249296751472-103536258"/>
  <a:temMorada rdf:resource="#i-1249296751472-186807117"/>
</a:Terceiro>
<a:Conta rdf:ID="i-1249296751470-70919791"
  a:NumContaPOC="2113 1"/>
<a:Conta rdf:ID="i-1249296751472-103536258"
  a:NumContaPOC="2113 2"/>
<a:Conta rdf:ID="i-1249296751470-1481906237"
  a:NumContaPOC="2112 1"/>
<a:Conta rdf:ID="i-1249296751475-76641701"
  a:NumContaPOC="2113 1"/>
<a:Morada rdf:ID="i-1249296751467-1792237266"/>
<a:Morada rdf:ID="i-1249296751470-1524121177"/>
<a:Morada rdf:ID="i-1249296751472-186807117"/>
<a:Morada rdf:ID="i-1249296751472-729275891"/>

```

Figura 24 - Extracto da migração de dados com mapeamento

O resultado da transformação em termos tabulares é apresentado de seguida:

legado:Terceiro				
CodTerceiro	Grupo	Nome	temConta	temMorada
1	15	HM Mauritz	i-1249296751470-70919791	i-1249296751470-1524121177
1	15	HM Mauritz	i-1249296751470-1481906237	i-1249296751467-1792237266
1	15	HM Albanru	i-1249296751475-76641701	i-1249296751472-729275891
2	10	GAP	i-1249296751472-103536258	i-1249296751472-186807117

legado:Conta	
ID	NumContaPOC
i-1249296751470-70919791	2113 1
i-1249296751470-1481906237	2112 1
i-1249296751475-76641701	2113 1
i-1249296751472-103536258	2113 2

legado:Morada	
ID	
i-1249296751467-1792237266	
i-1249296751470-1524121177	
i-1249296751472-186807117	
i-1249296751472-729275891	

Figura 25 - Resultado da transformação sobre os dados legados

Como se constata, e era expectável, são criados vários Terceiro e Conta no sistema SCM que não obedecem à semântica descrita anteriormente.

Este é um problema de identidade de objectos e que é tratado pelo componente de Pós-processamento.

4.4.3 Pós-processamento

O pós-processamento é a fase da migração de dados que processa os dados resultantes da transformação antes destes serem convertidos para a linguagem e modelo de dados do repositório de destino.

O MAFRA Toolkit tem originalmente suporte para garantir a semântica na identidade de objectos, mas a especificidade do cenário em causa exige que o serviço existente seja adaptado para este cenário de forma a ser capaz de identificar e de resolver as situações convenientemente.

Começando pelo conceito Terceiro, a identidade do objecto é definida pelo atributo Grupo, i.e. a chave única da tabela Terceiro é Grupo. Assim, das instâncias geradas pelo processo de transformação com determinado valor de Grupo, deve ser escolhida uma aleatoriamente, sendo todas as outras removidas. Contudo, a restante informação de Terceiro (i.e. temConta e temMorada) são diferentes de instância para instância e devem ser mantidos. O que se pretende portanto é que o mesmo Terceiro tenha várias relações temConta e várias temMorada.

Como exemplo e em termos tabulares, a Tabela 15 representa os dados anteriores pós-processados.

legado:Terceiro				
CodTerceiro	Grupo	Nome	temConta	temMorada
1	15	HM	i-1249296751470-70919791	i-1249296751470-1524121177
		Mauritz	i-1249296751470-1481906237	i-1249296751467-1792237266
		HM	i-1249296751475-76641701	i-1249296751472-729275891
2	10	GAP	i-1249296751472-103536258	i-1249296751472-186807117

Tabela 15 - Terceiros do Sistema Legado pós-processamento

Analisando agora a tabela Conta, percebe-se que várias contas estão semanticamente repetidas, pelo que têm de ser removidas.

Cada instância de Conta tem como identidade o atributo NumContaPOC, pelo que apenas uma das instâncias com valor repetido deve ser mantida. Contudo, essa linha deve ser identificada para que o valor do atributo temConta (do conceito Terceiro) possa ser actualizado com o identificador da linha que foi mantida.

legado:Conta	
ID	NumContaPOC
i-1249296751470-70919791	2113 1
i-1249296751470-1481906237	2112 1
i-1249296751475-76641701	2113 1
i-1249296751472-103536258	2113 2

Tabela 16 - Tabela de contas

Assim, a tabela Terceiro é actualizada no que diz respeito ao atributo temConta, da forma como se apresenta de seguida:

legado:Terceiro				
CodTerceiro	Grupo	Nome	temConta	temMorada
1	15	HM	i-1249296751470-70919791	i-1249296751470-1524121177
		Mauritz	i-1249296751470-1481906237	i-1249296751467-1792237266
		HM Albanru	i-1249296751475-76641701	i-1249296751472-729275891
2	10	GAP	i-1249296751472-103536258	i-1249296751472-186807117

Tabela 17 - Tabela de Terceiros atualizada

Finalmente, note-se que o nome de terceiro “HM Albanru” vai ser descartado, já que ele se identifica por pertencer ao mesmo grupo de “HM Mauritz”, portanto é o mesmo terceiro.

A especificação deste processo é realizada ao nível do RDF pois o MAFRA Toolkit não tem interface gráfica que o suporte. A Tabela 18 representa o código RDF correspondente:

```
<post:SimpleMasterRelationIdentity rdf:ID="TerceiroPOC">
  <post:inScopeOf rdf:resource="&a;Terceiro"/>
  <post:givenBy rdf:resource="&a;Grupo"/>

  <post:masterRelation rdf:resource="&a;temConta"/>
  <post:sourceProperty rdf:resource="&a;CodTerceiro"/>
  <post:masterProperty rdf:resource="&a;NumContaPOC"/>
  <post:regularExpression>\s[0-9]+</post:regularExpression>
</post:SimpleMasterRelationIdentity >
```

Tabela 18 - Especificação RDF do processo pós-processamento

Sendo declarativa, a especificação é rápida, mas torna-se bastante susceptível ao erro devido à linguagem de representação XML/RDF.

4.5 Conclusão

Usando a abordagem declarativa baseada no MAFRA Toolkit, conseguiu-se resolver um subconjunto dos problemas identificados e descritos na secção 3.2.3.2, nomeadamente a identificação de terceiros usando o pressuposto de Código Agrupador único, e da identificação e associação de endereços a terceiros.

Trata-se duma abordagem declarativa, em que as relações semânticas entre os repositórios são estabelecidas entre os seus esquemas/ontologias, sendo posteriormente aplicadas por um motor de integração de dados na execução da transformação dos dados entre repositórios, e.g. (Almeida et al. 2008).

Apesar da declaratividade da abordagem, devido à heterogeneidade estrutural e semântica dos repositórios, torna-se habitualmente necessário o desenvolvimento de componentes particulares de especificação de equivalência semântica. Esta necessidade surgiu também no caso em apreço, sendo que a complexidade e extensão da resolução ficou em muito aquém da resolução descrita no capítulo anterior.

Esta abordagem resolveu os problemas de mapeamento de terceiros, contas POC e moradas, não solucionados nos processos automático e manual com base em folhas de cálculo (secção 3.2.3). Contudo, não corrigiu os problemas de limpeza de dados (e.g. endereços parecidos que deveriam ser o mesmo ou países inseridos em atributos errados) mas que, para os quais se preconizam abordagens e metodologias (Oliveira 2008), para no futuro se desenvolverem sistemas complementares que não estão previstos na ferramenta de mapeamento MAFRA Toolkit.

Este esforço estava inicialmente previsto, mas alterações substanciais nos requisitos do processo de adaptação dos sistemas legado e SCM veio alterar substancialmente o rumo.

5 Migração continuada

Durante o desenvolvimento do processo de integração surgiram outros desafios a acrescentar aos problemas não resolvidos nas abordagens anteriormente apresentadas.

Desafios lançados por:

- alterações orgânicas e funcionais na organização - novos modelos de negócio desviam processos anteriormente executados na sede para as filiais obrigando a uma reformulação da codificação de materiais, entidades e documentos no sistema SCM, assim como a migração da utilização de aplicações não previstas para o sistema SCM;
- obrigações legais – a produção de um novo documento electrónico fiscal SAFT-PT¹¹(SAFTPT.COM 2009)(Diário da República 2007) que inclui transacções e contabilidade residentes no sistema legado, o que obriga a migrações de dados não previstas anteriormente para o sistema SCM, e com um carácter incremental;
- o próprio processo de migração em desenvolvimento – a migração de dados, agora expandida, deverá gerar no sistema SCM um estado inicial, e ser um processo iterativo que actualize o sistema SCM com dados gerados no sistema legado incrementalmente.

Assim passamos a ter três vectores ou eixos de preocupação:

- Problemas não resolvidos até então e que ainda se mantêm – em especial os erros de mapeamentos na migração de terceiros;
- Migração de utilização de aplicações para o SCM não previstas anteriormente;
- Operação em simultâneo dos dois sistemas em produção no período de migração, sendo os dados armazenados no sistema SCM gerados pelas aplicações do SCM e migrados continuamente do sistema legado.

Neste capítulo exploramos as consequências desta adaptação organizacional no processo de migração, e descrevemos o desenvolvimento das ferramentas de migração.

5.1 Revisão

No processo migração ad hoc descrita no Capítulo 3, sustentado pela manipulação de dados em folhas de cálculo, houve necessidade de intervenção humana demorada (e.g. identificação dos vários registos do mesmo terceiro do sistema legado a serem mapeados para um único no sistema SCM; selecção do

¹¹ Os ficheiros SAFT-PT têm estrutura XML definida pela inspecção-geral de impostos, são gerados em qualquer momento e armazenam dados contabilísticos da empresa num período, geralmente anual. Gerados a pedido de um inspector de finanças serão enviados ao fisco.

nome de terceiro a migrar) que se mostraram ineficazes dado o volume de dados a transformar. Em particular, esse processo apresentou as seguintes falhas:

- não manteve um dos números sequenciais de identificação de terceiro do sistema legado na identificação do terceiro no sistema SCM, como inicialmente planeado, por incompatibilidades do tipo descrito na Tabela 5;
- não fez uma equivalência inequívoca dos terceiros entre os sistemas, já que não houve um registo sistemático no sistema SCM, para um mesmo terceiro, de todas as chaves dos registos do sistema legado;
- o integrador não tem mecanismos para uma nova iteração da migração que evite a geração de registos duplicados no sistema SCM.

Por outro lado, a migração declarativa descrita no Capítulo 4 é uma réplica do processo ad hoc descrita no Capítulo 3, pelo que excluindo a intervenção humana requerida nesse, todos os restantes problemas se mantêm neste.

Em conclusão, os processos automático e manual da migração ad hoc mostraram-se ineficazes, e não tiveram impacto na utilização dos sistemas, i.e. não chegaram a motivar o início da utilização do novo sistema. O sistema declarativo é mais eficaz mas não responde a todos requisitos.

5.2 Novo contexto

No final do processo de migração declarativa novos requisitos foram introduzidos num novo contexto envolvente, condicionando a solução e requerendo novas abordagens promovendo uma migração construída “em produção” que responda aos desafios enumerados com soluções nos três vectores descritos na introdução do capítulo.

Além disso, e como se verá pela descrição nas secções seguintes, o processo de migração ad hoc não acompanha as mudanças estratégicas observadas no novo contexto.

As subsecções seguintes descrevem detalhadamente três requisitos orientadores a ter em conta nesta nova abordagem.

5.2.1 Migração de outras áreas

A curto e médio prazo, a utilização de aplicações nos dois sistemas far-se-á da seguinte forma:

- o sistema legado continuará a ser usado para:
 - armazenar especificações técnicas de malhas;
 - consultas históricas de dados técnicos de artigos de vestuário;
 - processamento de vencimentos;

- processamento e expedição das encomendas no sistema legado (até se esgotarem as encomendas registadas neste sistema);
- o sistema SCM continuará a ser usado para:
 - especificação de novos artigos de vestuário;
 - processamento de novas encomendas;
 - planeamento e operação da cadeia de fornecimento;
- o sistema SCM será adicionalmente usado para:
 - produção do documento electrónico SAFT-PT;
 - facturação a clientes e empresas do grupo;
 - contabilidade, incluindo dados de anos anteriores.

É de realçar que este processo de migração e integração alonga-se no tempo para além do esperado no início dos trabalhos da migração, com processos iterativos de identificação e resolução de problemas de mapeamento e conseqüentemente de migração.

Na Figura 26 dispomos de um cronograma que nos dá uma panorâmica da integração, de onde se destacam as três abordagens à migração observadas:

- Migração ad hoc automática;
- Migração ad hoc manual;
- Migração continuada.

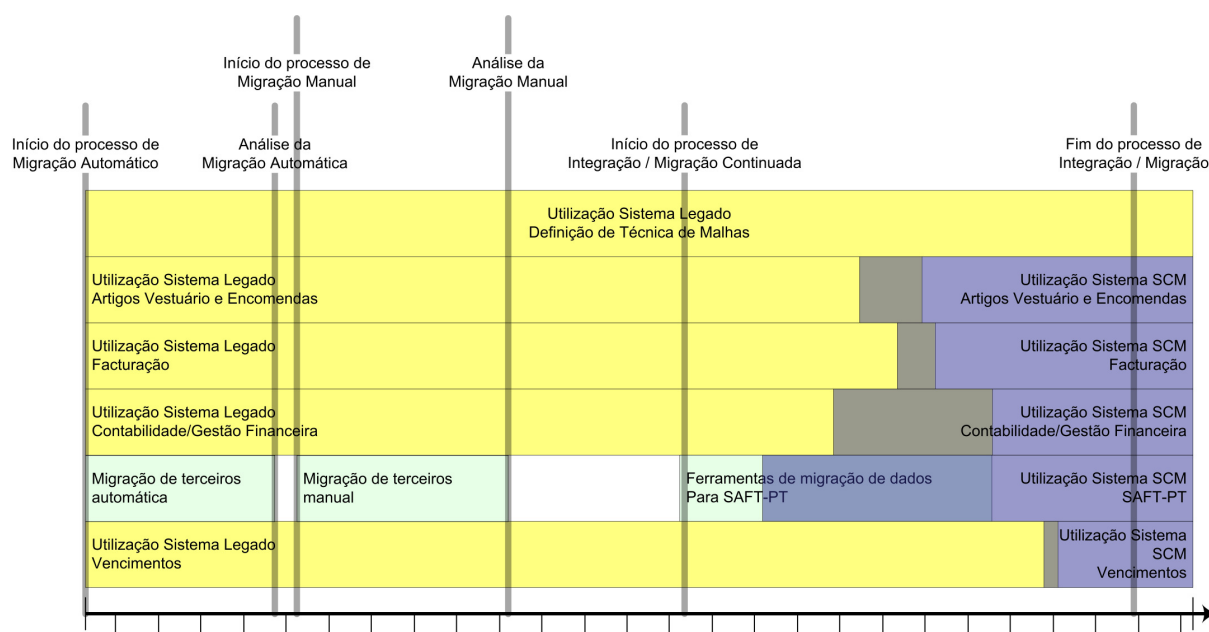


Figura 26 - Cronograma das diferentes fases e tentativas de migração

Este cronograma é uma evolução do apresentado na Figura 14 (pp.49). É notório que as abordagens ad hoc, automática e manual, não motivaram a utilização das aplicações no sistema SCM nem o abandono da utilização do sistema legado, pelos problemas encontrados (já referidos), e pelo cruzamento com as mudanças de contexto descritas nesta secção.

A barra de migração desceu da terceira para a quinta posição, ilustrando a nova dependência que as áreas aplicacionais de “Artigos de Vestuário e Encomendas”, “Facturação” e “Contabilidade/Gestão Financeira” têm relativamente ao resultado da Migração Continuada, e apresenta agora uma nova aplicação, o SAFT-PT, só presente no sistema SCM.

Mantém-se a utilização no sistema legado da “Descrição Técnica de Malhas” e as áreas aplicacionais “Vencimentos” não dependem desta ferramenta, sendo que esta última será alvo de migração no futuro.

5.2.2 Sobreposição de funções entre sistemas

Neste processo de transição das operações do sistema legado para o novo SCM, deve existir um período em que a utilização dos dois sistemas se faça de forma continuada porque:

- quando se iniciar o processamento de encomendas e facturação no sistema SCM, continuarão a produzir-se dados de gestão no sistema legado para as encomendas registadas nesse sistema e ainda em processamento;
- por imposição legal, existirão emissões no novo sistema SCM de ficheiros SAFT-PT sobre dados que foram gerados no sistema legado, e.g. facturação e movimentos contabilísticos.

Porque não haverá interrupção nas operações, a par do início de utilização gradual das diferentes áreas no sistema SCM, continuarão a inserir-se terceiros e movimentos (e.g. facturas e outros documentos contabilísticos) no sistema legado, pelo que estes dados serão migrados para o sistema SCM.

Em particular, na facturação apresentam-se os cenários:

- 1) migração com sobreposição – existirão facturas a ser geradas em simultâneo no sistema legado e no SCM. Por força do SAFT-PT as facturas geradas no sistema legado serão replicadas para o SCM, não podendo haver conflitos de numeração e data, i.e. a sequência de datas de emissão deverá ser sincronizadamente crescente com a numeração. Para contornar este problema poder-se-á:
 - a) criar um serviço no sistema SCM que forneça o número de documento a ser utilizado pelo sistema legado sempre que este necessitar de criar um documento, centralizando a numeração de documentos, permitindo assim que quando o documento for copiado do sistema legado para o novo sistema SCM seja inserido sem os problemas de numeração referidos;
 - b) emitir facturas com séries distintas no sistema legado e no sistema SCM, i.e. os sistemas terão numeradores distintos e na impressão de documentos serão impressos “dísticos” ou “siglas” que clarifiquem que série se imprime, e.g. “FA-S1 – 0223”; “FA-S2 – 0223”.

O sistema legado não irá conhecer as facturas geradas no sistema SCM. Situação semelhante se encontra com os movimentos contabilísticos.

2) migração sem sobreposição – neste cenário a facturação é um processo mutuamente exclusivo relativamente à sua utilização nos dois sistemas. Uma vez iniciada no sistema SCM é interrompida no sistema legado. Isto implica que todo o circuito das encomendas registadas no sistema legado termine com a facturação dessas encomendas antes de se iniciar a facturação no sistema SCM. Existirá uma paragem dos processos no sistema SCM, até que se esgotem as encomendas a processar no sistema legado. Este cenário prevê um período superior a uma semana de paragem no processamento de dados e nas operações, o que é crítico para o negócio, e reprova à partida esta abordagem.

Os documentos registados até então no sistema legado, migrarão para o sistema SCM para possibilitar a produção dos ficheiros SAFT-PT para esse período.

Ou seja, no cenário 1) existirá, a necessidade de usar determinados dados do sistema legado no novo sistema SCM, gerados após o momento de início da migração e de operação do novo sistema SCM, e que por via das ferramentas de migração de dados para SAFT-PT irão migrar e integrar-se no novo SCM.

Assim, nesse momento os dois sistemas começam a divergir no seu conteúdo, mas nas áreas de facturação, contabilidade e gestão financeira têm de ser usados em determinadas aplicações como se do mesmo conteúdo se tratasse. O processo de migração e integração é contínuo.

No cenário 2) encontramos três fases distintas e mutuamente exclusivas: produção no sistema legado; migração para o sistema SCM; produção no sistema SCM.

5.2.3 Mudança de layout organizacional

Por outro lado a organização sofreu uma alteração no modelo das operações envolvendo as empresas do grupo. Na continuidade das motivações do processo de desverticalização referido em 1.1.2 (pp. 18), as operações de encomendas de materiais e subcontratação de serviços de fabricação passaram a ser responsabilidade funcional das empresas satélite orientadas para a fabricação, assumindo a sede os papéis de especificação técnica de produto e gestão de grupo, colocando as encomendas de cliente nas empresas fabricantes (Fab 1; Fab 2; Fab 3; Fab 4). Estas adquirem os materiais necessários, subcontratam serviços, fabricam e expedem para o cliente final. No entanto, as operações sobre o sistema serão desempenhadas em concreto, independentemente da empresa que gera e emite documentos de requisição e facturação, por funcionários de todo o grupo.

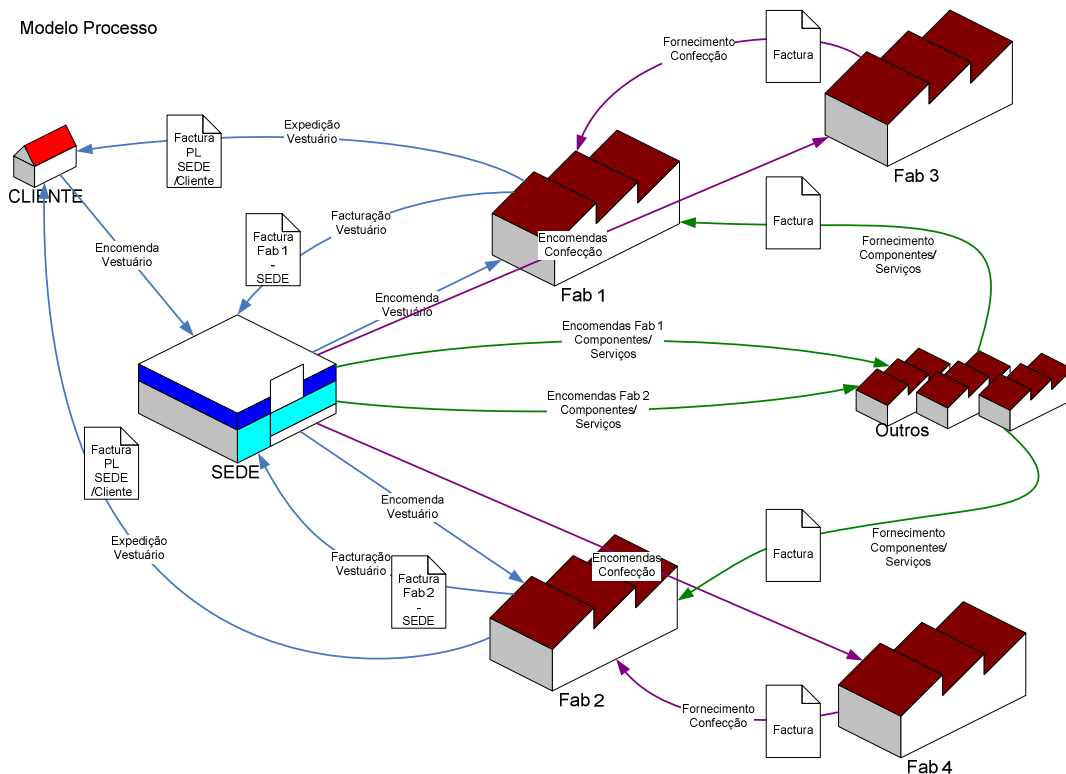


Figura 27 - Novo modelo do processo no grupo

O cenário descrito na Figura 27 implica uma deslocalização da utilização de aplicações do sistema legado e uma adaptação do sistema SCM. Neste último vão-se operar mudanças ao nível da partilha de base de dados pelas organizações que compõem o grupo, configuração de documentos e formulários, e codificação de terceiros e materiais.

Estas mudanças reflectem-se no processo de migração continuada, refinando-se os mapeamentos e transformações operadas na migração de dados.

No Anexo 3 descreve-se com detalhe o modo de operação no novo layout assim como implicações de implementação no sistema SCM e na abordagem à migração continuada.

5.3 Vista global

Com vista à resolução dos problemas impostos neste novo contexto, foi decidido adoptar um processo de migração continuada, caracterizado por não ter um termo previsto, ser construída “em produção” e apresentar iterações de tarefas ao longo do tempo em dois planos:

- Especificação: Os modelos, métodos e processos vão sendo definidos e alterados ao longo do tempo, de acordo com uma especificação inicial, evoluindo com a própria experiência de migração;
- Execução: Os processos envolvidos vão sendo executados iterativamente porque os dados do repositório fonte sofrem actualizações, e porque as regras de transformação para o repositório destino podem também mudar.

Esta abordagem preconiza uma metodologia em que o desenvolvimento e a especificação da migração evolui com a experiência de migração sobre dados reais, desenvolvendo enquanto se executa. Além disso, existe uma nova aplicação, destinada unicamente a facilitar o processo de Migração Contínua.

Na Figura 28 descrevemos o cronograma do processo, com foco na Migração Contínua.

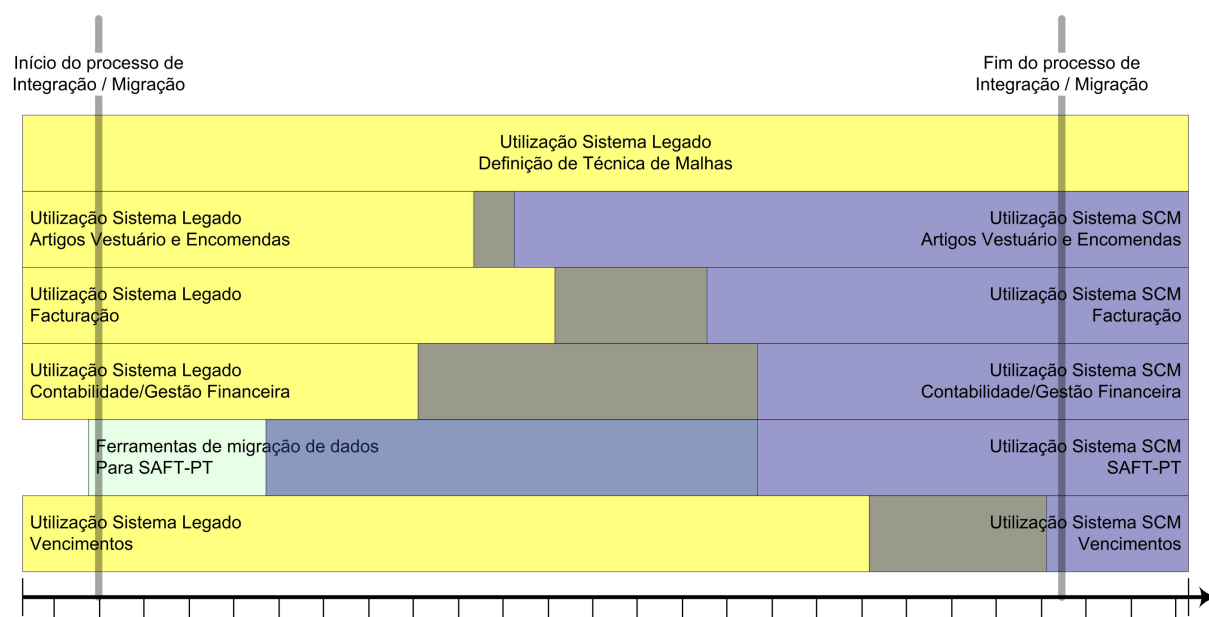


Figura 28 - Cronograma da Integração/Migração Contínua

Este cronograma apresenta uma perspectiva ampliada da sobreposição da utilização dos dois sistemas juntamente com a utilização da ferramenta de migração de dados para SAFT-PT, da sequência do início de operação das diferentes áreas no sistema SCM e dos momentos relativos de abandono do sistema legado.

Assim, observa-se que após o início da utilização da ferramenta de migração SAFT-PT inicia-se a utilização dessa aplicação no sistema SCM, antes ainda da utilização das restantes áreas aplicacionais dependentes (parte cinzenta da linha 5 do cronograma).

Gradualmente, e consecutivamente, deixar-se-á de utilizar no sistema legado as aplicações de “Artigos de Vestuário e Encomendas”, “Facturação” e “Contabilidade/Gestão Financeira”, o que resulta do esgotamento das encomendas registadas no sistema legado.

5.4 Ferramenta de mapeamento e migração

Considerando os três novos requisitos, adoptou-se um processo de integração entre o sistema legado e o SCM para dados contabilísticos e terceiros que, com apoio de intervenção humana, irá mapear de forma inequívoca os “terceiros” entre os sistemas, permitindo um processo que se desenrola no tempo, e à medida que a integração é necessária.

Este processo configura, em parte, o modelo de “Integração de catálogo” esquematizado na Figura 4, onde num processo continuado, tradutores inserem informação de diversos repositórios num repositório central, acedido para “consultas” por um tradutor que os insere num repositório final, de produção.

Há contudo algumas diferenças em relação à integração de catálogos originalmente descrita. A principal prende-se com a existência dum operador humano. Na Figura 29 esquematiza-se o modelo adoptado, suportado em tabelas e lógica de mapeamento.

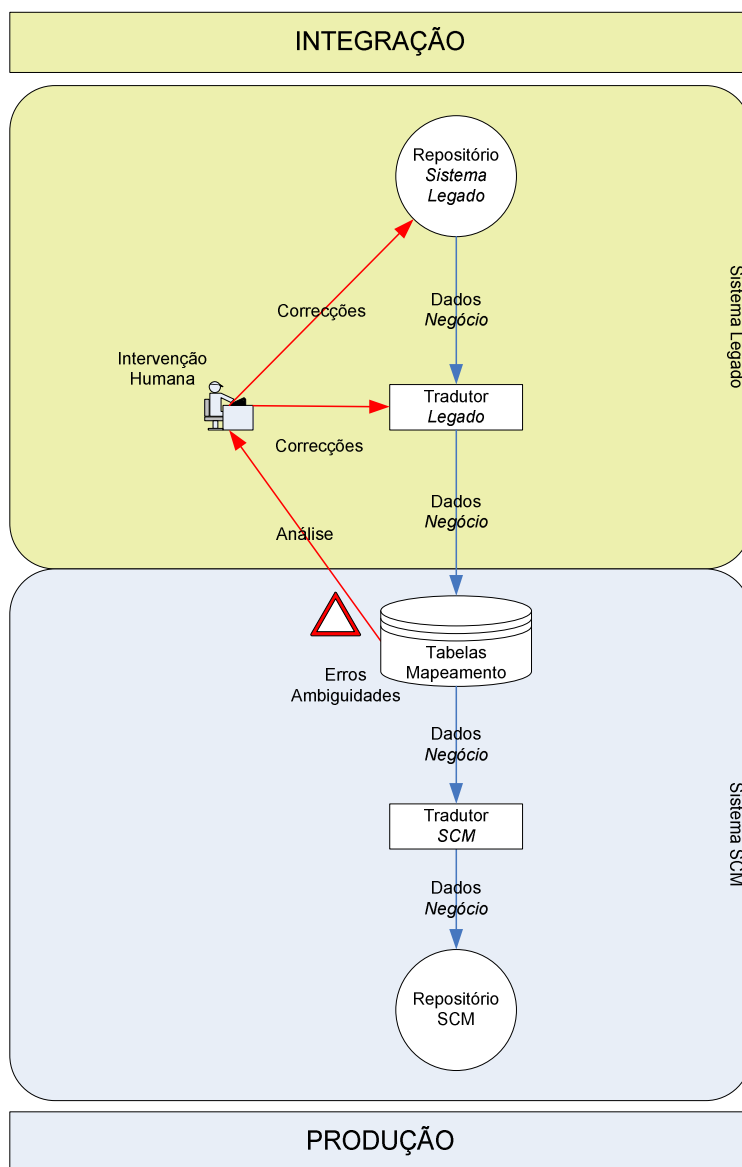


Figura 29 - Novo processo de integração de dados do negócio “em produção”

Este processo iterativo é executado em dois passos:

- 1) migração do repositório do sistema legado para as tabelas de mapeamento;
- 2) migração das tabelas de mapeamento para o repositório do sistema SCM, contemplando a inserção de novos registos, assim como a actualização dos já migrados anteriormente.

O processo é suportado por duas aplicações:

- Tradutor Legado. Esta aplicação interactiva lê registos seleccionados no sistema legado e insere os registos transformados em tabelas, aqui designadas de “tabelas de mapeamento”. Este processo é executado pelo operador periodicamente, e compreende tarefas por lotes e interactivas;
- Tradutor SCM. Utilitários desenvolvidos pelo fornecedor do sistema SCM, executados por lotes periodicamente, lêem as tabelas de mapeamento e inserem os dados processados nas tabelas de dados de produção do sistema SCM.

Esta abordagem permite uma migração contínua com os sistemas em produção. Será executada numa operação inicial que fará o estado inicial do sistema SCM, e iterações subsequentes alimentarão incrementalmente o sistema SCM com dados do sistema legado até se esgotarem as encomendas em curso neste último sistema.

O utilizador desencadeia a migração de forma repetida ao longo do período de migração executando o Tradutor Legado. Intervém no processo de migração de terceiros gerindo as identificações de terceiros no sistema SCM e corrigindo irregularidades.

O Tradutor Legado dispõe de dois modos de operação:

- modo de ensaio, fazendo a passagem de dados para folhas de cálculo;
- modo de produção, fazendo-a para as tabelas de migração no servidor SQL Server.

Os dados são então inseridos nas tabelas de mapeamento, ou actualizam os já existentes.

O Tradutor SCM é executado periodicamente de forma automática, lendo as inserções e alterações registadas nas tabelas de mapeamento, e aplicando-as às tabelas do sistema SCM.

A ferramenta está munida de funcionalidades que por um lado automatizam algumas tarefas repetitivas e sem aproveitamento das competências do operador (e.g. numeração automática de novos terceiros, migração automática de movimentos contabilísticos, identificação de erros), e por outro que facilitam a navegação e manipulação das linhas da tabela de mapeamento de terceiros para a identificação de terceiros já existentes no sistema SCM.

No anexo 4 descrevem-se detalhadamente casos de migração salientando-se o modo de operação assim como os problemas, dificuldades e soluções encontradas.

A prática veio a influenciar as funcionalidades da ferramenta, que sofreu sucessivos refinamentos para responder aos casos de erro que surgiram.

Uma das alterações é de âmbito conceptual: houve necessidade de expandir o modelo de mapeamento da Figura 13. Veio-se a verificar que um endereço de uma entidade pode ser endereço de destino de

uma outra entidade, assim a Figura 30 ilustra, de forma genérica, esta adaptação do modelo à realidade.

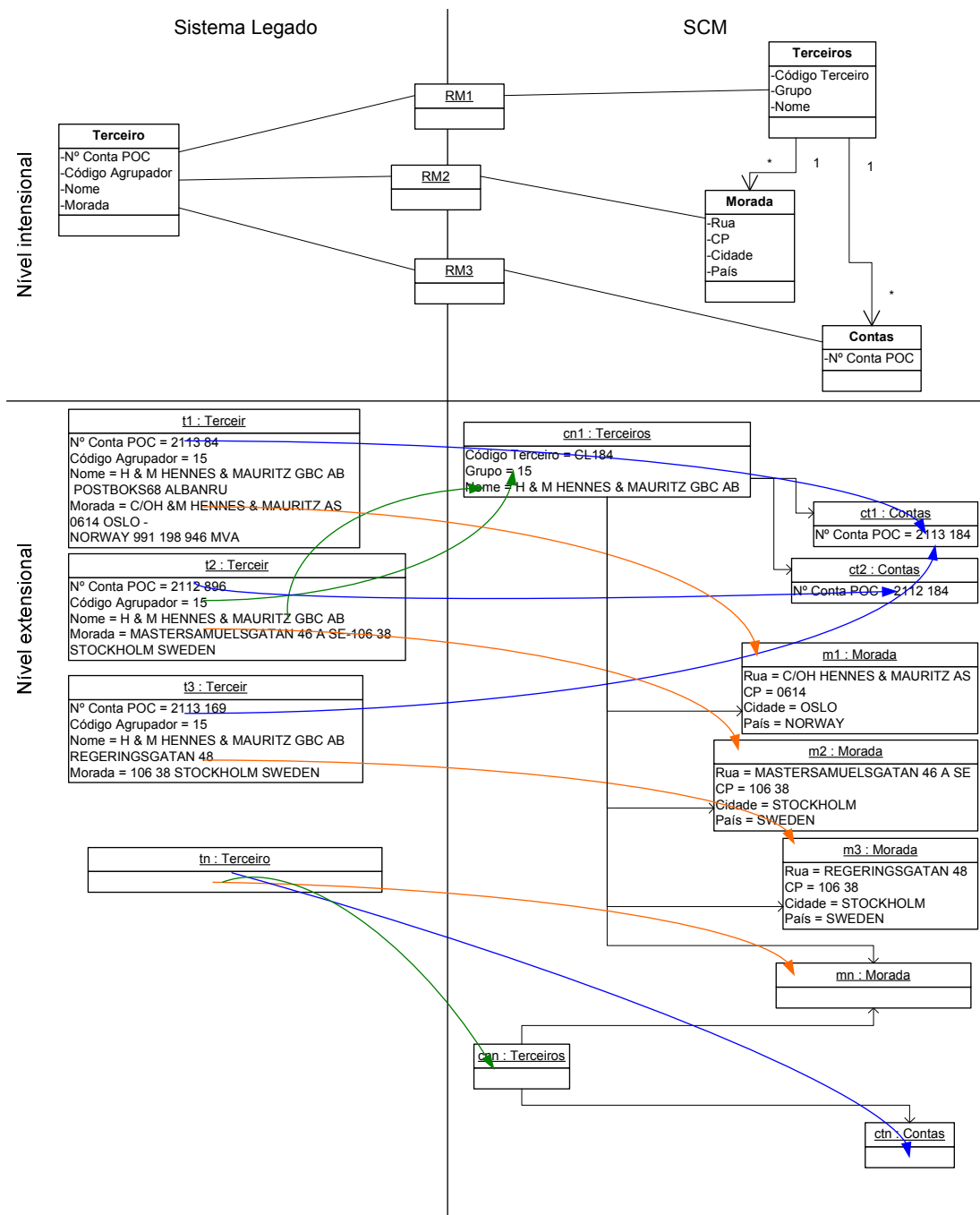


Figura 30 - Extensão a Mapeamento de terceiros do Sistema Legado para o SCM

Um terceiro “tn”, mapeado para “cn”, partilha o endereço “mn” com o terceiro “cn1”, i.e, o endereço “nm” de “cn” é destino de mercadoria do terceiro “cn1”.

Está garantida a rastreabilidade dos mapeamentos entre os dois sistemas.

5.5 Conclusões

Este método de integração revelou-se satisfatório no contexto organizacional e nos resultados obtidos, salientando-se que:

- Resolve-se o problema recorrente da diferente granularidade entre os sistemas quanto a Terceiros;
- Encontra-se uma forma bijectiva de mapeamento de terceiros entre ambos os sistemas;
- Total autonomia, pós-processamento manual, para a conversão, migração ou integração de dados entre os sistemas que agreguem identificação de terceiros, i.e. documentos contabilísticos.

Contudo apresenta as seguintes condicionantes:

- Processamento manual, i.e. dependência de conhecimento humano;
- Geração automática da identificação do terceiro no SCM, com opção de identificação manual;
- Os terceiros continuarão a ser criados no sistema legado e este processo irá migrá-los para o SCM quando se verificarem movimentos contabilísticos, financeiros ou de negócio que justifiquem a migração;
- As alterações de dados de terceiros do sistema legado serão repercutidas no sistema SCM de forma automática, mas as alterações em documentos (e.g. facturação) no sistema legado deverão ter um processo manual de actualização no SCM.
- O termo da utilização da ferramenta de migração dá-se com o termo da utilização de todas as aplicações no sistema legado dependentes (linha vertical direita na Figura 28).

6 Conclusão

As mudanças observadas no mundo dos negócios impelem mudanças estruturais, nomeadamente nos sistemas de informação, já por si difíceis de implementar neste ambiente conjuntural. Fazer acompanhar estas mudanças da preservação de dados disponíveis, e por consequência da cultura da organização residente nos sistemas, representa um desafio aos decisores na área da gestão da informação e operações de negócio.

Por outro lado, uma mudança de sistemas de informação far-se-á de forma mais suave reduzindo a natural inércia dos utilizadores à utilização dos novos paradigmas e sistemas, quanto mais informações históricas e conhecidas as novas interfaces revelarem aos utilizadores. De facto, encontrar um sistema vazio ou um sistema com elementos já familiares embora reformatados, provoca impactos diferentes para um utilizador com repercussões na sua interação com a organização. É desejável portanto o crescimento dos sistemas de informação com integração de sistemas legados.

O caso de estudo, caracteriza-se por se tratar de uma PME na ITV, num contexto pouco favorável de negócio, onde as reestruturações se sucedem. A organização opera mudanças no layout, reformulando a organização no sentido da agilização e refinamento de processos, focalização no cliente e abertura à subcontratação no mercado. Estas alterações motivaram a adopção de um novo sistema que cubra novas funcionalidades e suporte o negócio no novo paradigma de cadeia de fornecimento. No entanto, a empresa dispõe no sistema legado de bastante informação corrente, actual e necessária para a continuidade da laboração.

A opção por não integrar o sistema legado significaria enfrentar um dos cenários:

- a ausência dessa informação no novo sistema, o que obrigaria à digitação dos dados assim que necessários (e.g. entidades, matérias-primas), com impactos na operação, nomeadamente tempos de espera para informação complementar (e.g. a expedição teria que aguardar que os serviços administrativos digitassem informação complementar de facturação; a colocação de requisições a fornecedores aguardaria pela especificação técnica de malhas). A natureza das encomendas implica a existência de muitos subprodutos com bastantes especificações, assim como muitas especificações no trato das expedições e exportações, estando já estas informações em sistema. Digitá-las a pedido implica consumos de tempo incompatíveis com a urgência do processo de expedição.

Neste cenário também se incorria na irregularidade legal de não dispor de ficheiros SAFT-PT a partir do ano de 2008, estando esta funcionalidade disponível só no novo sistema SCM.

- digitação da informação no novo sistema. Este cenário resume-se à inserção de enormes quantidades de informação relativas a entidades, matérias-primas e semi-acabadas, dados contabilísticos e facturação. Esta prática representaria um custo considerável uma vez que o

volume de dados e o tempo de digitação ocupariam vários colaboradores, retirando-lhes disponibilidade para a força de vendas e subcontratação. Este modelo implicaria praticamente uma paragem na operação, ainda não avaliada.

Durante o processo descrito neste documento, tomou-se conhecimento de casos semelhantes em que os responsáveis optaram por não integrar os sistemas legados, motivados quer pelos custos envolvidos, quer pela complexidade na sua implementação.

Como exemplo surge-nos nesta área de negócio a State of Art®¹², uma organização na ITV com origem na Holanda que detém uma cadeia de lojas de roupa para homem em várias cidades da Europa. Num cenário de mudança de direcção dos serviços de informática e evolução do modelo de negócio, apoiada por uma administração pouco comprometida com o processo de integração e mais entusiasmada com os benefícios espectáveis com a adopção dos novos sistemas, implementou um novo ERP, o Microsoft Dynamics Nav. Apesar de dispor de um sistema legado rico em dados técnicos dos produtos, configuração de lojas e armazéns, registo de fornecedores, clientes e movimentos, não considerou nem a integração do sistema legado nem a migração de dados. Optou por iniciar a utilização do novo ERP a partir de um sistema “vazio”, como numa nova instalação, focando-se nas novas funcionalidades. Esta medida resultou em tarefas demoradas de inserção de dados já conhecidos e na abdicação da informação acumulada. A organização ficou a operar sobre um sistema completamente novo em dados e aplicações.

Este trabalho foi também uma oportunidade para um estudo da conceptualização de integração de sistemas de informação, analisando as propostas e abordagens mais recentes numa visão comparativa, e que sistematizou a concepção da solução final encontrada. O caso de estudo proporcionou três abordagens ao problema de integração, apresentados nos capítulos 4, 5 e 6.

6.1 Migração ad hoc

O processo de migração ad hoc descrito no capítulo 3 é representativo da realidade da evolução dos sistemas de informação no meio empresarial das PME.

Tratou-se de uma abordagem pragmática, *ad hoc*, iterativa e refinante, na medida em que:

- partiu-se rapidamente para os procedimentos práticos sem estudos prévios aprofundados;
- assumiu-se inicialmente que seria uma prática única com o objectivo de realizar uma só migração;
- os processos de migração vão sendo repetidamente iniciados e corrigidos numa lógica de *tentativa-falha* à medida que vão ocorrendo erros na transformação dos dados;

¹² <http://www.stateofart.com/> - Caso apresentado numa acção comercial de um representante do Microsoft Dynamics Nav.

- o processo é refinado gradualmente até conseguir terminar sem erros, ou constatar da impossibilidade da migração parcial ou total;
- não garante resultados, prazos, eficácia, qualidade.

As dificuldades encontradas nesta experiência em particular, prenderam-se com:

- o desconhecimento do domínio dos dados concretos em alguns atributos (e.g. Código agrupador de terceiros);
- a impossibilidade de mapear de forma unívoca os registos de terceiros, já que a natureza das chaves é distinta, conta POC no sistema legado (duplicação de entidades reais) e número único no sistema SCM;
- as designações de empresas escritas de forma diferente para vários registos do mesmo Terceiro no sistema legado e encontrar uma designação mais correcta para inserir no registo único de entidade no SCM;
- a inconsistência nos dados, e.g. o atributo código postal é alternadamente encontrado com nome de país, cidade e código postal de facto;
- a migração de terceiros sofreu bastantes manipulações em folha de cálculo, para corrigir erros e seleccionar os registos válidos. A intervenção humana na Migração Continuada resolveu equívocos não resolvidos automaticamente;
- o tempo inicialmente desejado para terminar a migração foi sistematicamente e bastante ultrapassado.

Um estudo prévio em profundidade poderia obviar algumas das dificuldades e evitar contratempos durante a migração.

A ausência de um estudo prévio, ou a pouca profundidade e rigor quando realizado, sintomático nesta faixa do tecido empresarial, deve-se a:

- constrangimentos orçamentais;
- pouco comprometimento da administração para com o processo;
- urgência em ver “obra feita”;
- o processo de migração é encarado no meio, como “cópia simples de tabelas”.

6.2 Migração declarativa

A Migração declarativa, descrita no capítulo 4, mostrou-se uma ferramenta sistémica que com suporte na ferramenta MAFRA Toolkit resolveu o problema de mapeamento de Terceiros.

De facto, partindo de uma tabela do sistema legado que contém num registo a informação de Terceiro, Conta POC e morada, geraram-se três tabelas no sistema SCM com uma identificação inequívoca da entidade Terceiro, contas POC e moradas, garantindo a unicidade de cada elemento.

Este ensaio revela que esta abordagem é um caminho a considerar no desenvolvimento de ferramentas que facilitem de forma económica a migração de dados.

6.3 Migração continuada

O processo de reorganização empresarial descrito no capítulo 5 promovendo alterações no layout da organização, deslocando funcionalidades para empresas satélite (e.g. facturação), e as imposições legais de produção dos ficheiros SAFT-PT, repercutiram-se em mudanças estratégicas no processo de migração dos sistemas de informação.

Áreas sem migração prevista tiveram que ser reconsideradas, como facturação, contabilidade e artigos transaccionados.

Por outro lado as asserções relativas ao domínio de Terceiros a migrar caíram por terra, sabendo-se que um movimento contabilístico no futuro pode implicar a migração de um terceiro não contemplado na “primeira migração”. Houve então necessidade de desenvolver uma ferramenta que estabelecesse os mapeamentos necessários à migração entre os dois repositórios, e que permitisse a intervenção humana para as situações ambíguas, em especial no mapeamento de terceiros. A abordagem Migração Continuada, apresentada no capítulo 5, foi a solução encontrada para este processo.

A Migração Continuada, sendo um processo iterativo e interactivo, permite uma migração em conjunto numa fase inicial, e migrações parciais em fases sucessivas com os sistemas em produção, migrando os registos novos e as actualizações. Esta solução garante que só são migrados os conjuntos de dados necessários, permite um estado inicial do novo sistema SCM na fase de arranque, tido como importante para a participação dos colaboradores envolvidos na sua utilização. Além disso, garante um mapeamento nos dois sentidos das entidades e documentos contabilísticos migrados, permitindo desta forma a rastreabilidade dos documentos entre os sistemas.

6.4 Notas finais

O contexto sócio-económico actual obriga às restrições orçamentais, proibindo investimentos avultados em integração de sistemas legados, em especial nas PME.

Por outro lado, a necessidade de reutilização de dados pelas PME na mudança de sistemas de computação é premente, observando-se metodologias não sistémicas, muitas vezes improvisadas, de migração de dados, frequentemente fracassadas. Esta tarefa envolve, de facto muitas vezes, programação, correcções manuais, digitação e manipulação de dados. É com frequência que são criadas estruturas intermédias, como tabelas e folhas de cálculo, que auxiliam na manipulação e transformação dos dados no sentido do sistema legado para o novo sistema. São por norma processos iterativos, numa lógica de *tentativa-falha* até se conseguir um conjunto de dados aceitável no novo esquema.

O engenheiro integrador é sistematicamente requisitado para resolver ambiguidades encontradas nos dados impossíveis de serem inseridos no novo sistema, ao contrário do que seria desejável com a utilização somente de ferramentas de integração.

A complexidade desta operação pode resultar no abandono parcial ou total do processo de migração, e quando bem sucedida, pode incorrer na abdicação de informação, capacidade de cálculo ou de processamentos anteriormente realizados.

Os projectos e abordagens entusiastas de integração de sistemas legados descritos no capítulo 0, com o envolvimento de grandes grupos empresariais são demonstrativos do interesse que esta área de investigação e desenvolvimento tem na continuidade dos diferentes domínios de conhecimento. Contudo, as abordagens descritas ou são resultado de investigação recente ou são levadas a cabo por medida por organizações cujo negócio central é esse. O foco da investigação está na tentativa de generalização dessas abordagens, mas fica-se com a convicção, ainda mais com a experiência relatada neste documento, de que o processo é muito dependente do domínio de aplicação e por conseguinte centrado no ser humano.

7 Trabalho Futuro

A seguir apresenta-se a enumeração de algumas dificuldades encontradas, assim como orientações para o que aparentemente seria desejável dispor como ferramentas de apoio à integração.

Também, relacionado com o caso apresentado, é apresentada uma previsão do futuro próximo.

7.1 Domínio de conhecimento

Ficou patente nesta dissertação que os sistemas de informação encapsulam uma abstracção do conhecimento de domínio de uma organização. Essa abstracção assume formas e representações ímpares de acordo com os sistemas de computação utilizados, o desenvolvimento de cada sistema e as pessoas envolvidas na construção de cada sistema.

Poderemos então partir da asserção de que existe um domínio de meta-conhecimento, que transcende a sua representação, e que é reconhecido por todos num ramo de negócio, portanto num domínio de conhecimento.

Por outro lado, as organizações vão especializando esse domínio em particularidades relacionadas com a vivência da própria organização, e dos colaboradores que são actores na construção desse conhecimento. Como referido em OMIS (secção 2.4.2) a integração deste conhecimento organizacional disperso com as tarefas de negócio é designada de Memória Organizacional.

A Memória Organizacional apresenta-se neste projecto como uma evolução natural dos sistemas de informação organizacionais, enquanto informação tangível (e.g. informação administrativa, de processo empresarial em base de dados) integra-se com informação intangível (e.g. conhecimento pericial, competências). Formaliza, descreve e relaciona a experiência da organização no tempo com a informação tradicionalmente suportada nos sistemas de informação, o que se pode tornar num facilitador para integração entre sistemas num domínio.

A geração de ferramentas que descrevam em linguagem declarativa este conhecimento em padrões aceite pela comunidade de desenvolvimento, e que permitam uma interface com as implementações de sistemas de produção das organizações, e.g. por meio de tecnologias afectas a bases de dados ou a serviços, pode ser uma perspectiva animadora no futuro da integração de sistemas em contexto de domínio de conhecimento.

Contudo, a abordagem seguida no caso de integração descrito está longe de se endereçar a este nível de conhecimento, e na conjuntura financeira actual as organizações parecem estar em contra-ciclo com esta preocupação.

7.2 Conhecimento dos dados

No sentido de potenciar o conhecimento da natureza dos dados que se pretendem migrar/integrar, a utilização de ferramentas tipicamente utilizadas em Datawarehouse pode desempenhar um papel de relevo na análise de semântica e conteúdos. A inclusão destas técnicas em ferramentas orientadas à integração, permitindo um estudo prévio com rigor, é espectável que evitaria muitas das iterações no processo de tentativa-falha já referido, nomeadamente despistando, no caso estudado, o Código Agrupador como unificador dos Terceiros Clientes do sistema legado.

7.3 Limpeza dos dados e correcções

No que concerne à problemática da limpeza de dados mostra-se importante a criação de ferramentas que permitam a identificação de erros em dados registados, partindo de pressupostos de preenchimento e padrões esperados em atributos.

Em concreto, e no caso estudado, para os problemas relacionados com o endereço, aponta-se como solução um algoritmo que concatene os atributos que compõem o endereço, e por análise léxica, sintáctica, estrutural e semântica identifique rua, código postal, cidade e país. A abordagem de (Oliveira 2008) mostra-se uma importante contribuição para esta problemática.

7.4 Conhecimento dos processos

O esquema de bases de dados e a análise de amostras de dados, não é suficiente para o levantamento conceptual de um sistema legado.

A análise dos processos que operam sobre os dados pode revelar bastante conhecimento de domínio, mas neste campo entramos numa dimensão sem escala já que as ferramentas de desenvolvimento são inúmeras e de semântica e sintaxe bastante heterogéneas. Além disso a análise de código de programas aparenta ser uma tarefa de difícil execução, e com resultados que duvidosamente contribuam para o esclarecimento dos objectivos das funcionalidades, permitindo descrever o conhecimento de domínio. Contudo, os sistemas legados poderão não ser tão “ingratos” como são tidos na generalidade.

Algum conhecimento técnico sobre as plataformas dos sistemas legados e uma inspecção às fontes das aplicações, pode fornecer pistas que apoiem a interpretação dos dados encontrados em 7.2.

Enumerando alguns:

- Os sistemas desenvolvidos com recurso a ferramentas CASE fornecem um modo estruturado e declarativo de especificação, recorrendo nomeadamente à imposição/definição de regras de negócio e/ou classificação de entidades.

Tomando para exemplo o Synon, uma ferramenta CASE para IBM AS/400, na linguagem disponível ao utilizador para embutir lógica de negócio nos programas que serão gerados pela ferramenta, as condições são definidas por classificadores (*conditions*) que associam a valores instanciados em variáveis ou atributos proposições que tomarão o valor de verdadeiro ou falso.

Assim, não existem condições ou testes aos valores das instâncias declarados nos “programas” ou rotinas, mas antes a verificação do estado do classificador sobre determinada instância. Portanto, sob um ponto de vista conceptual, este classificador é como uma classe, cujas instâncias obedecem aos requisitos para que o classificador retorne verdade para a instância. e.g. num processo não serão tratados registos de guias em estado “Desdobrada”, o que é indicado pelo preenchimento com “D” do campo com descrição “INDIC. DE TIPO”:

EDIT FIELD CONDITIONS		GERAL -	
Field name.	: INDIC. DE TIPO	Attr. :	STS
Position	: esdobrada	LST	
Enter condition . . .	:	and type to add new cond	
type . . .	:	(Type: LST, VAL)	
? Condition	Type Op	File/From value	Display/To value
Branco	VAL	*BLANK	*BLANK
Complementar	VAL	C	C
Desdobrada	VAL	D	D

Figura 31 - Condições em Synon

```

. // Guias desdobradas são irrelevantes
. -CASE
. - NOTc1
. | - c1: DB1.INDIC. DE TIPO is Desdobrada
. | -
. -CASE
. | - c1 OR c2
. | | - c1: DB1.INDIC. DE SITUAÇÃO is Cortado
. | | - c2: DB1.INDIC. DE SITUAÇÃO is Em armazém
. | -
. | Ler cortado por lote - Guia Confeção Det. PLDGT *

```

Tabela 19 - Utilização das condições em Synon

- Em programação na linguagem COBOL é frequente encontrar-se classificadores associados a variáveis ou atributos, normalmente declarações de dados nível 88 que são indicadores de valores lógicos. E.g. o atributo tipo de documento pode ser definido como alfanumérico de dois bytes na forma:

```

02 tipo-doc pic(XX).
88 factura value 'FT'.
88 nota-credito value 'NC'.
88 nota-debito value 'NB'.

```

Esta análise pode ser bastante esclarecedora para valores “incompreensíveis” encontrados nas tabelas.

- Em geral as consultas em SQL, ou SQL View, revelam as condições para que linhas de tabelas possam pertencer à vista que produzem. Uma vez mais, aqui podem ser colhidos bastantes dados sobre a conceptualização dos sistemas, e dos elementos do conhecimento de domínio que estão representados.

Esta problemática é também uma preocupação patente em alguns dos casos apresentados na secção 2.4, e sugere ser uma área de investigação promissora. Certamente, a serem produzidas ferramentas, serão um utensílio de apoio fundamental na integração de sistemas.

7.5 Outras áreas de integração no caso estudado

Outras áreas serão alvo de um processo de integração/migração como:

- Fichas técnicas de malhas - que serão mantidas no sistema legado pela riqueza de especificações, mas integradas no novo sistema SCM permitindo a sua utilização na especificação de artigos;
- Recursos humanos (salários e prémios de produção);
- O conjunto de informações de artigos, facturação e contabilidade de outras empresas do grupo que são geridas nos mesmos sistemas de computação;
- A lei fiscal nacional vai impor num futuro próximo, a partir de 2010, a utilização de um novo modelo para contabilidade designado de SNC - Sistemas de Normalização Contabilística, o que implica:
 - Reformulação do POC da organização;
 - Mapeamento de contas;
 - Novos programas relacionados com contabilidade e facturação que contemplam o novo processo contabilístico.

O fornecedor do sistema SCM reconhece a mais-valia da implementação de modelos e ferramentas de mapeamento e integração no âmbito do trabalho aqui descrito, que serão adoptados no sistema, permitindo dispor às organizações a contabilidade nos dois modelos contabilísticos.

A metodologia será a utilizada na Migração Continuada, com processos iterativos de migração/integração, com pontos de acesso interactivos para a solução de ambiguidades pelo utilizador, como descrito.

À medida que as encomendas registadas no sistema legado se forem esgotando, também deixaram de ser produzidas neste sistema facturas, movimentos de stock e contabilísticos. Então deixa de ser necessária a utilização das ferramentas de mapeamento e integração desenvolvidas, já que as operações serão suportadas pelo sistema SCM.

Contudo existe uma excepção no que diz respeito às malhas, que motivado pela riqueza de especificação existente no sistema legado, continuarão a ser descritas neste sistema, e integradas no sistema SCM. Esta ferramenta não requer intervenção do utilizador sendo completamente automatizada.

Referências bibliográficas

- Almeida, Ricardo, João Pedro Silva, Nuno Silva, e João Rocha. 2008. “Uma Solução Genérica e Faseada de Integração de Informação (CIAWI’08).” Lisboa: Instituto Superior de Engenharia do Porto
[http://www.dei.isep.ipp.pt/~nsilva/R&D/Publications/2008/Uma%20Solucao%20Generica%20e%20Faseada%20de%20Integracao%20de%20Informacao%20\(CIAWI%2708\)%20Queryin g.pdf](http://www.dei.isep.ipp.pt/~nsilva/R&D/Publications/2008/Uma%20Solucao%20Generica%20e%20Faseada%20de%20Integracao%20de%20Informacao%20(CIAWI%2708)%20Queryin g.pdf) (Acedido Março 17, 2009).
- Bontcheva, Kalina. 2008. “ISTweb - Content - Knowledge & Content Technologies - Projects - TAO.” <http://cordis.europa.eu/> (Acedido Agosto 10, 2009).
- Bontcheva, Kalina. 2007. “TAO - Transitioning Applications to Ontologies.” *TAO Fact Sheet*. <http://www.tao-project.eu/resources/publicity-materials/tao-factsheet-310507.pdf> (Acedido Março 30, 2008).
- Bontcheva, Kalina. 2008. “TAO - Transitioning Applications to Ontologies.” *TAO - Transitioning Applications to Ontologies*. <http://www.tao-project.eu/> (Acedido Março 30, 2008).
- Comissão de Normalização Contabilística. *Plano Oficial de Contabilidade*. <http://www.cnc.min-financas.pt/POC/POContabilidade.pdf> (Acedido Outubro 12, 2008).
- Diário da República. 2007. *Portaria n.o 321-A/2007 (SAFT-PT)*. <http://www.saftpt.com/docs/SAFT-T%20Portaria%20n.o%20321-A2007.pdf>.
- Frans Coenen, Barry Eaglestone, e Mick J. Ridley. 2001. “Verification, validation, and integrity issues in expert and database systems: Two perspectives.” *Int. J. Intell. Syst.*, 425-447.
- Georgakopoulos , Diimitrios , Mark Hornick, e Amit Sheth. 2008. “An Overview of Workflow Management: From Process Modeling to Workflow Automation Infrastructure.” pdf, GTE Laboratories Incorporated, <https://eprints.kfupm.edu.sa/25487/1/25487.pdf> (Acedido Setembro 13, 2008).

- Hull, Richard. 1997. "Managing Semantic Heterogeneity in Databases : A Theoretical Perspective." pdf, Bell Laboratories <http://w3.msi.vxu.se/~per/IVC743/p51-hull.pdf> (Acedido Outubro 3, 2008).
- Jérôme Euzenat, e Pavel Shvaiko. 2007. *Ontology Matching*. 1.º ed. Heidelberg, Germany: Springer-Verlag.
- Martins, Hélio Artur Mendes. 2007. "Suporte a conversão e introdução de conteúdos baseado em ontologias." Lic, Instituto Superior de Engenharia do Porto.
- Mikhail Simonov, Aldo Gangemi, e Massimo Soroldoni. 2004. "Ontology-driven Natural Language Access to Legacy and Web Services in the Insurance Domain." P. 10 em *BUSINESS INFORMATION SYSTEMS - BIS 2004*. Poland: Witold Abramowicz http://www.loa-cnr.it/Papers/BIS_2004.pdf.
- Nicola Guarino, Massimiliano Carrara, e Pierdaniele Giaretta. 1994. "An Ontology of Meta-Level Categories." LADSEB-CNR Int. <http://www.mcox.org/introspect/meta-level-ontologies.pdf> (Acedido Outubro 14, 2009).
- Oliveira, Paulo Jorge. 2008. "Detecção e Correção de Problemas de Qualidade dos Dados: Modelo, Sintaxe e Semântica." PhD, Universidade de Minho <http://repositorium.sdum.uminho.pt/bitstream/1822/9158/1/Tese%20Final.pdf> (Acedido Junho 8, 2009).
- Philippe Thiran, Jean-Luc Hainaut, Geert-Jan Houben, e Djamal Benslimane. 2006. "Wrapper-Based Evolution of Legacy Information Systems." *ACM Transactions on Software Engineering and Methodology*, Outubro, 329-259.
- SAFTPT.COM. 2009. "Portal SAFTPT.com." *Portal SAFT*. <http://www.saftpt.com/> (Acedido Abril 9, 2009).
- Sheth, Amit, e James Larsen. 1990. "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases'." http://delivery.acm.org/10.1145/100000/96604/p183-sheth.pdf?key1=96604&key2=4166133221&coll=GUIDE&dl=GUIDE&CFID=5347665&CF_TOKEN=26499503 (Acedido Outubro 6, 2008).
- Silva. 2004. "Multi-Dimensional Service-Oriented Ontology Mapping." PhD, Universidade de Trás-os-Montes e Alto Douro <http://www.dei.isep.ipp.pt/%7Eensilva/R&D/PhD/PhDThesis.pdf> (Acedido Novembro 4, 2008).
- Vasconcelos, José Braga, Feliz Ribeiro Gouveia, e Chris Kimble. 2002. "An Organizational Memory Information System using Ontologies." P. 17 em. Coimbra, Portugal <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.13.2342> (Acedido Outubro 17, 2008).

Anexo 1

Plano de Contas

No sistema legado, a identificação de terceiros usa a semântica da estrutura hierárquica usual do plano de contas (POC¹³), conforme se representa de seguida (Comissão de Normalização Contabilística):

Dígito de classe (e.g. 2)

Dígitos de conta (e.g. 111)

Dígitos para número de identificação discreta dentro da conta.

As classes agrupam as contas quanto à natureza, tal como:

Classe	Natureza
1	Disponibilidades
2	Terceiros
3	Existências
4	Imobilizações
5	Capital, reservas e resultados transitados
6	Custos e perdas
7	Proveitos e ganhos
8	Resultados
9	Contabilidade analítica

Tabela 20 - Classes de contas do POC

Esta classificação permite, observando os valores acumulados em cada classe num período, ter uma imagem da posição financeira e dos resultados das operações das organizações (Comissão de Normalização Contabilística).

¹³ Plano Oficial de Contas

Cada classe desdobra-se em contas que detalham os valores acumulados.

Tomando a classe 1 – Disponibilidades encontram-se, entre outras, as contas:

- 11 – Caixa;
- 12 – Depósitos à ordem;
- 13 – Depósitos a Prazo;
- 15 – Títulos negociáveis.

Estas contas poderão subdividir-se em sub-contas, e.g. 111 – Caixa Porto (1); 112 – Caixa Lisboa (2), materializando assim uma conta abstracta numa entidade real, por via da utilização de um dígito para a identificação discreta da entidade.

Este último não está presente em todas as contas. Só naquelas onde existe a individualização de um departamento, entidade ou instituição.

Tomando agora a classe 2 – Terceiros, encontram-se as contas:

- 21 – Clientes;
 - 211 – Clientes Conta Corrente (C/C);
 - 212 – Clientes – Títulos a receber;
- 22 – Fornecedores;
 - 221 – Fornecedores C/C;
 - 222 – Títulos a pagar;
- 23 – Empréstimos Obtidos;
- 24 – Estado e outros entes públicos;
- 25 – Accionistas (sócios);
- 26 – Outros devedores e credores;
 - 261 – Fornecedores de imobilizado;
 - 262 – Pessoal;
 - ...

No presente estudo de caso temos especial interesse na classe 2 de terceiros, nomeadamente em clientes e fornecedores.

No caso dos fornecedores observa-se o seguinte:

- a conta 22 inclui os que fornecem materiais relacionados com o produto (e.g. malhas, botões, fechos, serviços);
- a conta 261 aqueles que fornecem equipamentos para imobilizado (e.g. equipamentos industriais, para escritório, informática).

Um fornecedor pode ter movimentos contabilísticos nas duas contas se fornecer serviços e equipamentos.

De forma semelhante, um cliente será também fornecedor, quando fornece, por exemplo, etiquetas de marca para afixar nos artigos, e agregará contas de mercados distintos se se representar em países nesses mercados.

Por outro lado, um fornecedor poderá também ser cliente. É prática corrente, um fornecedor de confecção comprar linhas de costura à organização.

As contas desta classe (2) descem na hierarquia até ao nível da identificação de entidades com a indicação de um número de terceiro.

Assim, um terceiro é identificado pela classe 2, conta do POC e número de terceiro, e.g. 2113 82, onde:

- 2113, é a conta (de 3º grau) de cliente “Conta corrente Mercado não comunitário” da classe 2 – “Terceiros”¹⁴;
- 82, o “número único de terceiro dentro da conta de 3º grau 2113 pertencente à classe 2”¹⁵.

Além disso:

- Um Cliente é identificado por pertencer à conta de contabilidade 21, da classe 2;
- Um Fornecedor é identificado por pertencer à conta de contabilidade 22, 261 ou 268 da classe 2.

Portanto, um terceiro é caracterizado por:

- Classe;
- Conta;
- Número de terceiro.

Na instalação em estudo, as contas de terceiros apresentam a distribuição e nomenclatura expostas nas tabelas que se seguem.

Tipo de Terceiro	Conta	Mercado	Produtos
Cliente	2111	Nacional	Artigos de vestuário
Cliente	2161	Nacional	Malhas
Cliente	2112	Comunitário	Artigos de vestuário
Cliente	2162	Comunitário	Malhas
Cliente	2113	Extra-comunitário	Artigos de vestuário
Cliente	2163	Extra-comunitário	Malhas
Cliente	2151	Nacional	Tecidos
Cliente	2152	Comunitário	Tecidos
Cliente	2153	Extra-comunitário	Tecidos
Fornecedor	2211	Nacional	Serviços e materiais
Fornecedor	2212	Comunitário	Serviços e materiais
Fornecedor	2213	Extra-comunitário	Serviços e materiais
Fornecedor	26111	Nacional	Imobilizado
Fornecedor	26112	Comunitário	Imobilizado

¹⁴ Expressão retirada do POC

¹⁵ “Classe” é o termo usado no domínio contabilístico para categorizar/caracterizar contas do POC.

Fornecedor	26113	Extra-comunitário	Imobilizado
Fornecedor	2682	Nacional	Prestações de Serviços
Fornecedor	2689		Outros devedores e credores

Tabela 21 - Contas de Terceiros por mercado e produto

Mercado	Artigos de Vestuário	Malhas	Tecidos
Nacional	2111	2161	2151
Comunitário	2112	2162	2152
Extra-comunitário	2113	2163	2153

Tabela 22 - Contas de Clientes por mercados e produtos

Mercado	Serviços e materiais	Imobilizado	Prestação de serviços	Outros
Nacional	2211	26111	2682	2689
Comunitário	2212	26112	-	-
Extra-comunitário	2213	26113	-	-

Tabela 23 - Contas de Fornecedores por mercado e natureza

Em conclusão, um terceiro pode ter várias contas no plano de contas devido a uma das seguintes razões:

- Está implantado em vários países;
- Tem endereços de entrega ou de facturação distintas em mercados ou países distintos, e.g. mercado nacional, externo comunitário, externo não comunitário;
- Tem endereços de entrega ou de facturação distintas no mesmo país e com o mesmo número fiscal;
- Tem endereços de destino noutras entidades, portanto com número fiscal distinto;
- É fornecedor de serviços e materiais, obrigando por isso a ter contas de fornecedor no POC diferentes;
- É simultaneamente cliente e fornecedor, portanto terá contas diferentes na classe 2.

Anexo 2

Casos atípicos nos documentos com reflexos na identificação de terceiros

No sistema legado existe uma situação particular que origina uma prática específica na construção e emissão de documentos:

Factura com endereço e número de contribuinte diferente do cliente.

Quando se expede mercadoria para uma organização (B) que não é o cliente final (A), em especial quando o cliente e o destinatário pertencem a mercados diferentes, e.g. factura um cliente de mercado comunitário – conta 2112, mas a entrega é feita no mercado nacional – conta 2111, o documento de factura apresenta a designação do cliente (A) com o endereço da sede (de A), local de entrega a morada da entidade (B) com o número de contribuinte de (B).

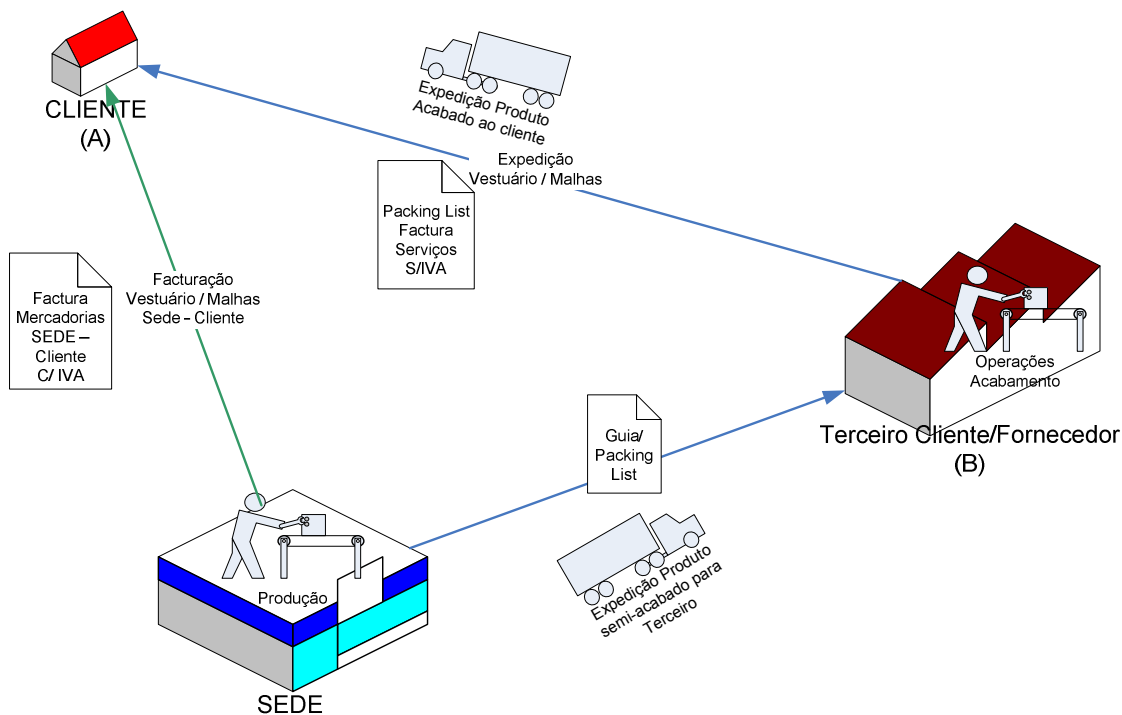


Figura 32 - Expedição e Facturação para destinos diferentes.

Para efeitos de IVA, os documentos devem evidenciar o número de contribuinte do destino da mercadoria, sendo que as contas a serem movimentadas pelas vendas diferem de mercado (neste caso cliente comunitário) das movimentadas em termos de impostos (mercado nacional), i.e. há facturação para mercado comunitário, e impostos – IVA – para mercado nacional. Na prática resulta que o cliente estrangeiro comunitário pagará IVA à organização, ao contrário do habitual nas expedições para o seu país, uma vez que a mercadoria foi entregue num destino nacional.

Na aplicação de facturação do sistema legado, a inscrição do número de contribuinte nacional fará com que sejam calculados e emitidos os valores de IVA. No entanto a identificação e conta de contabilidade do terceiro é uma conta de mercado comunitário, portanto o sistema não contabilizará o IVA. A operadora responsável pelos registos contabilísticos irá proceder a essa contabilização de forma manual.

Por outro lado, existem acordos alfandegários, como o do porto de Algeciras entre Espanha e Marrocos, que fazem considerar os trânsitos por este porto com destino a Portugal trânsitos em mercado comunitário, em vez de mercado extra-comunitário, o que tem implicações no processo administrativo.

Como o sistema legado não suporta estas exceções, encontramos neste sistema registos de terceiros com o número de contribuinte de outro terceiro, que foi destino de mercadoria a cargo do primeiro. No documento de expedição o local de entrega é escrito com recurso a linhas de comentário.

Esta problemática processual descrita está correctamente contemplada nos módulos de facturação do sistema SCM, permitindo indicar que um terceiro do mercado nacional, com o seu número de contribuinte, é um destino de entrega de um terceiro de mercado comunitário. Esta facilidade abrange qualquer mercado e quaisquer terceiros.

Anexo 3

Novo layout organizacional

A organização sofreu uma alteração no modelo das operações envolvendo as empresas do grupo. As operações de encomendas de materiais e subcontratação de serviços de fabricação passaram a ser responsabilidade funcional das empresas satélite orientadas para a fabricação, ficando a sede responsável pela especificação técnica de produto e gestão de grupo, colocação das encomendas de cliente nas empresas fabricantes (Fab 1; Fab 2; Fab 3; Fab 4). Estas últimas procedem à aquisição dos materiais necessários, subcontratação de serviços, fabricação e expedição para o cliente final. As operações sobre o sistema serão realizadas, independentemente da empresa que gera e emite documentos de requisição e facturação, por colaboradores de todo o grupo.

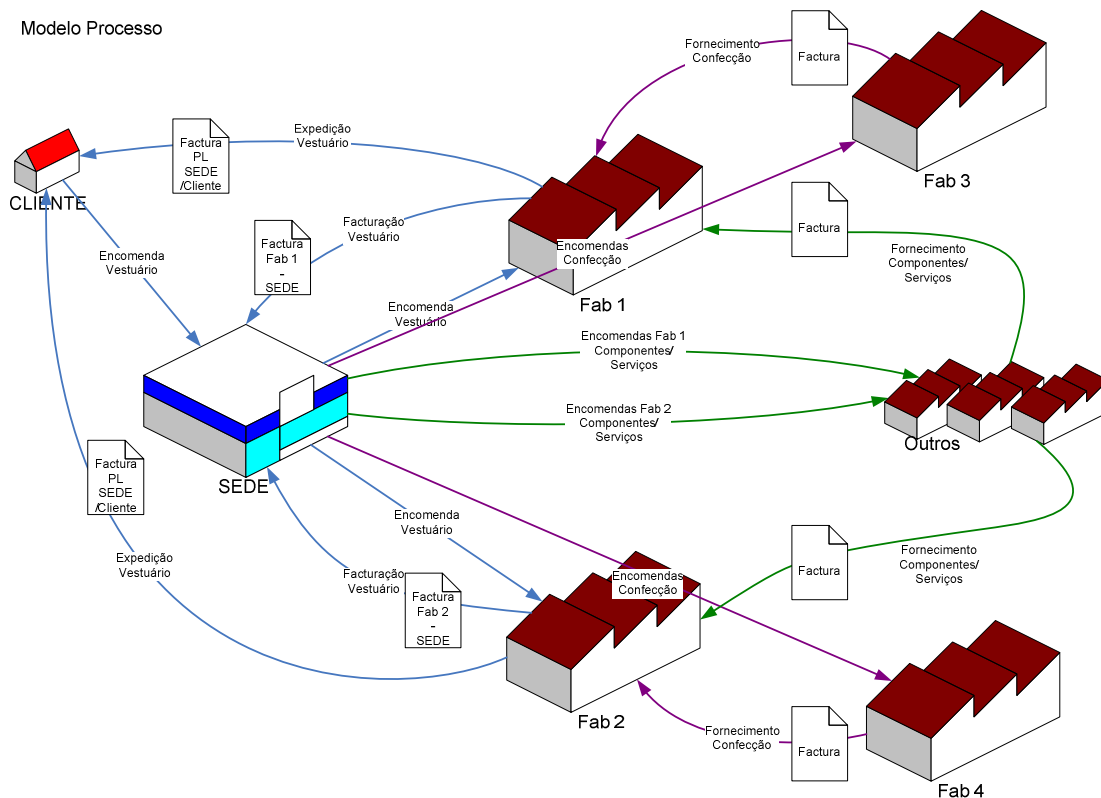


Figura 33 - Novo modelo do processo no grupo

Existirão grupos de funcionários na sede que estudam o produto, colocam as encomendas nas empresas Fab 1 a 4, produzem em nome dessas empresas as encomendas de materiais e serviços. Por outro lado, os funcionários das empresas Fab 1 a 4 recebem mercadorias, fazem a expedição e facturação em nome da sede para o cliente final. Este modo de operação obriga a que se mantenha uma base de dados comum a todas as empresas do grupo.

No plano dos sistemas de informação esta abordagem implica:

- uma base de dados comum partilhada por todas as empresas contendo entre outros:
 - descrição de artigos;
 - materiais;
 - encomendas;
 - terceiros;
- o estudo de produto na sede;
- a adjudicação e colocação das encomendas nas empresas fabricantes do grupo;
- a gestão das encomendas na sede em nome das empresas fabricantes;
- os processos de recepção de materiais e subcontratos nas empresas fabricantes;
- a expedição para o cliente final a partir das empresas fabricantes, em nome da sede;
- um refinamento no sistema de segurança e atribuição de autoridades aos utilizadores orientada ao documento, funcionalidade e organização que opera.

Motivados por esta partilha de informação optou-se por uma mudança na identificação dos terceiros. Uma entidade de terceiro pode ser cliente e fornecedor, e porque a base de dados é a mesma para várias empresas do grupo, abandona-se a sigla CL e FN, para adoptar uma sigla que indique a que organização o terceiro pertence, e.g. “TE” para a sede, “ML” para uma das empresas fabricantes. À custa de alguma semântica nos valores dos registos, esta abordagem permite manter registos discretos por empresa na mesma base de dados, respeitantes a terceiros, facturação, requisições, movimentos de armazéns, etc.

Neste novo paradigma organizacional continuam a ser relevantes e merecer especial atenção os problemas anteriormente referidos:

- a selecção ou geração automática de um número de identificação para a identificação do terceiro único a inserir no sistema SCM;
- a selecção de um dos nomes encontrados no sistema legado;
- correcção da descrição do país e cidade;
- correcção do conteúdo inadequado de alguns campos, como os que compõem o nome, morada e telefone;
- a criação de registos de endereços para as contas criadas por efeito de moradas de envio distintas;
- a identificação de marcas e atribuição nos locais certos no sistema SCM.

Por imposição do SAFT-PT, continuarão a migrar dados das aplicações de facturação, contabilidade e gestão financeira do sistema legado para o sistema SCM (tendo em atenção os novos formalismos de codificação e conceptualização da base de dados) o que se materializa em:

- movimentos de facturação;
- artigos movimentados;
- planos de contas;
- movimentos contabilísticos;
- movimentos de IVA.

Anexo 4

Experiência de Migração Continuada – utilização do Tradutor do Sistema Legado

Neste anexo explora-se a ferramenta desenvolvida para o Tradutor Legado da Figura 29.

Na Figura 34 mostra-se a interface da aplicação que permite migrar do sistema legado para o novo sistema SCM movimentos de facturação, artigos, contas do plano e movimentos contabilísticos, sincronizados com os terceiros.

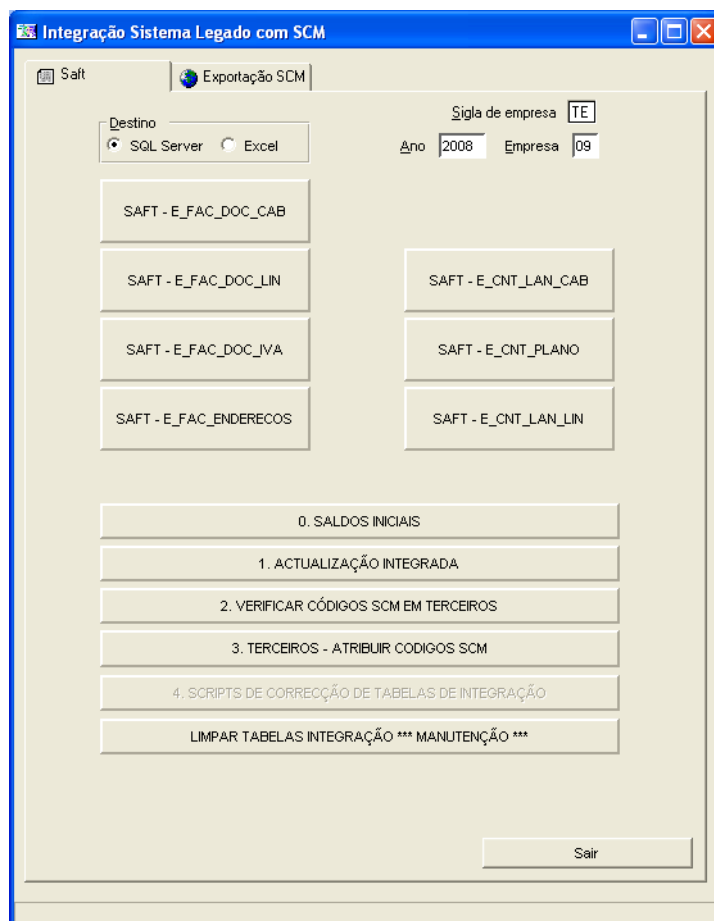


Figura 34 - Ferramenta para Integração do Sistema Legado com SCM

Em particular, no caso dos terceiros, estes migram por via dos movimentos de facturação e lançamentos de contabilidade. Pelo facto de figurarem nestes movimentos, são então inscritos na tabela de migração de terceiros.

De forma semelhante migram-se artigos transaccionados, sejam de vestuário, acessórios, malhas ou outros que tenham sido adquiridos ou vendidos.

Depois dos processos de migração serem executados utiliza-se a área de mapeamento de terceiros, representada na Figura 35.

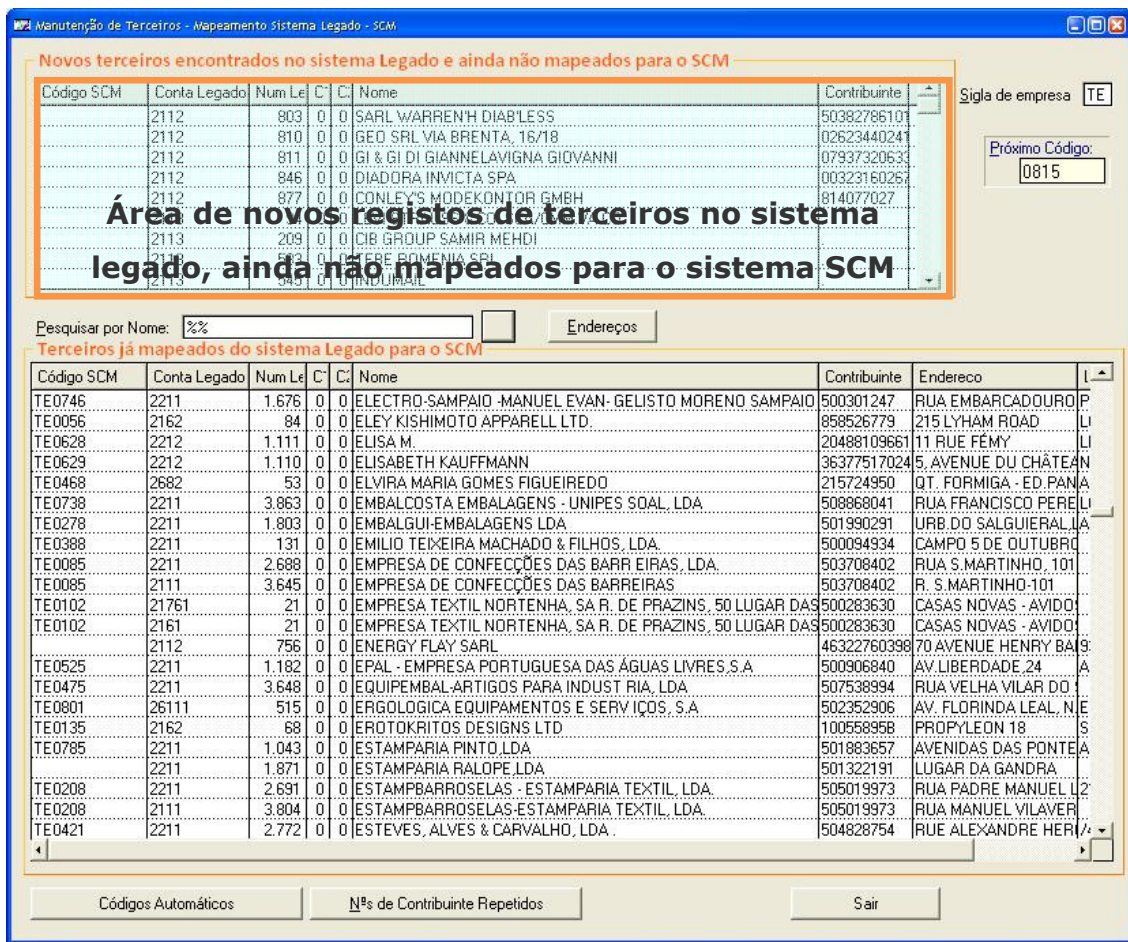


Figura 35 - Interface para a tabela de mapeamento de terceiros entre sistema Legado e SCM

A aplicação carrega automaticamente na tabela de mapeamento de terceiros todos os novos terceiros do sistema legado ainda não representados no SCM, que surgem na área de novos registos de terceiros. O sistema tem um numerador automático, e informa sempre o próximo número de terceiro a atribuir. Existe a opção de numerar automaticamente todos os terceiros ainda sem número atribuído no SCM premindo “Códigos automáticos”.

A interface disponibiliza ordenação da tabela de terceiros por qualquer das colunas apresentadas, e critérios de pesquisa/posicionamento por nome, facilitando a identificação de registos da mesma entidade, pela semelhança do nome ou igualdade de número de contribuinte.

Num registo ainda não mapeado, o operador poderá atribuir o próximo número disponível (mostrado no canto superior direito), ou o mesmo número já atribuído a outro registo da mesma entidade.

Neste exemplo, salientam-se os terceiros “TE0085” e “TE0208” pela diferente escrita do nome do Terceiro.

Na primeira execução do processo de migração, quando todos os registos de terceiros são novos para o sistema SCM, i.e. não têm código SCM, o operador deve utilizar o numerador automático, premindo “Números Automáticos”. O sistema atribui um número de identificação a cada registo ainda sem

número, portanto ficando todos os registos identificados com números distintos. Este procedimento permite um consumo menor de tempo já que só se alterarão os códigos para os casos dos vários registos para uma mesma entidade, alterando assim um menor conjunto de registos.

Em iterações futuras dever-se-á inspeccionar a lista de terceiros ainda com o código não preenchido, facilitando a identificação dos registos novos, e depois de codificados todos os registos novos de terceiros já existentes como entidade no SCM, então acciona-se a numeração automática, codificando assim todos os registos que não encontraram pares associados para o mesmo terceiro.

Portanto, o utilizador inspecciona a tabela de mapeamento de terceiros e quando identifica vários registos do mesmo terceiro no sistema legado, atribui-lhes o mesmo número identificador de terceiro SCM relacionando-os com uma só entidade. Para a identificação de registos do mesmo terceiro contribuem diversas condições, mas poucas são necessárias e suficientes. Destas destacam-se as seguintes:

- registos com o mesmo nome de entidade;
- registos com o mesmo número de contribuinte – critério inútil nos casos de terceiros do mercado extra-comunitário já que a grande maioria não contém valor representativo neste campo;
- registos dos terceiros para os quais se pretende uma conta-corrente única – um critério que depende da experiência do utilizador;
- registos que apresentam o mesmo número de contribuinte, quando este existe.

Nestes casos, o tradutor para o repositório SCM agrega à mesma entidade as contas de contabilidade e endereços ao mesmo terceiro.

Estas condições representam os casos mais comuns, mas encontramos excepções no processo administrativo que dificultam esta associação.

Testes e Experiências

A seguir apresentam-se alguns testes e experiências no processo iterativo de mapeamento de terceiros que ilustram o funcionamento da ferramenta de software desenvolvida, alguns mapeamentos e os melhoramentos realizados ao software.

O processo que se descreve de seguida foi executado iterativamente sobre todos os registos de terceiros não mapeados no SCM num determinado momento, para terceiros movimentados em 2008 e 2009. A identificação dos registos da mesma entidade foi realizada com base na semelhança do nome, igualdade do número de contribuinte e conhecimento do operador sobre o negócio.

No exemplo da Figura 36, os dois registos de “J.M.CACHINA, LDA” e “J.M.CACHINA-JOSE MANUEL CACHINA DE MORAIS”, apesar da semelhança de nomes, representam entidades distintas, e apresentam números de contribuinte diferentes.

Os dois registos “JACADI, S.A” e “JACADI S.A. SERVICE CONPTABILITE” apesar do grau de semelhança equivalente ao exemplo anterior, são mapeados para a mesma entidade no sistema SCM, “TE0387”. Apresentam o mesmo número de contribuinte, e moradas distintas. Darão origem a dois registos de contas, de classes diferentes (2212 e 2112) e a dois registos de endereço.

Os registos de Código SCM “TE692” são de semelhança total no nome e número de contribuinte e gerarão um registo de conta e dois de endereço.

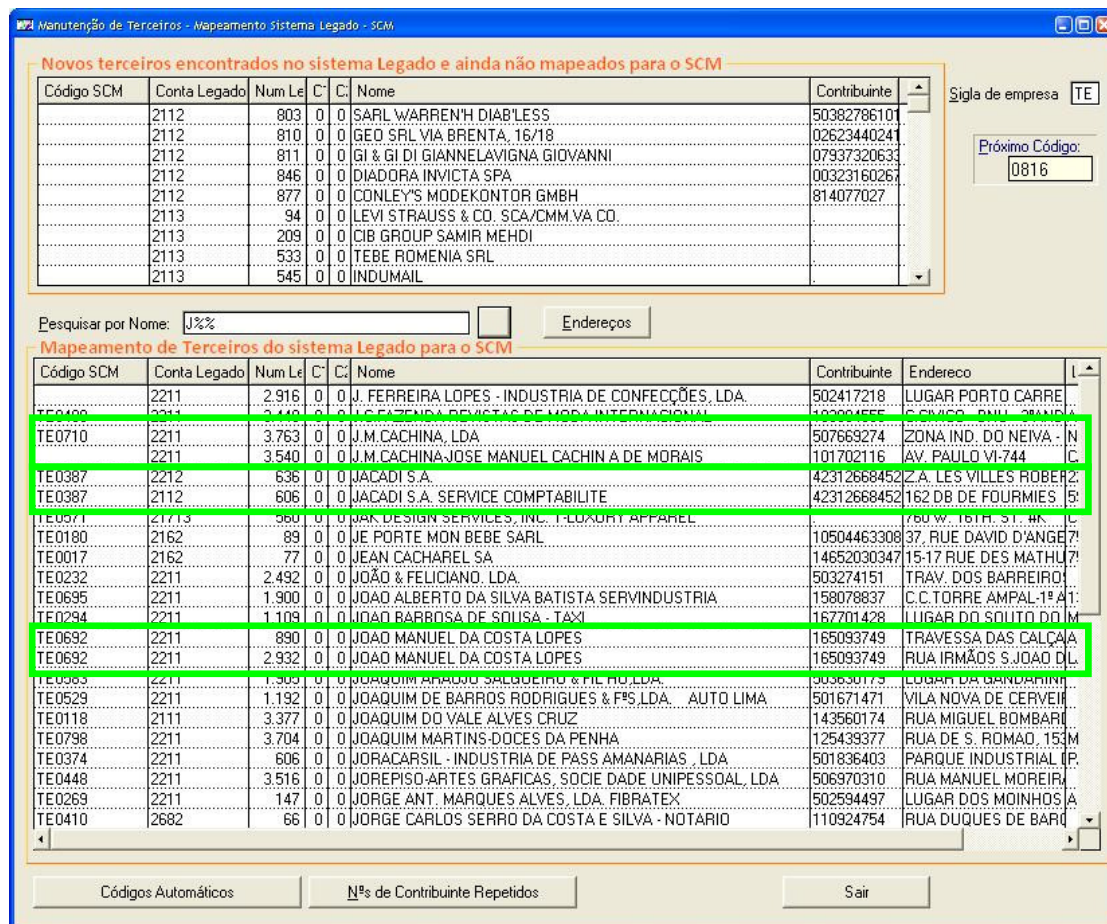


Figura 36 - Identificação de mapeamentos de terceiros

Neste exemplo demonstra-se que uma identificação de semelhança automatizada poderá falhar, dada a dualidade das opções nos casos semelhantes “JACADI” e “J.M. CACHADA”.

O caso apresentado a vermelho na Figura 37 é um exemplo do descrito no Anexo 2, em que existiram mercadorias vendidas ao cliente francês “ITO LAB SARL”, mas que foram entregues na empresa “O C M – INDÚSTRIA DE CONFECÇÕES, LDA.” em Portugal, provavelmente para trabalhos adicionais na mercadoria.

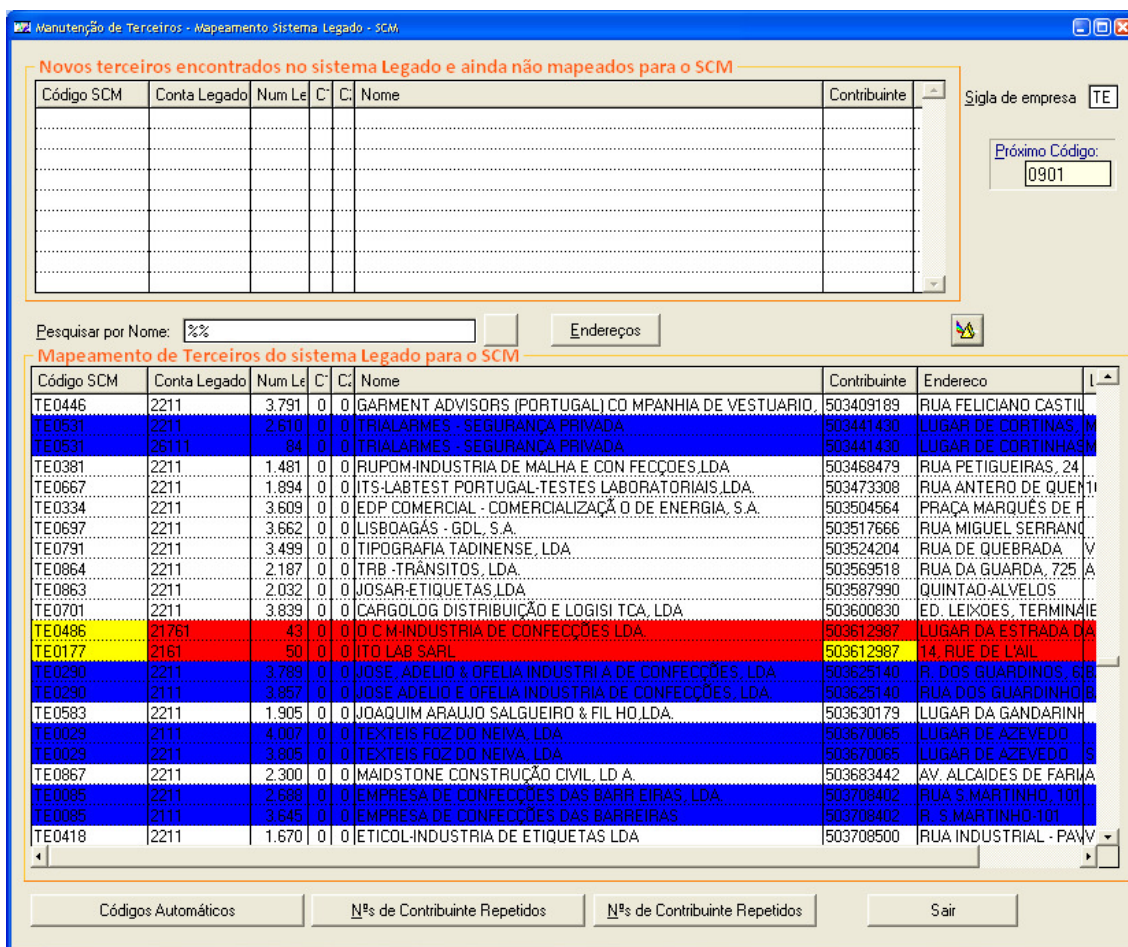


Figura 37 - Número de contribuinte igual para clientes diferentes (ITO e O C M)

Repetição de Contribuinte: 503612987
 TE0486 / 21761 / 43 / O C M-INDUSTRIA DE CONFECÇÕES LDA.
 TE0177 / 2161 / 50 / ITO LAB SARL

Figura 38 - Duplicação de um NIF em entidade diferentes

Por imposição legal, no documento têm que figurar a designação e morada do cliente francês (ITO), o local de entrega (O C M) com nome e morada, assim como o número de contribuinte do destinatário, O C M. Por limitações da aplicação no sistema legado, o operador inseriu o número de contribuinte de O C M na ficha do cliente ITO para produzir cálculo de IVA no documento e surgir no local previsto na impressão. No âmbito dos casos atípicos descrito no Anexo 2, aquando da transacção, o operador da área contabilística teve que manualmente proceder a movimentações de IVA, enquanto o cliente francês aceitou liquidar esse montante fiscal.

A cor vermelha generalizada nas duas linhas, com o amarelo no código SCM e no Contribuinte, indicam ao operador existe o mesmo número de contribuinte para entidades diferentes. Trata-se portanto de uma excepção que requer a análise do operador.

Note-se que o registo ITO (Cliente Francês), Figura 40, está com o número de contribuinte de O C M (Cliente Português) porque houve uma expedição, a ser paga por ITO, entregue nas instalações em Portugal da O C M.

Novos terceiros encontrados no sistema Legado e ainda não mapeados para o SCM

Código SCM	Conta Legado	Num Le	C	C	Nome	Contribuinte

Buscar por Nome: Endereços

Mapeamento de Terceiros do sistema Legado para o SCM

Código SCM	Conta Legado	Num Le	C	C	Nome	Contribuinte	Endereço
TE0469	2211	2.870	0	0	IPORTAL SERVIÇOS DE INTERNET E REDES, LDA	504837958	RUA SANTOS POUSSAD...
TE0594	2211	858	0	0	IRMAOS DIAS, LDA. - FABRICA DE TECIDOS DE SEDA	500141401	APARTADO 36 - GANDR...
TE0183	2161	34	0	0	IRMÃOS RODRIGUES CONFECÇÕES SA	501852433	LUGAR DO MONTE APTG...
TE0320	2211	3.792	0	0	IRMÃOS VILA NOVA II - INDÚSTRIA DE CONFECÇÕES, S.A.	501075682	RUA 5 DE OUTUBRO, NA...
TE0709	2211	3.571	0	0	ISABEL ALEXANDRA MENDES MAND	201188678	RUA DO QUEIMADO NºV...
TE0139	2111	4.014	0	0	ISABEL MARIA RIBEIRO PONTES	231150717	LUGAR PENEDA
TE0139	2211	3.749	0	0	ISABEL MARIA RIBEIRO PONTES	231150717	LUGAR DA PENEDA
TE0241	2211	3.504	0	0	ISAURA MARIA F. PERES FILIPE	156111632	LUGAR DA RUA
TE0522	2211	1.443	0	0	ISMAEL ANTUNES DE SOUSA, LDA EQUIPAMENTOS INDUSTRIAIS	501279040	RUA JOAQUIM DIAS SAV...
TE0528	2213	392	0	0	ISTANBULETIKET TEKSTIL MATBAA SAN TIC LTD STI	0	BASIN EKSPRES YOLU Y...
TE0730	26111	208	0	0	ISTC-INFORMATICA SISTEMAS & TELECOMUNICACOES LDA	501908439	RUA DANIEL SANTOS, 2E...
TE0730	2211	1.473	0	0	ISTC-INFORMATICA SISTEMAS & TELECOMUNICACOES LDA	501908439	RUA DANIEL SANTOS, 2E...
TE0532	2211	2.366	0	0	ITALCO-MODA ITALIANA, LDA	502608285	PRAÇA DUQUE DE SAL...
TE0177	2161	50	0	0	ITO LAB SARL	503612987	14, RUE DE L'AIL
TE0177	2162	88	0	0	ITO LAB SARL	21445340292	14, RUE DE L'AIL
TE0667	2211	1.894	0	0	ITS-LABTEST PORTUGAL-TESTES LABORATORIAIS, LDA	503473308	RUA ANTERO DE QUEN11...
TE0875	2211	2.916	0	0	J. FERREIRA LOPES - INDUSTRIA DE CONFECÇÕES, LDA	502417218	LUGAR PORTO CARRE
TE0489	2211	3.448	0	0	J.C.FAZENDA REVISTAS DE MODA INTERNACIONAL	103094555	C.CIVICO - BNU - 3ªANDA...
TE0710	2211	3.763	0	0	J.M.CACHINA, LDA	507669274	ZONA IND. DO NEIVA - N...
TE0879	2211	3.540	0	0	J.M.CACHINA-JOSE MANUEL CACHINA DE MORAIS	101702116	AV. PAULO VI-744
TE0387	2212	636	0	0	JACADI S.A.	42312668452	Z.A. LES VILLES ROBER2...
TE0387	2112	606	0	0	JACADI S.A. SERVICE COMPTABILITE	42312668452	162 DB DE FOURMIES 15...

Códigos Automáticos N.ºs de Contribuinte Repetidos N.ºs de Contribuinte Repetidos e Cód diferente Sair

Figura 40 - Registos do terceiro ITO com números de contribuinte diferentes

Assim o número de contribuinte para o terceiro ITO foi mudado nesta conta para o nº de O C M porque a lei obriga a que o número de contribuinte seja o do endereço de expedição. O cliente Francês pagou o IVA devido.

O registo 2162 88 tem o número de contribuinte francês correcto.

No sistema SCM deveremos ter uma só entidade com o número de contribuinte francês, sendo que existirá um destino de envio neste terceiro com os dados do terceiro O C M.

Este cenário estende a conceptualização dos mapeamentos descritos na Figura 13 - Associação ou Mapeamento de Terceiros do Sistema Legado para o SCM (pp.48).

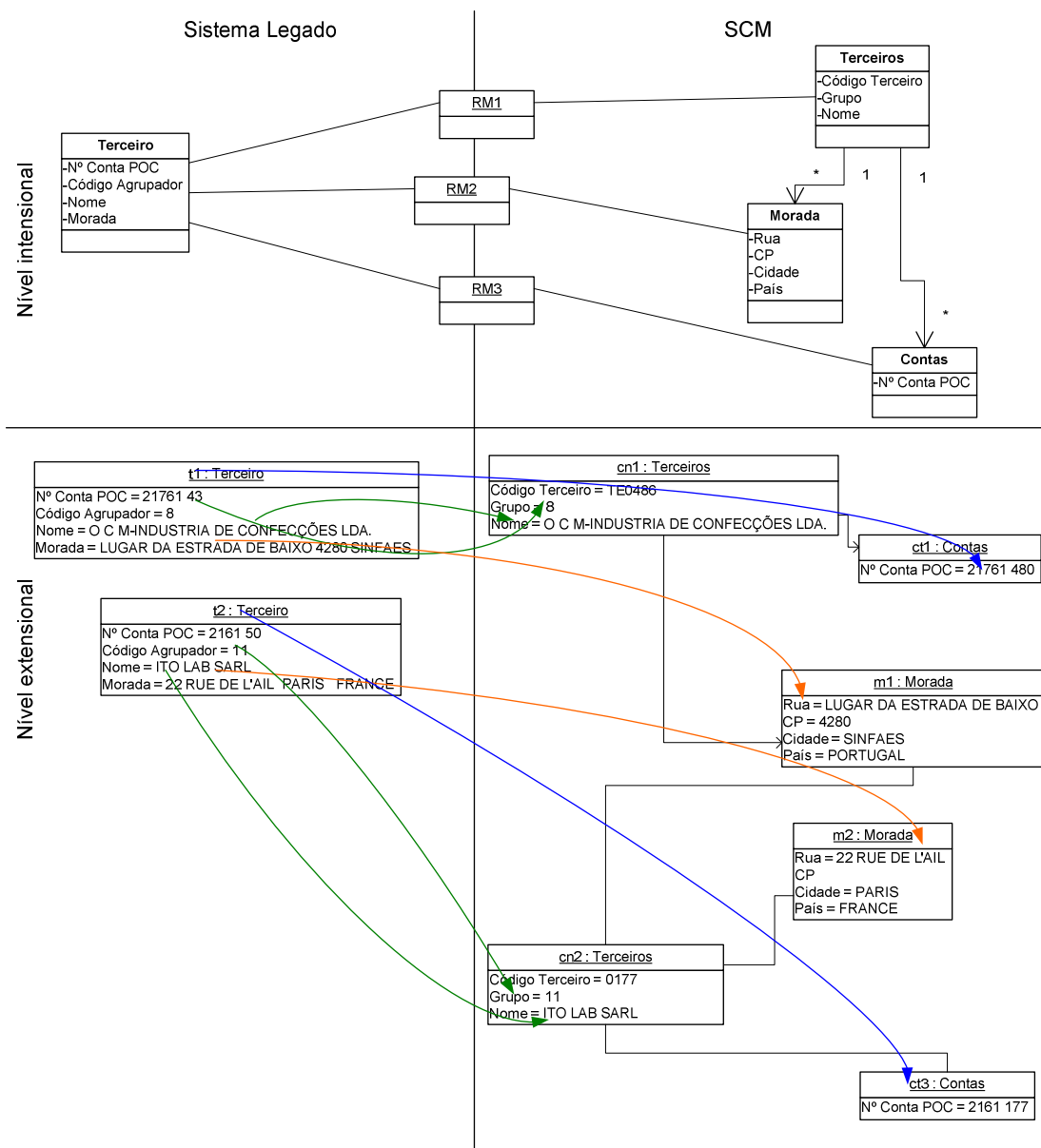


Figura 41 - Extensão a Mapeamento de terceiros do Sistema Legado para o SCM

O endereço “m1” do terceiro “cn1”, “O C M” é um endereço de destino para o terceiro “cn2” “ITO”, contudo este endereço “m1” não está referenciado no sistema legado para “t2”, uma vez que esta informação é extraída de um documento de faturação.

Refinamentos

Durante a utilização da aplicação surgiu a necessidade de desenvolver pequenos refinamentos na aplicação, no sentido de melhor suportar a tarefa de interação humana no processo. As correcções e melhorias desenvolvidas foram motivadas por dificuldades na utilização da interface e por valores não previstos encontrados nos registos em migração.

Na Figura 43 mostram-se os mesmos registos ordenados por Código SCM, ressaltando as diferenças de semântica no Nome de Terceiro para a mesma entidade, assim como o NIF diferente.

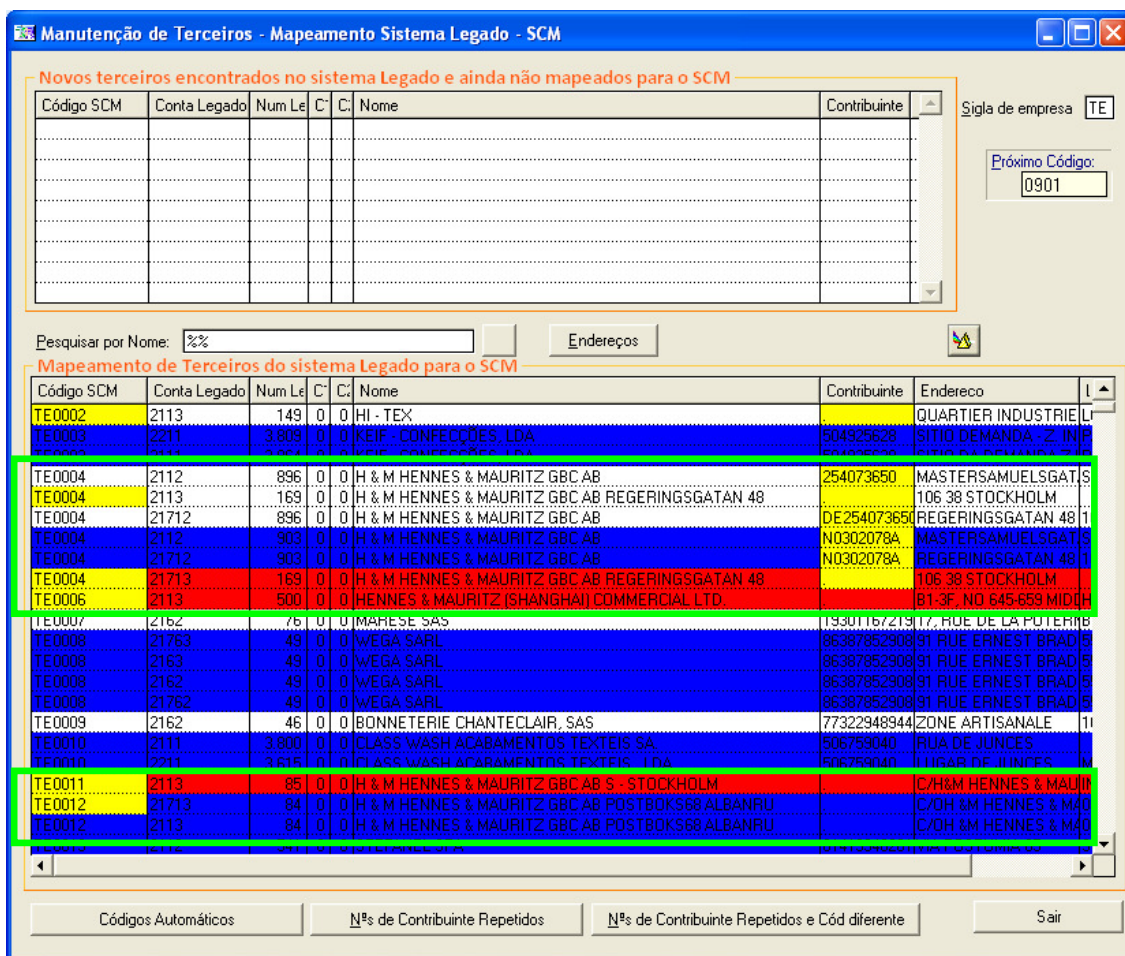


Figura 43 - Mapeamento de Terceiros ordenado por Código SCM

Desenvolvendo a facilidade de ordenação por todas as colunas e a codificação de cores, permite-se um trabalho de identificação/associação mais eficaz.

Na lista ordenada por nome, as cores indicam para o cliente “H&M” diferenças de valores em Contribuinte e Código SCM. Na Tabela 25 podemos comparar as associações realizadas pelo operador que em alguns casos não coincide com a associação prevista para o mesmo Grupo, ou Código Agrupador.

CodSCM	Conta	Num	Grupo	Nome
TE0004	2112	903	21	H & M HENNES & MAURITZ GBC AB
TE0004	2113	169	11	H & M HENNES & MAURITZ GBC AB REGERINGSGATAN 48
TE0004	21712	896	21	H & M HENNES & MAURITZ GBC AB
TE0004	21712	903	21	H & M HENNES & MAURITZ GBC AB
TE0004	21713	169	11	H & M HENNES & MAURITZ GBC AB REGERINGSGATAN 48
TE0004	2112	896	21	H & M HENNES & MAURITZ GBC AB
TE0006	2113	500	25	HENNES & MAURITZ (SHANGHAI) COMMERCIAL LTD.
TE0011	2113	85	15	H & M HENNES & MAURITZ GBC AB S - STOCKHOLM
TE0012	2113	84	15	H & M HENNES & MAURITZ GBC AB POSTBOKS68 ALBANRU
TE0012	21713	84	15	H & M HENNES & MAURITZ GBC AB POSTBOKS68 ALBANRU

Tabela 25 - Associação de Terceiros e Código Agrupador

Constata-se que o operador não identificou os mesmos terceiros pelo Grupo, como esperado e tomado por pressuposto no trabalho descrito na secção 4.4, o que revela problemas na natureza dos dados relacionados com actualização ou conceptualização, i.e. ou o código agrupador não está actualizado ou não representa de facto um unificador de terceiro na conceptualização aceite nesta integração.

Nesta amostra verifica-se:

- As mesmas dificuldades semânticas em identificar registos do mesmo terceiro, nas instâncias do sistema legado;
- O código Agrupador não pode funcionar como chave de terceiro, conforme se esperava, e conforme demonstrado.

Assim, no final do primeiro processo de migração com a aplicação teve-se a preocupação de procurar casos de registos que foram associados ou não associados de forma incorrecta, nomeadamente entidades diferentes com o mesmo número de contribuinte.

Como os registos foram acedidos por ordem ascendente do nome, existe a hipótese de se terem gerado casos de terceiros com identificação diferente que partilham o mesmo número de contribuinte.

No sentido de verificar a existência destes casos pesquisou-se a tabela de mapeamento contando o número de entidades diferentes por número de contribuinte com o script SQL:

```
select res.[FEnt_N_Cnt], count(res.[FEnt_N_Cnt] )
from
(SELECT distinct
[FEnt_Codigo_Macwin] CODSCM,
[FEnt_N_Cnt],
(select count(*) from [GMIC_TEBE].[MACWIN].[I_Fac_Entidades] e2
where e2.[FEnt_Codigo_Macwin]=e1.[FEnt_Codigo_Macwin] and
e2.[FEnt_N_Cnt]=e1.[FEnt_N_Cnt]) as nterceiros
FROM [GMIC_TEBE].[MACWIN].[I_Fac_Entidades] e1
--where fent_n_cnt='0' or fent_n_cnt='.'
)res
group by res.[FEnt_N_Cnt]
```

Figura 44 - Script SQL que detecta Entidades diferentes com mesmo NIF

Obteve-se uma lista com o número de repetições de números de contribuinte (NIF) para diferentes códigos de entidade, como se representa num excerto na Tabela 26:

NIF	Número de Entidades
506911314	1
506927288	2
506931161	1
506944484	1

Tabela 26 - Verificação da duplicação de NIF após mapeamento assistido

Este resultado indica que estes casos existem.

Uma actualização na ferramenta de mapeamento permitiu obter uma lista dos casos identificados:

Repetição de n.ºs de contribuinte por terceiros com id diferente.
 Repetição de Contribuinte: 500052999
 TE0637 / 2211 / 3752 / POSTO BP SANTAREM NASCENTE GEST24 SOC.UNIP.LDA
 TE0719 / 2211 / 1828 / A.S. BP MEALHADA-CANTANHEDE CARDOL LDA.

Repetição de Contribuinte: 500684278
 TE0704 / 2211 / 3021 / FERNANDO VALENTE & CA.SA
 TE0740 / 2211 / 3538 / N UTILIZAR FERNANDO VALENTE & C^a,S.A.

Repetição de Contribuinte: 500792887
 TE0632 / 2689 / 51 / I.N.C.M., S.A.
 TE0315 / 2211 / 1370 / INCM IMPRENSA NACIONAL - CASA DA MOEDA

Repetição de Contribuinte: 503612987
 TE0486 / 21761 / 43 / O C M-INDUSTRIA DE CONFECÇÕES LDA.
 TE0177 / 2161 / 50 / ITO LAB SARL

Repetição de Contribuinte: 503625140
 TE0758 / 2111 / 3857 / JOSE ADELIO E OFELIA INDUSTRIA DE CONFECÇÕES, LDA.
 TE0290 / 2211 / 3789 / JOSE, ADELIO & OFELIA INDUSTRIA DE CONFECÇÕES, LDA

Repetição de Contribuinte: 503930253
 TE0223 / 2211 / 2479 / VOBIS INFOFIELD-INFORMAT.,S.A.
 TE0519 / 26111 / 474 / INFOFIELD-INFORMATICA,S.A

Repetição de Contribuinte: 504160451
 TE0534 / 2211 / 3815 / CONFECÇÕES DE MODESTO PEREIRA DA SILVA, LDA
 TE0058 / 2111 / 3872 / MODESTO PEREIRA DA SILVA, LDA.

Repetição de Contribuinte: 506927288
 TE0670 / 2211 / 3765 / STAR II - SISTEMAS DE SEGURANÇ A, LDA
 TE0802 / 26111 / 516 / C.S.P.-COM E INST DE SIST. DE ALARM,ELECTRO E TELEVIGIL,LDA

Figura 45 - Lista de Terceiros com o mesmo NIF

A irregularidade aqui descrita aconteceu porque o operador estava focado em nomes semelhantes, e não em números de Contribuinte quando atribuía identificação a terceiros para o SCM.

Confirmando-se que se trata da mesma entidade, para corrigir estes casos o operador usa o identificador de terceiro do que apresenta movimentos mais recentes.

Atribuindo a identificação de entidade correcta, i.e. um só identificador no SCM para os dois registos do sistema legado da mesma entidade, o componente Tradutor SCM (Figura 29, pp. 83) irá identificar o segundo registo como sendo um endereço adicional do terceiro.

Após algumas correcções obtém-se a lista (Tabela 27):

NIF	Número de Entidades
.	75
0	16
503612987	2
506927288	2

Tabela 27 - Entidades diferentes com o mesmo NIF, após primeira verificação

Os casos mostrados com número de Contribuinte igual a “.” ou a “0” não são representativos, já que se trata de registos de terceiros do mercado extra-comunitário, que não exigem número fiscal.

Num universo de 11.566 registos de terceiros foram migrados 923 (7,98%) porque registaram transacções no período em integração, e destes, foram encontrados 16 registos com NIF partilhado – 8 registos com NIF incorrecto.

Conforme descrito na Tabela 28, os registos migrados representam 758 Terceiros no sistema SCM, onde se destacam um Terceiro com 6 registos na tabela de mapeamento e outro com 5 registos; 6 Terceiros com 4; e os restantes com 1 a 3 registos.

Nº de registos de Terceiros no Sistema Legado	Nº de Terceiros no Sistema SCM
1	619
2	124
3	7
4	6
5	1
6	1
	758

Tabela 28 - Número de registos de terceiro mapeados o sistema SCM

É importante verificar estas ocorrências porque neste modelo de integração em produção existe a possibilidade de vir a integrar registos de terceiros do sistema legado, que não apresentavam movimentos aquando do início do processo de integração, e entretanto apresentam actividade. E.g. um registo de uma conta de um terceiro inactivo desde 2002 se for movimentado em 2009 aparece como um novo terceiro na tabela de mapeamento.