



Universidade de Évora
Departamento de Gestão

Mestrado em Gestão

Área de Especialização em Organização e Sistemas de Informação

**AVALIAR A QUALIDADE DOS DADOS
EM SISTEMAS DE INFORMAÇÃO**

Paula Cristiana Leal Rodrigues Coelho Inês
Orientador: Professor Doutor Paulo Fernando Lopes Resende da Silva

Évora, 2010



Universidade de Évora
Departamento de Gestão

Mestrado em Gestão

Área de Especialização em Organização e Sistemas de Informação

**AVALIAR A QUALIDADE DOS DADOS
EM SISTEMAS DE INFORMAÇÃO**



186288

Paula Cristiana Leal Rodrigues Coelho Inês
Orientador: Professor Doutor Paulo Fernando Lopes Resende da Silva

Évora, 2010

Agradecimentos

Ao meu Orientador, Professor Doutor Paulo Resende da Silva, que tem a capacidade extraordinária de tornar tão simples aquilo que por vezes para mim é tão complicado, pelo dinamismo e profissionalismo demonstrado ao longo deste percurso, transmitindo-me confiança numa convivência agradável e salutar.

À minha família, em especial aos meus pais pelo seu apoio incondicional, ao meu esposo e companheiro pela sua paciência e força e aos meus filhos pela sua resignação de nem sempre contarem com a minha presença nas idas ao parque.

Ao organismo da Administração Pública onde exerço funções, especialmente à Dr^a Ana Margarida Fernandes por me ter proporcionado condições de trabalho sem as quais teria sido muito mais difícil concluir a presente investigação.

A todos os que contribuíram directa ou indirectamente para que este trabalho pudesse ser realizado.

Resumo

Avaliar a qualidade dos dados em sistemas de informação

A informação, como fonte do conhecimento, representa nos dias de hoje um recurso precioso dentro da organização e, como tal, deve ser encarada como um produto que deve ser definido, medido, analisado e melhorado. Por sua vez, o sistema de informação pode ser considerado como um conjunto gerido de recursos humanos e materiais, destinado a realizar as actividades de adquirir, armazenar, processar e difundir informação. Nesta medida, é essencial que o sistema de informação consiga tratar e produzir informação com qualidade, para que esta possa contribuir de forma eficaz para o sucesso organizacional. Nesta dissertação é efectuada uma revisão dos modelos existentes de qualidade dos dados e é proposto um modelo de avaliação da qualidade dos dados, apontando uma metodologia que poderá integrá-lo e ser utilizada em auditoria de sistemas de informação. O modelo de avaliação encontra-se assente nas quatro dimensões de qualidade dos dados: intrínseca, representação, contextual e acessibilidade.

Abstract

Assessing data quality in information systems

Information, as a source of knowledge, represents today a precious resource within the organization and, as such, should be faced like a product that must be defined, measured, analyzed and improved. In turn, the information system can be considered as a managed set of human and material resources, to perform the activities of acquiring, storing, prosecuting and spreading information. In this measure, it is essential that the information system is able to treat and produce information with quality, so that this one can contribute in the efficient form to the organizational success. This dissertation conducted a revision of the existent models of quality of the data and proposed a model for evaluating the quality of the data, pointing to a methodology that will be able to integrate it and to be used in auditing of information systems. The evaluation model is established in four quality dimensions of the data: intrinsic, representation, contextual and accessibility.

ÍNDICE

| | |
|--|-------------|
| ÍNDICE DE FIGURAS | viii |
| ÍNDICE DE QUADROS | ix |
| LISTA DE SIGLAS..... | x |
| INTRODUÇÃO | 1 |
| CAPITULO 1 – A qualidade do sistema de informação..... | 4 |
| 1.1. A INFORMAÇÃO E O SISTEMA DE INFORMAÇÃO NO CONTEXTO ORGANIZACIONAL | 4 |
| 1.2. O VALOR E A QUALIDADE DE UM SISTEMA DE INFORMAÇÃO | 8 |
| CAPITULO 2 – A qualidade da informação | 12 |
| 2.1. MODELOS DE QUALIDADE DA INFORMAÇÃO | 15 |
| 2.1.1. Modelos orientados para os processos..... | 17 |
| 2.1.2. Modelos orientados para o sistema..... | 19 |
| 2.1.3. Modelos orientados para os utilizadores..... | 23 |
| 2.2. ESTUDO COMPARATIVO DOS MODELOS..... | 32 |
| CAPITULO 3 – Desenho metodológico do projecto | 36 |
| CAPITULO 4 – Modelo de avaliação de um sistema de informação | 41 |
| 4.1. O PROCESSO DE AUDITORIA E AS METODOLOGIAS DE AVALIAÇÃO DA QUALIDADE DA INFORMAÇÃO | 41 |
| 4.2. DESENHO DO MODELO DE AVALIAÇÃO..... | 45 |
| 4.3. DISCUSSÃO DO MODELO..... | 56 |
| 4.4. APLICAÇÃO REAL | 59 |
| CONCLUSÕES E TRABALHO FUTURO | 64 |
| BIBLIOGRAFIA | 67 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| FIGURA 1.1 - OS NÍVEIS HIERÁRQUICOS DA INFORMAÇÃO | 5 |
| FIGURA 1.2 - CICLO DE VIDA DA INFORMAÇÃO | 7 |
| FIGURA 1.3 - EQUAÇÃO DE VALOR DE UM SISTEMA DE INFORMAÇÃO | 8 |
| FIGURA 1.4 - MODELO DE SUCESSO DE UM SISTEMA DE INFORMAÇÃO | 10 |
| FIGURA 2.1 – MODELOS TEÓRICOS DE QUALIDADE DA INFORMAÇÃO..... | 16 |
| FIGURA 2.2 - HIERARQUIA DE DIMENSÕES DE QUALIDADE DOS DADOS..... | 19 |
| FIGURA 2.3 - REPRESENTAÇÕES DO SISTEMA REAL DE WAND E WANG (1996) | 20 |
| FIGURA 2.4 - CATEGORIAS E DIMENSÕES DE QUALIDADE DA INFORMAÇÃO | 24 |
| FIGURA 2.5 - A PERCEÇÃO DOS UTILIZADORES SOBRE AS CATEGORIAS DE QUALIDADE DOS DADOS..... | 28 |
| FIGURA 2.6 - HIERARQUIA DE DIMENSÕES DE QUALIDADE DOS DADOS..... | 32 |
| FIGURA 3.1 – MÉTODO DE PESQUISA DE CASO ESTRUTURADO | 39 |
| FIGURA 4.1 – METODOLOGIA DQA - MATRIZ DE CLASSIFICAÇÃO DA QUALIDADE | 45 |
| FIGURA 4.2 - INTEGRAÇÃO DOS MODELOS DE QUALIDADE DA INFORMAÇÃO COM O MODELO DE SUCESSO DE UM SISTEMA DE INFORMAÇÃO DE DELONE E MCLEAN (1992) | 46 |
| FIGURA 4.3 – MODELO DE AVALIAÇÃO DE UM SISTEMA DE INFORMAÇÃO..... | 47 |
| FIGURA 4.4 - ESPECIFICAÇÃO DO MODELO DE QUALIDADE DOS DADOS | 57 |
| FIGURA 4.5 - RESULTADOS OBTIDOS NA AVALIAÇÃO - EXEMPLO TEÓRICO | 60 |

ÍNDICE DE QUADROS

| | |
|--|-----------|
| QUADRO 2.1 – FORMAS DE CARACTERIZAÇÃO DAS METODOLOGIAS DE AVALIAÇÃO DA QUALIDADE DA INFORMAÇÃO..... | 13 |
| QUADRO 2.2 – COMPARAÇÃO DOS MODELOS DE QUALIDADE DOS DADOS | 33 |
| QUADRO 4.1 - ESPECIFICAÇÕES DAS METODOLOGIAS DE AUDITORIA | 42 |
| QUADRO 4.2 - TAXONOMIA DOS PROBLEMAS DE QUALIDADE DOS DADOS | 54 |
| QUADRO 4.3 – RESULTADOS OBTIDOS NA AVALIAÇÃO – EXEMPLO TEÓRICO..... | 61 |

LISTA DE SIGLAS

| | |
|---------------------------------|---|
| AI ₁ RI ₂ | Acessibilidade, Interpretabilidade, Relevância e Integridade |
| AIQM | A methodology for Information Quality Assessment |
| AMEQ | Activity -based Measuring and Evaluating of product information quality |
| CDQ | Comprehensive methodology for data quality management |
| CIHI | Canadian Institute for Health Information Methodology |
| COLDQ | Cost Effect Of Low Data Quality |
| DQA | Data Quality Assessment |
| DWQ | The Data Warehouse Quality Methodology |
| IQM | Information Quality Measurement |
| QAFD | Quality Assessment of Financial Data |
| SI | Sistema de Informação |
| TDQM | Total Data Quality Management |
| TIQM | Total Information Quality Management |

INTRODUÇÃO

Nas organizações actuais a informação assume uma importância primordial e crescente. Assiste-se a uma nova era de evolução organizacional, cuja ênfase é a gestão do conhecimento como chave para os novos modelos de estratégia organizacional.

Hoje gere-se o conhecimento, das pessoas e das organizações. Um novo conceito de estratégia organizacional emergiu e assiste-se assim ao crescimento das organizações aprendentes. Estas aprendem com a interacção com o meio envolvente, emergem “no terreno”, baseadas em *feedbacks* constantes dos sectores operacionais que interagem com a envolvente externa da organização. O nível estratégico desce aos níveis da gestão operacional, observando e escutando as camadas operacionais, aprendendo com estas através da análise e avaliação dos seus procedimentos organizacionais, flexibilizando os canais de comunicação e fazendo com que a informação flua por toda a organização. Ao interagir com estes níveis, mais facilmente a governação estratégica passa a sua mensagem e faz com que toda a organização conheça e comungue dos objectivos da mesma.

Não é por isso de estranhar a grande importância que tem a informação e a proliferação dos sistemas de informação no seio organizacional. Estes permitem que a informação flua de forma mais eficaz em toda a organização e, conseqüentemente, facilitam os processos de conhecimento organizacional e a aprendizagem organizacional, constituindo-se um dos recursos mais preciosos.

Os processos de auditoria visam garantir que os procedimentos adoptados nos níveis operacionais da organização permitem cumprir os objectivos organizacionais de forma eficiente e eficaz, contribuindo assim para o sucesso organizacional. Os sistemas de informação são hoje importantes ferramentas no seio da organização e deve por isso haver uma garantia de que estes cumprem todas as funções para as quais foram concebidos e que, simultaneamente, conseguem garantir informação de qualidade para os níveis de decisão.

É importante por isso definir o que é efectivamente um sistema de informação de qualidade e as características que este deve conter para que a informação produzida seja

de qualidade. Simultaneamente, para se poder auditar um sistema de informação é necessário não só conhecer o que significa qualidade em sistemas de informação como também dispor de ferramentas que permitam avaliar esta qualidade.

Apesar de já muito se ter falado sobre qualidade da informação, não existe um consenso sobre as características que a informação deve apresentar para ser considerada informação com qualidade.

Sabe-se que são cada vez maiores os investimentos das organizações em sistemas de informação. Contudo, muitas vezes estes investimentos não têm o retorno esperado, dada a má qualidade do sistema em que se investiu. Podem ser múltiplas as causas para este fenómeno, mas não raras vezes assiste-se a problemas relacionados com os dados. No entanto, será que se pode concluir que um sistema de informação tem qualidade, que garante os padrões e características de qualidade e a fiabilidade dos processos de negócio, avaliando a qualidade dos seus dados? Por isso a grande questão que se coloca é: *será possível avaliar a qualidade de um sistema de informação pela avaliação da qualidade dos seus dados?*

O sistema de informação é fortemente influenciado pela qualidade dos dados. Se o sistema de informação for capaz de garantir a qualidade dos dados imediatamente na fase de input, terá maiores probabilidades de produzir informações com qualidade. Os processos implementados num sistema de informação apenas poderão reflectir correctamente as operações e o negócio da organização se utilizarem dados com características de qualidade e assim contribuirão de forma eficaz para o desenvolvimento da organização e para a tomada de decisões.

É possível estabelecer padrões de qualidade que podem ser utilizados na avaliação de qualquer sistema. Apesar de não existir um consenso relativamente aos critérios que definem a qualidade, existem alguns como a fiabilidade, a integridade, a consistência, a actualidade e a disponibilidade, que são comumente citados por vários autores e que podem ser traduzidos em programas de trabalho a utilizar em auditoria informática.

Assim, os objectivos deste trabalho consistem em identificar o significado de qualidade dos dados em sistemas de informação, perceber a influência da qualidade dos dados nos órgãos de decisão e conseqüentemente nas políticas da organização, determinar o modo como a qualidade dos dados pode influenciar o desempenho de um sistema de informação, designadamente a fiabilidade dos seus processos de negócio e, finalmente,

investigar a aplicação de uma *framework* de avaliação da qualidade dos dados em processos de auditoria informática.

Para poder responder à questão e aos objectivos da investigação, ao longo deste trabalho será analisado um conjunto de documentos de autores que se destacaram na abordagem da temática em estudo. A partir desta análise crítica serão propostos níveis e critérios de qualidade e elaborado um modelo de avaliação da qualidade dos dados com base naqueles, modelo este que poderá ser aplicado em auditoria de sistemas de informação.

Este documento encontra-se estruturado em quatro capítulos: os dois primeiros pretendem dar uma visão geral do que representa a informação no contexto organizacional e do que representam a qualidade da informação e a qualidade do sistema de informação. No capítulo três é efectuada uma análise sobre o processo de auditoria e as metodologias de avaliação de dados mais utilizadas nesta área, identificando a que mais se adequa a esta investigação e apresenta-se o desenho metodológico do projecto que é desenvolvido no capítulo quatro. Neste capítulo encontra-se especificado o desenho do modelo de avaliação da qualidade dos dados em sistemas de informação acompanhado de um exemplo teórico onde se pretende ilustrar a possível aplicação do modelo.

CAPITULO 1 – A qualidade do sistema de informação

1.1. A informação e o sistema de informação no contexto organizacional

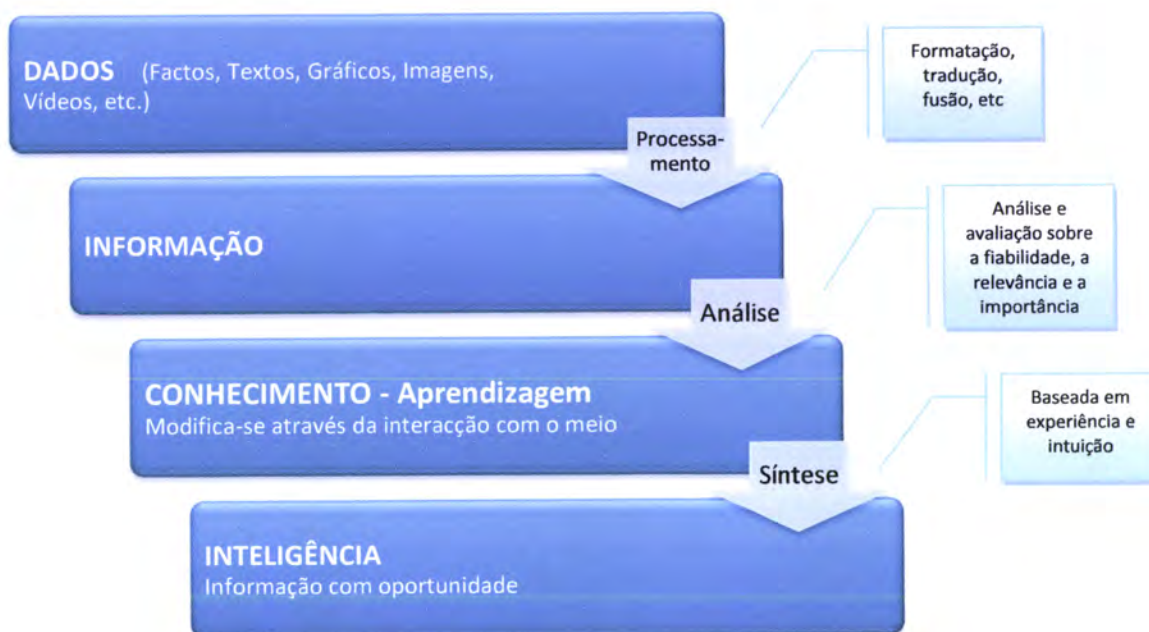
Para Oliveira e Amaral (1999), “*a informação não se compra nem se vende, o seu acesso é que pode ser adquirido, não se desgasta com o uso, o seu valor não é exacto, não se substitui informação e a sua replicação é ilimitada*”. Como tal, a informação é todos os dias utilizada pelas empresas para obter valor (Redman, 2005), o que demonstra a grande importância deste recurso para a organização.

A informação deve atender às necessidades de uma pessoa ou de um grupo, e, neste caso, ela só tem valor se for enviada à pessoa ou ao grupo certo, na hora e no local exacto e na forma correcta (Moresi, 2000; Redman, 2005). Por outro lado, a informação também deve satisfazer o domínio da organização e, neste caso, o seu valor está relacionado com o papel que desempenha na função de tomada de decisão. Assim, o seu valor será tanto maior quanto melhor a decisão que resultar do seu processamento (Moresi, 2000).

As necessidades de informação também diferem dentro da organização se tomarmos em consideração os vários níveis organizacionais: o institucional ou estratégico, o intermédio ou de gestão e o operacional. O nível institucional, onde são tomadas todas as decisões estratégicas da organização, necessitará de informação mais agregada que suporte o processo de tomada de decisão (dados com características qualitativas), enquanto o nível operacional necessitará de informação mais desagregada e com características mais quantitativas.

Assim, de acordo com Moresi (2000), é necessário dividir a informação numa hierarquia que consiga diferenciar as necessidades nas diversas situações, tomando como referência a classificação proposta por Urdaneta (in Moresi, 2000): dados, informação, conhecimento, inteligência (Figura 1.1).

FIGURA 1.1 - OS NÍVEIS HIERÁRQUICOS DA INFORMAÇÃO



Fonte: Adaptado de Moresi (2000)

Uma vez que a informação, o conhecimento e a inteligência provêm dos dados, é absolutamente necessário que estes tenham qualidade para que possam ser tomadas decisões mais adequadas na organização, podendo assim assumir-se que dados de qualidade representam uma vantagem competitiva organizacional.

Para Redman (1998) a má qualidade dos dados tem implicações nos vários níveis organizacionais, ou seja:

- ✓ Ao nível **operacional**, a fraca qualidade dos dados conduz directamente à não-satisfação do cliente, ao aumento do custo (porque se gasta tempo e outros recursos a detectar e a corrigir os erros) e à diminuição da satisfação no trabalho por parte do trabalhador.
- ✓ Ao nível de **gestão (tático)** a má qualidade dos dados compromete a tomada de decisão e aumenta a falta de confiança das organizações umas nas outras.
- ✓ Ao nível **estratégico** a falta de dados relevantes, completos, fiáveis e atempados – acerca de clientes, concorrentes, tecnologias, e outras questões da maior

relevância para o negócio – pode ser o maior obstáculo ao desenvolvimento de uma estratégia.

O impacto da má qualidade dos dados será mais subtil e os efeitos serão muito mais difíceis de observar directamente no nível estratégico do que nos níveis operacional e de gestão.

Assim, não é de estranhar que as organizações tenham cada vez maior preocupação em gerir a qualidade dos dados, encarando a informação como um recurso organizacional, e aumentado o seu investimento, quer ao nível tecnológico quer ao nível humano, para que a informação possa ter as características adequadas à sua utilização, ou seja, tenha qualidade (Oliveira e Amaral, 1999).

Um sistema de informação, independentemente de ser automatizado ou não, pode ser considerado como o conjunto gerido de recursos humanos e materiais, destinados a realizar as actividades de adquirir, armazenar, processar e difundir informação, cuja missão deverá ser a de fornecer informação com qualidade a quem dela necessita, sejam agentes internos ou externos à organização em causa (Oliveira e Amaral, 1999).

Moresi (2000) considera ainda que, uma vez que o fluxo de informação é um processo de agregação de valor, o sistema de informação constitui-se então como a cadeia de valor da informação numa organização, uma vez que serve de suporte para a produção e a transferência desta. Assim, o sistema de informação é *“uma combinação de processos relacionados ao ciclo informacional, de pessoas e de uma plataforma de tecnologias de informação, organizados para o alcance dos objectivos de uma organização”*.

Desta forma, para o desenho de um sistema de informação de qualidade, é necessário conhecer o ciclo de vida da informação. Este conceito é também abordado por Redman (1996), que considera o ciclo de vida da informação dividido em ciclo de aquisição, onde a informação é recebida e analisada, e ciclo de utilização, onde a informação é extraída para ser consumida. Como suporte do ciclo de vida da informação encontra-se a base de dados, onde os dados são armazenados para depois serem processados e finalmente distribuídos (Figura 1.2).

FIGURA 1.2 - CICLO DE VIDA DA INFORMAÇÃO



Fonte: Adaptado de Redman (1996)

Também Laudon e Laudon (2009) consideram que existem três actividades num sistema de informação capazes de produzir a informação que as organizações necessitam para tomar decisões, controlar as operações, analisar problemas e criar novos produtos ou serviços: actividades de *input*, actividades de processamento e actividades de *output*. Nesta medida, para os autores, os sistemas de informação assentam em definições claras e precisas de dados e de procedimentos para recolha, armazenamento, processamento, disseminação e utilização dos mesmos – ciclo sistémico da informação.

Neste contexto, um dos desafios das organizações actuais é compreender o valor estratégico da informação, de forma a poder potencializar a sua correcta utilização, através da implementação de uma gestão da informação eficaz. Para Choo (2003) a gestão da informação é absolutamente necessária à aprendizagem organizacional e compreende seis processos fundamentais que se desenvolvem de forma contínua: a identificação das necessidades de informação, a aquisição de informação, a organização e armazenamento da informação, o desenvolvimento de produtos e serviços de informação, a distribuição de informação e a utilização da informação, particularmente com vista à criação de conhecimento e à tomada de decisão. O autor acrescenta ainda um último processo – comportamento adaptativo – que inclui a selecção de novas

acções com base da informação disponibilizada que vão gerar novas necessidades informacionais – Ciclo gestivo da informação.

1.2. O valor e a qualidade de um sistema de informação

Enquanto suporte dos vários níveis organizacionais, o sistema de informação deve ser considerado um recurso organizacional, que deve por isso ser valorizado. Moresi (2000) considera quatro factores que permitem perceber o valor de um sistema de informação: qualidade, serviço, custo e tempo, dando enfoque particular à essência desse sistema, que é a informação. Nesta medida, os quatro factores referidos são transformados respectivamente em qualidade da informação, portfólio de produtos e serviços, custo e tempo de resposta. Um sistema de informação para fornecer valor superior deverá disponibilizar aos seus clientes todos estes factores em simultâneo.

Moresi (2000) apresenta no seu trabalho uma equação (figura 1.3), que embora metafórica, procura traduzir esta percepção de valor. Assim, “o aumento do valor do sistema estará condicionado à **maximização** do numerador e à **minimização** do denominador.”

FIGURA 1.3 - EQUAÇÃO DE VALOR DE UM SISTEMA DE INFORMAÇÃO

$$\text{Valor} = \frac{\text{Portfólio (Produtos e Serviços)} \times \text{Qualidade}}{\text{Custos} \times \text{Tempo de Resposta}}$$

Fonte: Adaptado de Moresi (2000)

O **portfólio de produtos e serviços** compreende todos os produtos tangíveis (documentos, relatórios, entre outros) e intangíveis (indexação, recuperação, entre outros) que estão relacionados com a matéria-prima informação. Existem múltiplos factores que podem contribuir para quantificar o valor destes produtos, designadamente: fiabilidade, integridade, precisão, clareza, disponibilidade e acessibilidade, entre outros. Já o factor **qualidade** realiza-se por meio de atributos ou características, o que, traduzindo para os sistemas de informação, implica a existência de um conjunto de

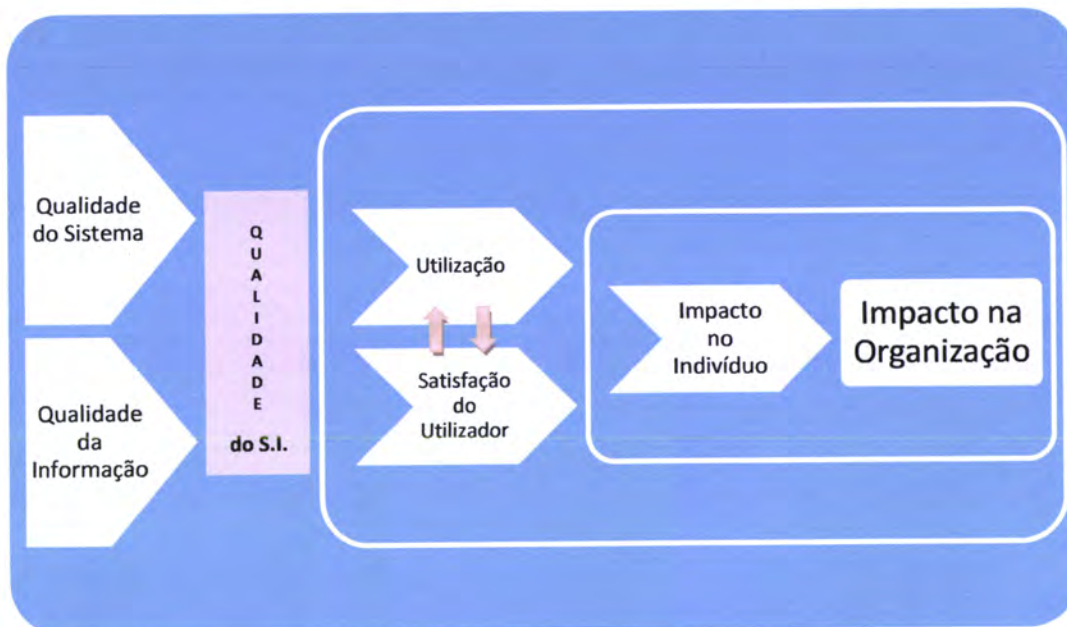
propriedades como a interpretabilidade, a representação ou a relevância, que têm que ser atendidas de modo a que o sistema satisfaça as necessidades dos utilizadores.

No que respeita ao factor **custo** de um sistema de informação, este pode ser medido de forma quantitativa (custo global do ciclo de vida) e de forma qualitativa. Esta última é mais difícil de ser medida e inclui os factores que permitem identificar a contribuição global do sistema de informação em relação às estratégias da organização. O **tempo de resposta** corresponde ao custo da oportunidade, o que implica que o sistema de informação só poderá ser eficaz se responder em tempo oportuno.

A equação de valor de um sistema de informação proposta por Moresi (2000) evidencia a necessidade de aumentar a qualidade do sistema (numerador) para conseguir superar os custos efectivos e de oportunidade (denominador).

DeLone e McLean (1992), baseados nos estudos de Shannon e Weaver (1949) e de Mason (1978), desenvolveram um modelo para o sucesso do sistema de informação (Figura 1.4) que lhes permitiu identificar a qualidade do sistema e a qualidade da informação como os antecedentes chave que, em conjunto, influenciam as medidas de percepção dos benefícios do sistema: a utilidade e a satisfação do utilizador. A quantidade de utilização pode afectar o grau de satisfação do utilizador e vice-versa. Estes são antecedentes directos do impacto na performance individual que condiciona, por sua vez, o impacto organizacional, positiva ou negativamente. Ou seja, se o utilizador estiver perante um sistema de qualidade, utilizá-lo-á mais e melhor, de modo a aumentar o seu desempenho no seio da organização e, conseqüentemente, contribuindo de forma mais eficaz para o desempenho organizacional.

FIGURA 1.4 - MODELO DE SUCESSO DE UM SISTEMA DE INFORMAÇÃO



Fonte: Adaptado de DeLone e McLean (1992)

Para DeLone e McLean (1992), algumas das medidas que mais significativamente poderão ter impacto no utilizador e que, por conseguinte, poderão ser utilizadas para medir a qualidade do sistema poderão ser o tempo de resposta, a facilidade de utilização, a fiabilidade do sistema e ainda medidas de performance, como a utilização de recursos e o retorno do investimento. Quanto às medidas utilizadas para medir a qualidade da informação, estas poderão ser a fiabilidade, a integridade, a relevância, a precisão e a disponibilidade da informação.

Posteriormente Seddon (1997) desenvolveu um estudo que pretendeu interpretar e clarificar o modelo de DeLone e McLean (1992), mais especificamente no que diz respeito à utilidade e à satisfação do utilizador, que define como a satisfação no uso. Este estudo mostrou que a qualidade do sistema e a qualidade da informação, em simultâneo, têm de facto a capacidade de influenciar as medidas de percepção do utilizador relativamente ao benefício do sistema.

Nelson, Todd e Wixom (2005) testaram este modelo em sistemas de *Data Warehouse*, através da análise da qualidade do sistema e da qualidade da informação. O seu objectivo primário neste estudo foi perceber as dimensões-chave da qualidade da informação e da qualidade do sistema, através da identificação daquilo que, neste

contexto, permite avaliar a qualidade de uma tecnologia de suporte ao sistema de informação.

Para tal, desenvolveram um modelo constituído por nove dimensões determinantes da qualidade em sistemas de informação (quatro delas relacionadas com a qualidade da informação e cinco com a qualidade do sistema). Posteriormente, utilizaram estas dimensões, através do teste de 465 utilizadores em sete sistemas de *Data Warehouse*, e concluíram que a fiabilidade e a exactidão são as dimensões-chave da qualidade do sistema e da qualidade da informação, respectivamente.

De acordo com esta análise um sistema de qualidade é aquele que é de fácil acesso (atributo acessibilidade), integrado (atributo integração) e flexível (atributo flexibilidade), e tem de fornecer informação exacta (atributo exactidão), completa (atributo completude/exaustividade) e bem formatada (atributo formatação/representação).

Assim, se olharmos atentamente os antecedentes chave para o sucesso do sistema de informação, é fácil reconhecer os factores que condicionam o valor do sistema de informação de Moresi (2000). Quanto melhor for a qualidade do sistema e da informação, maior será a satisfação do utilizador e, conseqüentemente, o valor individual percebido, que se traduzirá num valor para a organização, influenciando a sua competitividade no mercado.

De facto, parece inquestionável que de nada adianta termos um sistema que contenha informação com qualidade se esta não for disponibilizada em tempo útil, a quem dela necessita. Contudo, não é menos verdade que por melhor que o sistema seja tecnicamente, permitindo níveis de desempenho acima da média, este não será útil se a informação que disponibiliza não tiver qualidade. Ambas as situações irão causar insatisfação no utilizador e não teremos um sistema de informação de sucesso / qualidade, já que terá como consequência um impacto negativo na organização.

Importa salientar, no entanto, que provavelmente o maior impacto negativo que um sistema de informação sem qualidade terá para a organização deverá ser aquele cujo problema de base tem a ver com a qualidade da informação, já que uma má informação poderá ter consequências muito negativas aos vários níveis da organização.

CAPITULO 2 – A qualidade da informação

Uma vez assumida a informação como sendo o resultado do processamento dos dados, pode-se entender que esta apenas será de qualidade se os dados também tiverem essa mesma qualidade. Assim, ao falar-se de qualidade da informação está necessariamente a falar-se de qualidade dos dados.

Nesta perspectiva, é muito frequente na literatura sobre o tema encontrar-se referência à informação de forma abrangente, onde estão implícitos os dados. Deste modo, quando se fala em qualidade da informação fala-se simultaneamente em qualidade dos dados.

Pela importância que a informação tem para a organização, e sendo considerada um produto intangível, é necessário garantir a sua qualidade durante todo o processo de produção, à semelhança do que é efectuado para os produtos tangíveis. Tal como os produtos resultam de um sistema de processamento sobre as matérias-primas, também a informação resulta de um sistema de processamento sobre as suas matérias-primas – os dados (Wang, 1998).

Podemos assim presumir que tal como é necessário garantir a qualidade de um produto, em todo o seu ciclo de produção, da mesma forma é necessário garantir a qualidade de uma informação, em todo o seu ciclo de vida.

A pesquisa que tem sido realizada aponta no sentido da adaptação do ciclo de Deming (Deming, 1986) para a gestão da qualidade (PDCA): *Planear, Executar, Avaliar e Agir*¹; à gestão da qualidade de um produto de informação, transformando-o em *Definir, Medir, Analisar e Melhorar*² (Wang, 1998).

Nestas circunstâncias a componente *definir* identifica as características (dimensões) mais importantes na qualidade dos dados e os respectivos requisitos; a componente *medir* produz as métricas para medir as características seleccionadas, a componente *analisar* identifica as causas dos problemas da fraca qualidade da informação, e a componente *melhorar* fornece as técnicas para melhorar a qualidade dessa mesma informação.

¹ PDCA: Plan, Do, Check, Act.

² DMAI: Define, Measure, Analyse, Improve.

A disciplina da gestão da informação tem sido cada vez mais implementada nas organizações e, desde então, têm sido propostas por vários autores metodologias para a avaliação e melhoria da qualidade da informação, como forma de a gerir de maneira mais eficaz.

Batini et al (2009) desenvolveram um estudo que teve como principal objectivo fornecer uma descrição sistemática e comparativa das metodologias de qualidade dos dados existentes. Para tal, consideraram que estas podem ser caracterizadas sob várias perspectivas (Quadro 2.1):

QUADRO 2.1 – FORMAS DE CARACTERIZAÇÃO DAS METODOLOGIAS DE AVALIAÇÃO DA QUALIDADE DA INFORMAÇÃO

| | |
|---|--|
| Fases e passos que compõem a metodologia | <ul style="list-style-type: none"> a) Reconstituição dos estados (recolha de informação sobre a organização e os seus processos) b) Avaliação/medição c) Melhoramento |
| Estratégias e técnicas | <ul style="list-style-type: none"> a) Orientadas para os dados b) Orientadas para os processos |
| Dimensões e métricas | Não existe acordo relativamente a quais as dimensões que definem a qualidade dos dados, nem quanto ao seu significado. Cada metodologia adopta as suas dimensões, sendo que as mais consensuais são: <i>Exactidão/Precisão; Exaustividade/Compleitude; Consistência e Disponibilidade.</i> |
| Custos | <ul style="list-style-type: none"> a) Custo da fraca qualidade dos dados (custos do processo/custos de oportunidade) b) Custo da avaliação e do melhoramento da qualidade dos dados |
| Tipo de dados | <ul style="list-style-type: none"> a) Dados estruturados; b) Dados não estruturados; c) Dados semi-estruturados; |
| Tipo de sistema de informação | <ul style="list-style-type: none"> a) Sistema de informação monolítico; b) Data Warehouse; c) Sistema de informação distribuído; d) Sistema de informação cooperativo; e) Sistema de informação WEB; f) Sistema de informação P2P (peer to peer) |

Fonte: Elaboração Própria

Desta forma, os autores compararam treze metodologias de qualidade dos dados³ sob estas seis perspectivas e concluíram que cada metodologia focaliza-se tendencialmente num conjunto de características da qualidade dos dados, e classificaram-nas em quatro categorias (Batini et al, 2009, p26 a p27):

1. Metodologias *completas* (focalizam-se na avaliação e no melhoramento) – TIQM, CDQ;
2. Metodologias de *auditoria* (focalizam-se especialmente na fase de avaliação) – AIMQ, CIHI, DQA, AMEQ, QAFD, IQM;
3. Metodologias *operacionais* (focalizam-se apenas nas questões técnicas das fases de avaliação e de melhoramento) – TDQM, DWQ, ISTAT, DaQuinCIS;
4. Metodologias *económicas* (focalizam-se na avaliação dos custos) - COLDQ.

Em qualquer das metodologias o ponto-chave está na selecção das dimensões da qualidade a avaliar. Com a evolução tecnológica dos sistemas de informação, com o crescimento das redes e com a generalização da Internet, quer ao nível das empresas quer da própria administração pública, tornou-se necessário avaliar a qualidade dos sistemas também sob o ponto de vista da sua acessibilidade e reputação. Assim, as dimensões que caracterizam a qualidade da informação, para além de incluírem características de qualidade como a integridade, a exactidão, a exaustividade, entre outras, passaram a incluir características de utilização dos dados, como sejam a sua disponibilidade, a sua acessibilidade e ainda outras respeitantes à segurança. De facto, num sistema em que a base de dados pode estar em qualquer lugar a milhares de quilómetros do local onde o utilizador se encontra, é necessário que a informação desejada chegue correcta, completa e de uma forma célere. Por outro lado, se existem

³ TDQM – Total Data Quality Management (Wang, 1998)
 DWQ – The Data Warehouse Quality Methodology (Jeusfeld et al., 1998)
 TIQM – Total Information Quality Management (English, 1999)
 AIQM – A methodology for Information Quality Assessment (Lee et al., 2002)
 CIHI – Canadian Institute for Health Information Methodology (Long and Seko, 2005)
 DQA – Data Quality Assessment (Pipino et al., 2002)
 IQM – Information Quality Measurement (Eppler and Münzenmaier, 2002)
 ISTAT – ISTAT Methodology (Falorsi et al., 2003)
 AMEQ – Activity-based Measuring and Evaluating of product information quality (AMEQ) Methodology (Su and Jin, 2004)
 COLDQ – Cost Effect Of Low Data Quality (Loshin, 2004)
 DaQuinCIS – Data Quality in Cooperative Information Systems (Scannapieco et al., 2004)
 QAFD – Methodology for the Quality Assessment of Financial Data (De Amicis and Batini, 2004)
 CDQ – Comprehensive methodology for data quality management (Batini and Scannapieco, 2006)

múltiplas fontes de dados, uma vez que vários utilizadores, em vários locais em simultâneo podem inserir informação na base de dados, também é necessário que o sistema consiga garantir credibilidade aos dados para que estes tenham reputação para o utilizador que necessita de os utilizar com segurança.

2.1. Modelos de qualidade da informação

A percepção de qualidade é algo que não é inequívoco e consensual. A visão mais global sobre o conceito é aquela que aponta o consumidor do produto, neste caso trata-se de informação, como o centro de todo o processo de criação de qualidade. Contudo, ao contrário de outros produtos, cuja qualidade pode ser de imediato apurada, a informação é algo que poderá ter elevada qualidade para um utilizador mas não ter a mesma qualidade para outro, revelando-se portanto bastante subjectiva. Assim, a expressão “*fitness for use*”⁴ é utilizada pela grande parte dos autores para a definição de qualidade da informação acreditando-se que esta é a mais flexível e consistente (Bovee, Srivastava e Mak, 2003).

Numa visão mais generalista, a qualidade dos dados pode ser relacionada com um conjunto de “dimensões” que são normalmente definidas como propriedades ou características de qualidade (Scannapieco e Catarci, 2002).

Wand e Wang (1996) definem ainda, a propósito das características da qualidade de um produto de informação, a visão externa e a visão interna de um sistema de informação, sendo que a primeira foca-se no seu uso e no seu efeito, incluindo atributos extrínsecos. Já a segunda, a visão interna, foca-se na sua construção e na sua operação, fases necessárias para obter as funcionalidades requeridas de acordo com os requisitos definidos, que se vão reflectir na visão externa, e incluem atributos intrínsecos.

Dependendo do tipo de abordagem à qualidade da informação têm sido desenvolvidos modelos teóricos que definem as dimensões e os atributos que permitem medir e aferir da qualidade de um produto de informação, ou seja as características que definem a sua qualidade. Assim, é usual encontrar na literatura referência a modelos de qualidade da

⁴ A expressão indica que um sistema de informação terá qualidade se for apropriado para a sua utilização, ou seja, este deve ser capaz de fornecer ao utilizador toda a informação que ele necessite, sendo que esta deverá ser correctamente compreendida e interpretada, aplicável ao domínio de interesse do utilizador, num determinado contexto para o qual o sistema foi concebido, e credível (o utilizador deverá “acreditar” na informação que recebe para a poder utilizar).

informação orientados para os processos, modelos orientados para o sistema e modelos orientados para o utilizador (Bovee, Srivastava e Mak, 2003).

A figura 2.1 ilustra os vários modelos de avaliação da qualidade de um produto de informação.

FIGURA 2.1 – MODELOS TEÓRICOS DE QUALIDADE DA INFORMAÇÃO



Fonte: Elaboração Própria

Tal como a figura pretende ilustrar, os modelos orientados para os processos tendem a incluir mais atributos intrínsecos – de conformidade com os requisitos, como sejam a precisão, a consistência, a integridade e a objectividade, entre outros, mas cujo enfoque se situa sempre no processamento da informação. Nesta abordagem, a qualidade da informação é demonstrada pela qualidade dos processos de tratamento da informação e pela forma como estes estão de acordo com as especificações iniciais do negócio, o que implica que seja demonstrada por características objectivas. De modo inverso, os modelos orientados para o utilizador, embora também incluam os atributos intrínsecos, já incluem aqueles que são chamados de extrínsecos – aqueles que revelam as necessidades e as percepções dos utilizadores, acarretando um grau elevado de

subjectividade, e que estão mais ligados ao interface, como sejam, a fiabilidade, a disponibilidade, a acessibilidade e a relevância, entre outros.

Estes modelos resultam de uma evolução da investigação sobre a qualidade da informação/dados, sendo notória a gradual abertura para o mundo externo, que rodeia o sistema de informação e as bases de dados, que tem levado à tendencial inclusão da preocupação com as necessidades dos utilizadores e com as características de qualidade, do ponto de vista de utilização, que um produto de informação deve conter. A qualidade do sistema de informação deixou de ser encarada apenas sob o ponto de vista do seu *efeito* (se executa bem ou mal os processos para os quais foi criada) mas também passou a incluir a qualidade do seu *uso* (se executa os processos para os quais foi criada de maneira satisfatória para o utilizador). Só desta forma se poderá, de acordo com o modelo proposto por DeLone e McLean (1992) e desenvolvido por Seddon (1997), ter um sistema de informação de sucesso para a organização.

Esta evolução das abordagens à qualidade da informação resulta também da evolução tecnológica dos sistemas de informação que passaram de monolíticos (cuja arquitectura tecnológica consistia apenas num *mainframe* e numa única base de dados, onde os circuitos de informação eram simples e repetitivos podendo os erros existentes ter origem apenas em entradas de dados incorrectas ou em processamentos defeituosos) para sistemas distribuídos em redes, aumentando as fontes de dados e a sua complexidade, bem como as suas formas de representação (Batini et al., 2009).

2.1.1. Modelos orientados para os processos

Os modelos orientados para os processos consideram que a falta de qualidade dos dados se revela pela diferença que pode existir entre as expectativas do utilizador, relativamente ao sistema, e o desempenho que este é capaz de lhe fornecer.

Um modelo de referência na temática da qualidade da informação é o de Ballou e Pazer (1985), que avalia o impacto da qualidade dos dados e dos processos nos *outputs* de um sistema de informação, ou seja, a forma como os dados, desde que entram no sistema até serem apresentados ao utilizador, podem perder qualidade durante o seu processamento. Assim, o enfoque do modelo está mais orientado para os processos do que para os dados.

Uma vez que muitos dos processos necessários para gerar determinada informação utilizam *outputs* de outros processos, se não houver garantia de que estes estão correctos poder-se-á estar a utilizar dados alterados e os próprios processos introduzirem erros sucessivos uns nos outros, levando o sistema de informação a não produzir os resultados esperados.

Assim, o modelo de Ballou e Pazer (1985), pressupõe a existência de módulos de controlo da qualidade dos dados ao longo de todo o processamento, incluindo o controlo de quatro dimensões da qualidade: exactidão/precisão (o valor guardado está em conformidade com o valor real), actualidade (o valor guardado está actualizado), exaustividade/completude (todos os valores para uma determinada variável estão guardados) e consistência (a representação de um valor é o mesmo para casos diferentes).

Por sua vez, Wang, Reddy e Kon (1995) apresentam um modelo cujo enfoque é nos processos e na visão interna do sistema de informação, que revela como ideia base que a perspectiva da qualidade, na disciplina de gestão da informação, é um ponto crítico uma vez que é condição necessária que os utilizadores tenham confiança nos dados para que os possam utilizar, quer ao nível operacional quer ao nível da decisão.

Assim, os autores consideram que deve haver uma intervenção ao nível da modelagem dos processos do sistema de informação de modo a que a base de dados possa conter indicadores de qualidade inseridos na sua estrutura. Para tal é necessário primeiro determinar quais os parâmetros de qualidade importantes para o utilizador, para que este possa avaliar a qualidade dos dados antes da sua utilização. Estes parâmetros de qualidade são utilizados para determinar os indicadores de qualidade a serem associados aos dados, a sua forma de processamento, quando devem ser criados e o seu método de recolha (Wang, Kon e Madnick, 1993).

Com base nos indicadores o utilizador consegue efectuar uma melhor interpretação dos dados e determinar a sua fiabilidade.

Desta forma, Wang, Reddy e Kon (1995) desenvolveram uma hierarquia de dimensões de qualidade dos dados (Figura 2.2) e focalizaram a sua pesquisa em formas de garantir a fiabilidade e a interpretabilidade dos dados, através da modelagem dos processos.

Os indicadores de qualidade inseridos na estrutura da base de dados deverão representar as características relevantes para cada tipo de informação nela contida. Assim, por

exemplo, para determinada informação será importante saber quando ela foi registada, por quem e onde – estas são as características de qualidade para a informação que tem de estar actualizada, só pode ser alterada por utilizadores autorizados e apenas em alguns departamentos da organização.

FIGURA 2.2 - HIERARQUIA DE DIMENSÕES DE QUALIDADE DOS DADOS



Fonte: Adaptado de Wang, Reddy e Kon (1995)

A utilização deste tipo de indicadores de qualidade na base de dados poderá também ser de extrema importância na presença de erros detectados na base de dados, uma vez que poderão permitir a identificação de presumíveis autores de manipulações abusivas dos dados, ou até mesmo a identificação de uma causa interna do problema.

2.1.2. Modelos orientados para o sistema

Um modelo orientado para o sistema considera que a qualidade da informação depende não apenas dos processos de desenho e de produção envolvidos na sua construção, mas também do próprio desenho do sistema e da forma como este é capaz de representar o mundo real.

Wand e Wang (1996) basearam a sua abordagem na noção de que o objectivo de um sistema de informação é fornecer uma **representação** de um domínio aplicacional tal como ele é percebido pelo utilizador (visão do mundo real), concentrando a sua análise na **visão interna** do sistema de informação e, portanto, dando especial enfoque ao

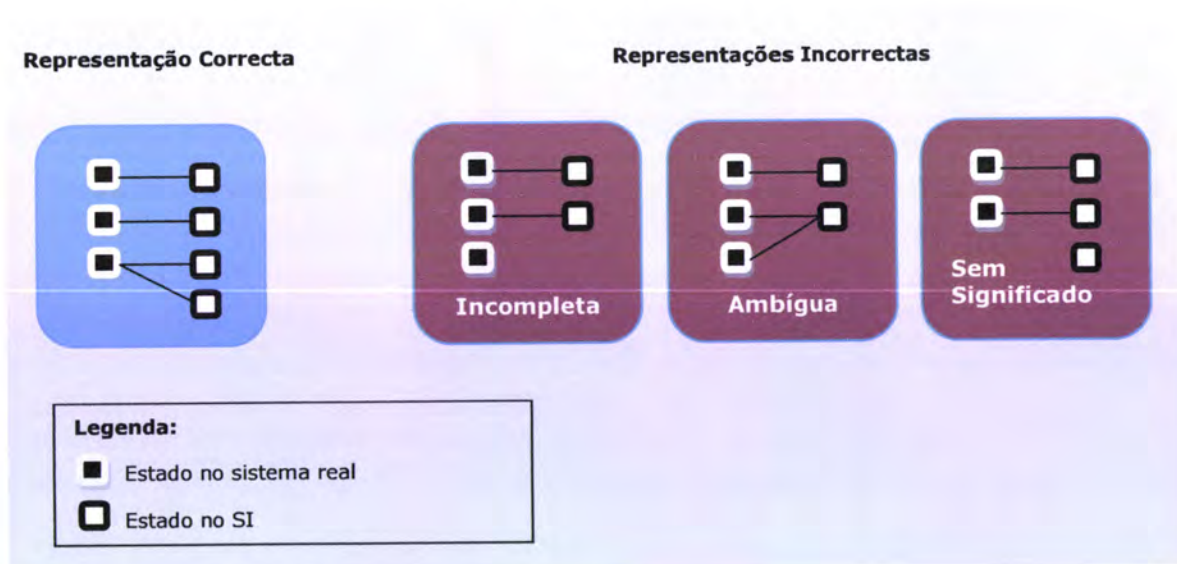
desenho do sistema e à produção dos dados. Nesta visão, como os atributos de qualidade dos dados são intrínsecos à aplicação, é possível compará-los entre várias aplicações.

Assim, é necessário definir aquilo que é directamente observado no domínio do mundo real para poder definir a forma como o sistema de informação deverá representar esse domínio.

Um sistema de informação é construído para ser utilizado pelos utilizadores, cuja visão do sistema do mundo real serviu de base para o seu desenho, criando uma representação **interpretável**, a partir da qual o utilizador consegue **inferir** uma visão do sistema real representado.

Tomando em consideração que o sistema real tem que estar representado de forma apropriada pelo sistema de informação, dever-se-á identificar as possíveis deficiências representacionais que poderão ocorrer durante o desenho do sistema e a produção dos dados. Assim, deste ponto de partida, são três as deficiências de representação que poderão ser encontradas, e que estão representadas na figura 2.3 (Wand e Wang, 1996, pág. 90):

FIGURA 2.3 - REPRESENTAÇÕES DO SISTEMA REAL DE WAND E WANG (1996)



Fonte: Adaptado de Wand e Wang (1996)

O conceito de “estado” está associado a todas as formas que os dados poderão ter em diferentes momentos, dentro de cada funcionalidade do sistema.

Uma representação é correcta ou apropriada se a todos os estados do sistema real corresponder pelo menos um estado no sistema de informação. Simultaneamente é necessário que qualquer estado no sistema de informação possa ser associado ao respectivo estado no sistema real.

As deficiências encontradas são todas as possíveis associações de estados entre o sistema real e o sistema de informação que não estejam nestas condições. São elas:

- ✓ **Representação incompleta** – Existem estados no sistema real que não estão cobertos pelo sistema de informação (um exemplo desta representação é o sistema de informação disponibilizar um campo de dados para inserir o estado civil e não permitir que se insira “divorciado”);
- ✓ **Representação ambígua** – A um estado do sistema de informação não poderá estar associado mais do que um estado do sistema real, uma vez que, se isto ocorrer, não se poderá inferir a qual estado do sistema real pertence o estado do sistema de informação (exemplo de ambiguidade é o caso de o sistema de informação apenas permitir um número de telefone sem definir se se trata de um telefone fixo ou de telemóvel - neste caso o sistema de informação deveria permitir inserir os dois números de telefone);
- ✓ **Representação sem significado** – Apesar de o sistema de informação representar apropriadamente o sistema real, esta deficiência é considerada como tal pelo facto de não constituir um bom desenho, uma vez que representa estados desnecessários (por exemplo um sistema de informação que se destine a recolher e tratar dados de crianças não necessita de ter um campo “estado civil” que permita todos os estados possíveis, uma vez que todas as crianças terão o estado civil de “solteiro” – neste caso é um dado desnecessário, embora reflecta o sistema real).

Wand e Wang (1996) definem ainda aquilo que consideram **deficiências de operação** que ocorrem tipicamente devido a acções incorrectas de utilização do sistema de informação (introdução de dados errados). Quando isto ocorre, basicamente o que se está a fazer é a associar estados do sistema real a estados incorrectos do sistema de informação.

Assim, com base nas deficiências encontradas é possível definir neste modelo as características dos dados de qualidade: estes deverão ser *completos (exaustivos), não ambíguos, com significado e correctos*.

Analisando as possíveis falhas das representações do sistema real no sistema de informação, Wang e Wand (1996) identificaram as dimensões da qualidade intrínseca dos dados:

- ✓ *Exactidão e precisão* – existe falta de exactidão sempre que exista uma associação de um estado do sistema real a um estado válido do SI, mas a associação feita não é a correcta.
 - Deficiências de operação
 - Representação ambígua (mais do que um estado do SI representa o mesmo estado real) – Falta de **precisão**
 - Representação incompleta
- ✓ *Fiabilidade* – Embora não exista uma definição clara de fiabilidade, para efeitos de aplicação do modelo ela é considerada como a correcção dos dados, ou seja, se a todos os estados do sistema real corresponder pelo menos um estado no sistema de informação.
- ✓ *Actualidade e valor* – no modelo, a actualidade refere-se ao intervalo de tempo entre uma alteração num estado do sistema real e a consequente alteração do estado no SI. O valor está intimamente relacionado com a actualidade uma vez que pode ser interpretado como o tempo de armazenamento de cada item de dados.

A falta de actualidade conduz a um estado no SI que pode corresponder a um estado passado do sistema real. Esta dimensão, apesar de poder ser considerada sob a visão externa do SI, conduz à existência de deficiências no sistema uma vez que se os dados estiverem desactualizados irão surgir representações incorrectas, ambíguas e sem significado, especialmente em sistemas integrados se não forem todos actualizados ao mesmo tempo.

- ✓ *Exaustividade/Completude* – É a capacidade do SI representar todos os estados do sistema real com significado.

- ✓ **Consistência** – Para que os dados sejam consistentes o valor de um determinado item de dados só poderá ser o mesmo se aplicado em situações semelhantes. Em termos do modelo, a existência de estados sem significado no SI poderão conduzir a inconsistência dos dados.

Para cada atributo da dimensão intrínseca da qualidade dos dados (exactidão, fiabilidade, actualidade, exaustividade e consistência) é possível identificar a deficiência de representação que lhe está na origem, tornando mais fácil encontrar soluções para a mitigação dos vários tipos de problemas relacionados com os dados.

Todos os problemas que possam ocorrer relativamente aos dados, do ponto de vista da visão interna de um sistema de informação inserem-se necessariamente dentro de uma destas dimensões e têm uma explicação teórica lógica de acordo com Wang e Wand (1996).

Contudo, o grande desafio para a qualidade dos dados em sistemas de informação é a sua capacidade de adaptação às alterações constantes do sistema real. Os dados na base de dados são estáticos enquanto os dados no mundo real estão sempre a alterar-se (Orr, 1998).

Assim, os problemas relacionados com a qualidade dos dados vão muito além da incorrecção dos seus valores, incluindo também problemas e erros de produção, problemas técnicos com o armazenamento e o acesso e problemas causados pelas alterações das necessidades de informação dos utilizadores (Strong, Lee e Wang, 1997a) que poderão ocorrer em todo o ciclo de vida da informação e que se inserem dentro da visão externa de um sistema de informação.

2.1.3. Modelos orientados para os utilizadores

Os modelos orientados para os utilizadores têm em consideração as características de qualidade dos dados na óptica do utilizador. Wang e Strong (1996) desenvolveram um estudo empírico qualitativo que pretendeu perceber o que efectivamente representa a qualidade para os utilizadores. Os autores consideram que um sistema de informação concebe produtos de informação cujo valor é transferido para o utilizador, e identificaram três intervenientes neste processo: os *produtores de informação* (quem

desenha, desenvolve e mantém os dados e a infra-estrutura do sistema), os *gestores da informação* (quem fornece e gere os recursos tecnológicos para armazenar e processar a informação) e os *consumidores da informação* (quem utiliza o produto de informação). O conceito de qualidade da informação para estes últimos deverá ser percebido pelos produtores e pelos gestores da informação para que estes possam conceber e manter um produto que satisfaça os utilizadores.

Os resultados do estudo efectuado por Wang e Strong (1996) mostraram que existem quatro grandes grupos, abrangendo quinze dimensões, que definem a qualidade de um produto de informação (figura 2.4).

FIGURA 2.4 - CATEGORIAS E DIMENSÕES DE QUALIDADE DA INFORMAÇÃO

| Categoria | Dimensão |
|-----------------------------|--|
| Qualidade Intrínseca | Exactidão/Precisão, Objectividade, Credibilidade, Reputação |
| Qualidade de Acessibilidade | Acessibilidade, Segurança |
| Qualidade Contextual | Relevância, Valor-Acrescentado, Oportunidade, Exaustividade/Completo, Quantidade de informação |
| Qualidade de Representação | Interpretação, facilidade de entendimento, representação concisa, representação consistente |

Fonte: Adaptado de Wang e Strong (1996)

A qualidade intrínseca inclui as dimensões relacionadas com a visão interna do sistema de informação que podem condicionar a sua utilização: a exactidão, a objectividade, a credibilidade e a reputação dos dados. O estudo revelou que os utilizadores consideram a credibilidade e a reputação dos dados como parte da qualidade intrínseca da informação, contrariamente ao que é percebido pelos restantes intervenientes no produto de informação, que apenas consideram a exactidão e a objectividade como parte desta dimensão. De facto, se o utilizador se deparar com dados que estão em conflito, são inconsistentes ou imprecisos, não “acreditará” neles e, conseqüentemente, terá dificuldades em utilizá-los. Nesta situação, a reputação dos dados e do sistema de informação em geral estará em causa.

A qualidade de acessibilidade diz respeito às dimensões que poderão afectar a utilização do sistema, porque o utilizador não conseguirá aceder a este de forma apropriada e satisfatória para a sua utilização e com segurança.

Relativamente à qualidade contextual, esta diz respeito ao contexto aplicacional e, deste modo, poderá variar de sistema para sistema. A relevância da informação e o valor-acrescentado que esta poderá trazer, a exaustividade, a oportunidade e a quantidade de informação são dimensões relevantes para esta categoria de qualidade. Estas características revelam que a qualidade da informação deverá ser considerada pelos produtores e gestores da informação dentro do contexto em que esta é utilizada, para que possa ser adequada aos utilizadores.

Quanto à qualidade de representação, está relacionada com o interface entre o sistema de informação e o utilizador, uma vez que este deverá conseguir interpretar correctamente a informação que lhe é apresentada para poder efectivamente utilizá-la. Inclui aspectos relacionados com o formato dos dados (representação concisa e consistente) e com o significado dos dados (interpretabilidade e facilidade de entendimento)

Para além das categorias e dimensões de qualidade mais relevantes para os utilizadores, o estudo de Wang e Strong (1996) permitiu ainda perceber os principais motivos que podem conduzir a problemas de qualidade dos dados e a forma como estes problemas são percebidos pelos utilizadores. São eles:

✓ **Problemas de produção da informação (*input*)**

1. *Múltiplas fontes para a mesma informação produzem valores diferentes* – A existência de múltiplas fontes para os mesmos dados ocorre frequentemente em sistemas multi-aplicacionais sempre que é necessário introduzir o mesmo dado em mais do que um local diferente e até por pessoas diferentes. Se os valores introduzidos para o mesmo dado não forem os mesmos, isto conduz a uma situação de incoerência dos dados gerando desconfiança sobre eles.
2. *Informação produzida utilizando julgamentos subjectivos conduz a informação enviesada* – É frequente em sistemas cuja utilização é dividida por várias pessoas. Se o dado a inserir em determinado campo gerar algumas dúvidas ao utilizador sobre o tipo de valores que este comporta, então este campo goza de subjectividade podendo conduzir a inconsistências na informação.
3. *Erros sistemáticos na produção da informação levam a perda de informação* – Estes erros estão associados aos erros de operação definidos por Wang e Wand



(1996) e consistem em formas de inserir os dados, ainda que incorrectamente, de forma sistemática, que sirvam para contornar algumas validações do sistema. Este problema está também associado a falhas na análise e desenho do sistema, uma vez que não consegue representar todos os estados do sistema real, levando a que o utilizador necessite de “contornar” o sistema. O facto de não se inserir a informação correctamente pode conduzir a situações de informação incompleta. Se o sistema não tiver toda a informação que necessita para produzir os *outputs* necessários ao utilizador, estes deixam de ser relevantes.

✓ **Problemas no armazenamento da informação (Processamento)**

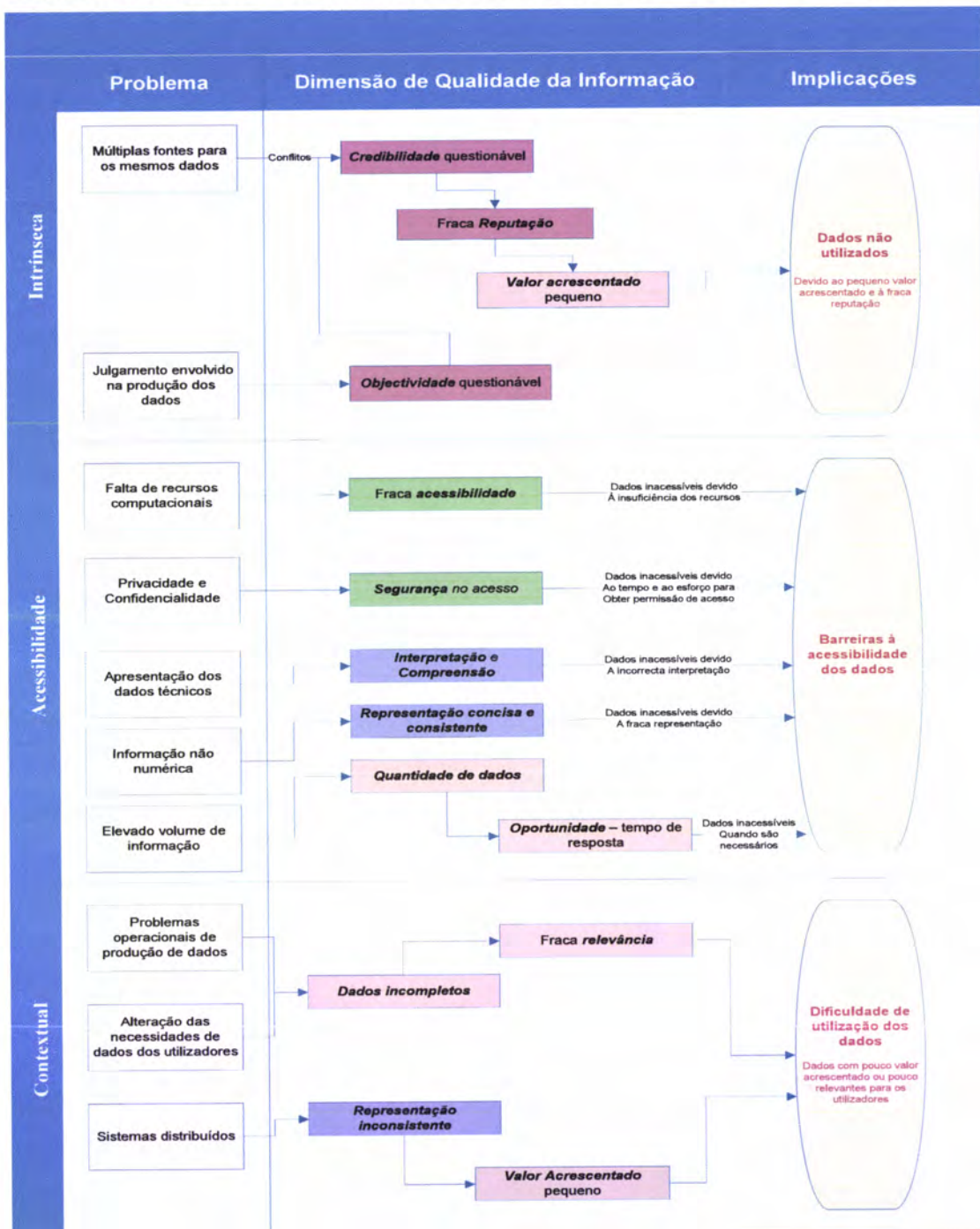
4. *O armazenamento de grandes volumes de informação torna difícil o acesso a essa informação em tempo razoável* – A informação armazenada deve ser apenas aquela que é necessária. A existência de um volume muito elevado de informação armazenada conduz a longos tempos de processamento e consequentemente à produção de informação que poderá chegar ao utilizador já desactualizada. Este é um problema que será percebido pelo utilizador como um problema de acessibilidade. Contudo, trata-se de um problema de qualidade contextual pois o utilizador não obtém a informação necessária, exacta e com valor-acrescentado.
5. *Sistemas distribuídos de forma heterogénea ao invés de sistemas integrados conduzem a definições, formatos e valores inconsistentes* – Quando não existe integração dos sistemas os dados podem ter diferentes representações, em termos quer de formato quer de valor, criando problemas de representação para o utilizador, pois é muito mais difícil a sua agregação para apresentar dados coerentes a quem deles necessita.
6. *A informação não numérica (como imagem e texto) é difícil de indexar* – A informação não numérica é mais difícil de tratar pelo sistema de informação, uma vez que não é possível efectuar tratamentos fiáveis desta, o que pode dar origem a problemas de representação e o utilizador irá ter dificuldades em tratar esta informação de modo a obter dados coerentes e oportunos para o seu trabalho. Este problema é percebido pelo utilizador como um problema de acessibilidade uma vez que este não irá ter acesso a dados correctamente representados, capazes de lhe proporcionar a informação que necessita.

✓ **Problemas de utilização da informação (*output*)**

7. *Apresentação dos dados técnicos* – Este problema reflecte-se tipicamente em sistemas que utilizam informação e terminologia técnica, como é o caso de sistemas que tratam dados clínicos, utilizando terminologia médica, mas que se destinam a ser utilizados não apenas por médicos mas por pessoal de outras áreas de especialidade. O sistema deve ser capaz de apresentar ao utilizador informação que seja interpretável por este e fácil de entender. Se isto não suceder, o utilizador entenderá esta dificuldade como um problema de acessibilidade, uma vez que não será possível aceder à informação de forma apropriada.
8. *As alterações das necessidades de informação dos utilizadores* – Este problema representa a inconsistência entre a informação que o utilizador necessita e aquela que lhe é fornecida pelo sistema de informação. O sistema deve não só ser capaz de satisfazer as necessidades dos utilizadores no momento como também estar suficientemente “aberto” a mudanças na informação. Caso contrário este apenas fornecerá ao utilizador informação incompleta e conseqüentemente pouco relevante.
9. *O acesso fácil à informação pode gerar conflitos com os requisitos de segurança, privacidade e confidencialidade* – Assegurar a privacidade, a confidencialidade e a segurança do sistema requer barreiras de acesso gerando um conflito entre estas dimensões da qualidade da informação.
10. *A falta de recursos computacionais suficientes limita o acesso à informação.* Trata-se exclusivamente de um problema técnico do sistema, uma vez que os seus recursos condicionam o utilizador que poderá ter dificuldade em aceder à informação necessária em tempo oportuno devido a falta de recursos do sistema.

A figura seguinte (figura 2.5) mostra a forma como os problemas enumerados anteriormente são percebidos pelos utilizadores. Salienta-se o facto da categoria de qualidade de representação nunca ser percebida como tal, uma vez que as dimensões da qualidade dos dados inseridas nesta categoria são percebidas como problemas de acessibilidade e contextual.

FIGURA 2.5 - A PERCEÇÃO DOS UTILIZADORES SOBRE AS CATEGORIAS DE QUALIDADE DOS DADOS



Fonte: Adaptado de Wang e Strong (1996) e de Strong, Lee e Wang (1997b)

LEGENDA

- Dimensões da categoria de qualidade intrínseca
- Dimensões da categoria de qualidade de acessibilidade
- Dimensões da categoria de qualidade de representação
- Dimensões da categoria de qualidade contextual

Os problemas de qualidade da informação e de qualidade do sistema condicionam a utilização do SI (DeLone e McLean, 1992; Seddon, 1997). Da figura apresentada pode-se concluir que todos os problemas relacionados com a qualidade do sistema de informação condicionam efectivamente a sua utilização, uma vez que, se o sistema apresentar problemas de categoria intrínseca o utilizador não utilizará os dados porque não confiará neles, se os problemas forem ao nível da categoria de acessibilidade os utilizadores sentirão barreiras à utilização dos dados, e se forem encontrados problemas de categoria contextual o utilizador terá dificuldade na utilização dos dados por estes não estarem de acordo com as suas necessidades.

A existência de *múltiplas fontes para os mesmos dados*, por exemplo uma aplicação obrigar à introdução do mesmo dado em locais ou funcionalidades distintas, pode levar a que esse dado não seja inserido da mesma forma nos dois locais, gerando conflitos entre os dois. O utilizador apercebe-se que os dados estão em conflito, mas não sabe a origem desse conflito, considerando apenas que os dados não são credíveis, uma vez que não sabe qual o que está correcto. Esta situação conduz a uma fraca reputação e consequentemente o utilizador não utiliza os dados. Por outro lado, o problema dos *juízos envolvidos na recolha de alguns dados*, quando não estão claras as regras de preenchimento, fará com que não exista objectividade no seu conteúdo, sendo o processo de produção de dados visto pelo utilizador como subjectivo, o que mais uma vez conduz à falta de credibilidade nos dados e à sua fraca reputação.

Em ambos os casos, existência de múltiplas fontes ou juízos subjectivos, a fraca reputação dos dados perante os utilizadores conduz a um problema que é visto por estes como sendo de qualidade intrínseca mas que na realidade constitui um problema de qualidade contextual, a informação deixa de ter valor para os utilizadores e portanto não é utilizada.

Existem cinco problemas que são vistos pelo utilizador como problemas relacionados com a acessibilidade dos dados, uma vez que como resultado final este encontra barreiras no acesso aos dados que necessita: falta de recursos computacionais, privacidade e confidencialidade, apresentação dos dados técnicos, informação não numérica e elevado volume de informação. Contudo, destes cinco problemas apenas a *falta de recursos computacionais* que conduz a uma fraca acessibilidade e as questões de *privacidade e confidencialidade* que podem conduzir a restrições de acesso constituem efectivamente problemas de acessibilidade. As falhas que se prendem com a

apresentação dos dados técnicos e com a *informação não numérica*, que são na realidade problemas de representação, uma vez que o utilizador tem dificuldades na interpretação e visualização dos dados, são percebidos por este como problemas de acessibilidade uma vez que condicionam o acesso aos dados que necessita. O *elevado volume de informação*, por sua vez, origina a diminuição da velocidade de processamento, conduzindo à falta de oportunidade na disponibilização da informação, o que na verdade constitui um problema contextual de actualidade.

No que diz respeito aos problemas que são percebidos pelos utilizadores como sendo de ordem contextual, pelo facto de conduzirem a dificuldades na utilização dos dados pela falta de relevância, também se verifica que apenas dois dos três apresentados são efectivamente desta categoria. Os *problemas operacionais de produção de dados* e a *alteração das necessidades dos utilizadores* levam a que a informação fornecida pelo sistema de informação seja incompleta e deixe de ser relevante. O problema dos sistemas distribuídos, que podem causar uma representação inconsistente conduzindo a uma informação de fraco valor para o utilizador, é encarado por este também como um problema contextual, quando na realidade se trata de um problema de representação. Na verdade, o sistema possui a informação necessária mas não a representa correctamente para o utilizador.

Assim, dos resultados obtidos no estudo de Wang e Strong (1996) é possível observar que a falta de qualidade intrínseca, tal como ela é percebida pelos utilizadores, tem origem tipicamente em problemas de *input* (quer *inputs* dos utilizadores, quer *inputs* internos de outros processos), a falta de qualidade de acessibilidade tem origem em problemas de utilização dos dados e a falta de qualidade contextual pode ter origem em problemas de *input*, de utilização e de armazenamento dos dados. Quanto à falta de qualidade de representação, ela nunca é percebida pelo utilizador como tal, uma vez que este não consegue distinguir entre o que está dentro do sistema e o que este efectivamente lhe “mostra”.

Desta forma, o problema de apresentação dos dados técnicos e o problema de armazenamento dos dados não numéricos, que na verdade afectam a qualidade de representação, são entendidos pelo utilizador como um problema de acessibilidade, uma vez que ao não conseguir interpretar e representar correctamente os dados ele não consegue aceder aos dados que necessita. Também o problema que pode ocorrer em sistemas distribuídos pode reflectir-se na qualidade de representação dos dados, embora

o utilizador o entenda como um problema de qualidade contextual uma vez que a informação que lhe chega pode não ser coerente e completa.

As categorias definidas neste modelo para a qualidade dos dados: qualidade intrínseca, qualidade do acesso, qualidade contextual e qualidade representacional; baseadas nas dimensões que os utilizadores consideram que melhor reflectem a qualidade dos dados, trouxeram uma nova visão sobre este tema, uma vez que até então a “exactidão/precisão⁵” era o atributo que mais preocupava os profissionais de sistemas de informação.

Por outro lado, este estudo permitiu também perceber que, embora a qualidade dos dados possa ser dividida em quatro categorias (intrínseca, de acessibilidade, contextual e de representação), os utilizadores não conseguem identificar correctamente cada uma delas, não conseguindo distinguir a qualidade de representação das restantes, o que mostra como as várias categorias estão de alguma forma interligadas.

Bovee, Srivastava e Mak (2003) desenvolveram um modelo de qualidade da informação que, à semelhança do modelo de Wang e Strong (1996), baseia-se na ideia de qualidade como “*fitness for use*”, incorporando por isso atributos extrínsecos de qualidade dos dados como a acessibilidade, a interpretabilidade e a relevância.

Para os autores deste modelo existem dois cenários a ter em consideração: a qualidade do ponto de vista da utilização da informação (atributos extrínsecos) e a qualidade do ponto de vista da garantia da existência efectiva de dados/informação de qualidade (atributos intrínsecos). A forma de garantir a qualidade neste último cenário passa pela validação externa por auditores. Assim, o que diferencia este modelo dos anteriores é a preocupação dos autores relativamente à garantia externa de qualidade da informação (auditoria).

Assim, o modelo AI₁RI₂ (Acessibilidade, Interpretabilidade, Relevância e Integridade) inclui os atributos **acessibilidade**, **interpretabilidade** e **relevância** que deverão ser avaliados do ponto de vista da utilização da informação (e portanto pelos utilizadores da informação) e o último atributo – **integridade**, que deverá ser avaliado por terceiros de modo a poder obter-se uma garantia externa de qualidade dos dados (Figura 2.6).

⁵ Em inglês *Accuracy*

FIGURA 2.6 - HIERARQUIA DE DIMENSÕES DE QUALIDADE DOS DADOS



Fonte: Adaptado de Bovee, Srivastava e Mak (2003)

A dimensão integridade incluída no modelo deverá ser validada externamente por auditores, apresentando vários atributos, entre eles o atributo “existência” referenciado pela primeira vez na literatura sobre a qualidade dos dados. Apesar do modelo teórico de Wand e Wang (1996) ter explicado os problemas intrínsecos que podem existir num sistema de informação, através das relações entre os estados do sistema real e os estados do sistema de informação, ele não inclui a preocupação da *existência* dos estados do sistema real. Para os autores, a informação que passe o teste da *existência* não é falsa nem redundante. Exemplos típicos de problemas de existência dos dados são dados por Bovee, Srivastava e Mak (2003), pág. 13: 1) a existência de registos de pessoas que na realidade não existem; 2) registos redundantes, isto é, em duplicado ou repetidos erradamente; 3) campos fictícios; ou 4) valores fictícios para campos válidos. Este sub-atributo é muito importante, do ponto de vista de auditoria, para desviar possíveis tentativas de fraude no manuseamento dos dados, e garantir assim a integridade efectiva.

2.2. Estudo comparativo dos modelos

Os vários modelos de qualidade da informação apresentados nos pontos anteriores ilustram a evolução que a investigação tem levado a cabo, tentando acompanhar também a evolução tecnológica dos sistemas de informação.

O quadro seguinte (Quadro 2.2) mostra os resultados dos modelos apresentados distribuídos pelas quatro dimensões da qualidade dos dados:

QUADRO 2.2 – COMPARAÇÃO DOS MODELOS DE QUALIDADE DOS DADOS

| | Qualidade Intrínseca | Qualidade de Acessibilidade | Qualidade Contextual | Qualidade de Representação |
|---------------------------------------|--|-----------------------------|---|---|
| Ballou e Pazer (1985) | Exactidão/Precisão Exaustividade/ /Compleitude Consistência | | Actualidade | |
| Wang, Reddy e Kon (1995) | Exactidão/Precisão Exaustividade/ /Compleitude Consistência Credibilidade da fonte | Disponibilidade | Relevância Actualidade | Sintaxe Semântica |
| Wand e Wang (1996) | Correcção (<i>fiabilidade</i>) Não ambiguidade (<i>Exactidão/Precisão</i>) | | Exaustividade/ /Compleitude | Significado |
| Wang e Strong (1996) | Exactidão/Precisão Objectividade Credibilidade Reputação | Acessibilidade Segurança | Relevância Valor-Acrescentado Oportunidade Exaustividade/ /Compleitude Quantidade de informação | Interpretabilidade Facilidade de entendimento Representação concisa Representação consistente |
| Bovee, Srivastava e Mak (2003) | Exactidão/Precisão Exaustividade/Compleitude Consistência Existência | Acessibilidade | Relevância | Interpretabilidade |

Fonte: Elaboração Própria

No que diz respeito à **qualidade intrínseca**, destaca-se o atributo *exactidão* em todos os modelos, sendo que no modelo de Wand e Wang (1996) esta é representada pela não ambiguidade. De facto, para estes autores a falta de *exactidão* representa estados no sistema real diferentes daqueles que estão representados no sistema de informação, ou seja, estados incorrectos ou ambíguos de acordo com o seu modelo teórico.

A *exaustividade/compleitude* (garantia de que a informação está completa) e a *consistência* surgem como atributo da qualidade intrínseca em três dos modelos apresentados e simultaneamente como atributo da qualidade contextual nos restantes modelos. Contudo, o seu significado é diferente uma vez que, quando aplicado na qualidade intrínseca, ele refere-se à garantia de que a informação que consta na base de

dados está completa e, quando aplicado à qualidade contextual, refere-se à garantia de que o sistema permite inserir toda a informação necessária. Ou seja, o sistema de informação deve permitir que a informação seja completa, de acordo com todos os estados do sistema real e, por outro lado a informação que efectivamente se encontra registada deve estar completa.

No modelo de Wang e Strong (1996) é possível identificar três atributos que não estão explicitamente presentes nos outros modelos. São eles a *objectividade*, a *credibilidade* e a *reputação*. A objectividade e a credibilidade estão directamente ligadas à introdução dos dados, uma vez que o modelo identificou duas situações capazes de conduzir à falta de objectividade e de credibilidade. De facto, se houver várias fontes para os mesmos dados, ou seja, a informação poder ser introduzida a partir de fontes de dados diferentes pode originar divergências nos dados e conduzir à falta de credibilidade. Por outro lado, se não estiver plenamente estabelecida a forma correcta de introdução dos dados pode dar origem a várias interpretações diferentes e conduzir assim à falta de objectividade da informação. Por fim, estas duas situações de falta de objectividade ou de falta de credibilidade podem conduzir à falta de reputação dos dados e à sua consequente não utilização. Assim, o atributo *reputação* é uma consequência da não existência dos atributos *objectividade* e *credibilidade*.

Bovee, Srivastava e Mak (2003) consideram que a credibilidade e a reputação de uma fonte de informação constituem uma evidência de *integridade*, não sendo por isso consideradas atributos, e podem ser avaliadas sem ser necessário examinar os dados. O mesmo já não se aplica ao atributo *credibilidade da fonte* de Wang, Reddy e Kon (1995), uma vez que este diz respeito a características de registo dos dados que permitam a qualquer entidade externa (auditor) aferir da credibilidade dos dados registados, como sejam datas de inserção e de alteração de registos e utilizadores que efectuaram as operações sobre os dados.

O atributo *existência* aparece apenas em um dos modelos apresentados que considera que ao passar o teste da existência é garantido que a informação não contém entidades ou campos falsos ou redundantes.

A *relevância* é o atributo que aparece em três dos modelos apresentados para a qualidade contextual. De facto, se se tiver informação que se possa entender e

interpretar, então é também necessário que ela seja relevante para o domínio de interesse num determinado contexto (Bovee, Srivastava e Mak, 2003).

Mas a informação só será relevante se for *actual* e se estiver *completa* ao nível contextual, pelo que estes atributos estão intimamente ligados.

O atributo de qualidade *acessibilidade* apenas é focado em modelos orientados para o utilizador, o que é facilmente verificável no quadro acima. De facto apenas os modelos de Wang e Strong (1996) e de Bovee, Srivastava e Mak (2003) consideram a acessibilidade como uma dimensão e um atributo crítico para o utilizador. Contudo, o modelo de Wang, Reddy e Kon (1995) já contemplava também o atributo *disponibilidade* que, embora seja visto mais do ponto de vista do sistema do que do utilizador, pode ser considerado, à luz dos modelos orientados para os utilizadores, um atributo de acessibilidade, uma vez que o utilizador apenas pode aceder ao sistema de informação se este estiver disponível.

Quanto à qualidade de representação, os atributos apresentados pelos vários modelos são vários, mas todos dependem da avaliação subjectiva do utilizador.

Assim, em síntese, os modelos de Wand e Wang (1996) e de Ballou e Pazer (1985) focalizam-se em atributos que podem ser medidos objectivamente, enquanto os modelos de Wang e Strong (1996), de Wang, Reddy e Kon (1995) e de Bovee, Srivastava e Mak (2003) incluem também atributos de avaliação mais subjectiva uma vez que estes dependem da subjectividade inerente às pessoas (utilizadores) que os possam avaliar.

CAPITULO 3 – Desenho metodológico do projecto

A informação assume nos dias de hoje uma importância primordial numa organização. Ela constitui um dos recursos mais importantes para o sucesso organizacional e, ao ser encarada como um recurso, “*significa que esta terá que ser gerida como tal, que será necessária a existência de uma estrutura capaz de assegurar que a informação esteja disponível no momento, na forma e na quantidade desejável para os seus consumidores, ou seja, que tenha qualidade*” (Oliveira e Amaral, 1999).

Todas as decisões dentro de uma organização são tomadas com base na informação disponível. Informação com fraca qualidade produzirá decisões inadequadas.

De acordo com Juran et al (1993) citados por Oliveira e Amaral (1999), a qualidade não existe por si só nas características intrínsecas dos objectos, mas antes na utilização ou aplicação desses objectos. Ou seja, a qualidade da informação só pode ser avaliada por quem a consome, e só terá qualidade se quem a consome a considerar como apropriada para as suas necessidades.

Por seu lado, a qualidade do próprio sistema de informação deverá estar interligada com a qualidade dos dados com que opera, que, por sua vez, influenciam a qualidade daqueles que produz.

Apesar do crescente investimento das organizações em sistemas de informação, este muitas vezes não tem o retorno esperado, devido à sua fraca qualidade. Poderão existir múltiplas causas para este fenómeno, mas não raras vezes assiste-se a problemas relacionados com os dados. Por isso, será que se pode concluir que um sistema de informação tem qualidade, que garante os padrões e características de qualidade e a fiabilidade dos processos de negócio, avaliando a qualidade dos seus dados? Por isso a grande questão que se coloca nesta investigação é: *será possível avaliar a qualidade de um sistema de informação pela avaliação da qualidade dos seus dados?*

Assim, para responder a esta questão, esta investigação inclui quatro objectivos. O primeiro consiste em *identificar o significado de qualidade dos dados em sistemas de informação*, sendo que o termo qualidade poderá não ser igual para todas as pessoas, e poderá depender do consumidor da informação. Neste sentido, pretende-se identificar as

características de qualidade que os dados deverão apresentar nas perspectivas dos vários utilizadores da informação.

O segundo objectivo consiste em *perceber a influência da qualidade dos dados nos órgãos de decisão e conseqüentemente nas políticas da organização*, uma vez que, sendo os dados a matéria-prima da informação, e esta a matéria-prima do conhecimento, é importante perceber a influência que os dados terão no processo de tomada de decisão. Como terceiro objectivo pretende-se *determinar o modo como a qualidade dos dados pode influenciar o desempenho de um sistema de informação, designadamente a fiabilidade dos seus processos de negócio*. Os dados que são inseridos num sistema de informação serão utilizados nos processos de negócio implementados, o que implica que se aqueles não tiverem qualidade, dificilmente o sistema de informação produzirá dados com qualidade.

O quarto e último objectivo é o de *investigar a aplicação de uma framework de avaliação da qualidade dos dados em processos de auditoria informática*, tentando criar um modelo de avaliação da qualidade dos dados que possa ser utilizado em auditoria de sistemas de informação e permita aferir se o sistema tem ou não qualidade, a partir da avaliação da qualidade dos seus dados.

Para a prossecução deste trabalho de investigação procurou-se desenvolver teoria, partindo da análise e interpretação de estudos já realizados por outros investigadores, e efectuando reflexões críticas de forma a encontrar a teoria emergente. Assim, a metodologia adoptada teve por base o paradigma interpretativo, dando origem a uma abordagem qualitativa. O elemento comum em toda a investigação qualitativa é a recolha de dados em forma de palavras ou imagens, a qual é analisada por métodos que não incluem estatística ou quantificação (Strauss e Corbin, 1990).

Desta forma, a metodologia utilizada para a construção deste projecto foi a *Grounded theory*, uma vez que esta permite construir explicações teóricas, que vão crescendo através da recolha e análise sistemática dos dados (Glaser e Strauss, 1967). Assim, foram recolhidos e analisados vários trabalhos de investigação que serviram de base para o estudo do tema escolhido e para a indução da teoria proposta - codificação aberta. A análise comparativa dos vários estudos permitiu chegar à codificação axial deste, ou seja, permitiu descobrir os significados por detrás dos conceitos, através da sua comparação e categorização que, juntas, apresentam padrões ou um conjunto de eixos

que explicam os dados e os relacionamentos entre os conceitos – codificação axial (Strauss e Corbin, 1990).

A indução analítica foi o método de análise escolhido para descobrir e desenvolver a teoria. O termo *indução* refere-se aos processos pelos quais os observadores reflectem sobre a sua experiência do fenómeno social e a partir daí tentam formular explicações que podem ser utilizadas para formar uma regra abstracta, ou um princípio orientador efectuando uma codificação selectiva do fenómeno estudado (Symon e Cassel, 1998).

Assim, ao contrário dos procedimentos dedutivos, a análise indutiva inclui procedimentos que se iniciam pela observação do fenómeno e terminam com a emergência de uma teoria que explica o fenómeno estudado.

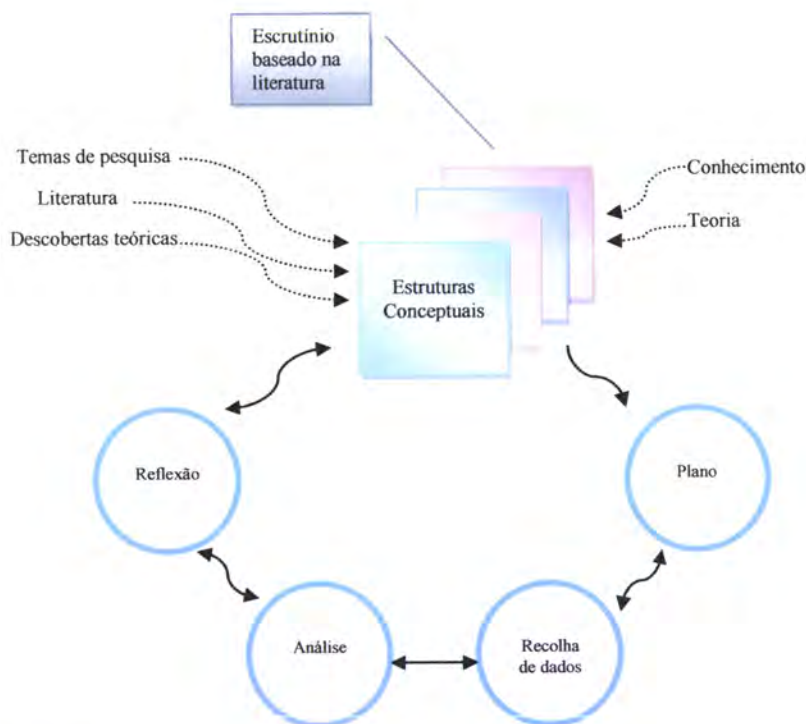
Desta forma, foi necessário encontrar um método que permitisse estruturar o trabalho de maneira a efectuar a recolha e a análise dos dados conducentes à construção da teoria. Foi utilizada a *framework structured-case*, que inclui a construção e articulação de uma estrutura conceptual preliminar, recolha e análise dos dados e reflexão sobre os resultados de forma a construir conhecimento e teoria (Carroll e Swatman, 2000). Esta *framework* metodológica encontra-se dividida em três elementos principais: estrutura conceptual, ciclos de pesquisa e construção da teoria. No primeiro elemento pretende-se que seja desenhado conceptualmente o objecto de estudo, através da explicação dos factores-chave, dos pressupostos e das variáveis que fazem parte do objecto de estudo e dos seus relacionamentos. A estrutura conceptual expressa o conhecimento do investigador sobre o tema da pesquisa e é o ponto de partida para o desenvolvimento exploratório da investigação.

Os ciclos de pesquisa permitem gerar conhecimento sobre o tema definido na estrutura conceptual do trabalho e incluem quatro fases: planeamento, recolha de dados, análise e reflexão. Os conceitos desenvolvidos na estrutura conceptual são utilizados nesta fase como codificação inicial, de forma a guiar a análise, podendo ser adicionados outros códigos que surjam e que se relacionem com o tema em investigação (Carroll e Swatman, 2000). Nesta fase, se forem encontrados novos casos, estes devem ser adicionados à estrutura conceptual inicial, construindo uma nova estrutura conceptual e iniciando novamente um ciclo de pesquisa, já com os novos casos incluídos. Apenas quando já não há casos a adicionar se poderá prosseguir para a construção da teoria.

No elemento de construção da teoria, através de um rigoroso processo de reflexão, os conceitos são clarificados e categorizados, e os relacionamentos entre as categorias especificados, de forma a que a teoria sobre o tema escolhido é induzida (Carroll e Swatman, 2000).

A figura seguinte (figura 3.1) ilustra o método de pesquisa de caso estruturado, utilizado neste trabalho:

FIGURA 3.1 – MÉTODO DE PESQUISA DE CASO ESTRUTURADO



Fonte: Carroll e Swatman (2000)

Um importante aspecto da metodologia *grounded theory* é que a análise é efectuada em simultâneo com a recolha de dados (Esteves, Ramos e Carvalho, 2007). Por esta razão estas duas fases dos ciclos de pesquisa funcionam de forma interactiva, uma vez que à medida que se vai analisando podem surgir novas variáveis para o tema, conduzindo à recolha de mais informação para poder relacionar estas variáveis com as já existentes.

Assim e de acordo com o método de pesquisa de caso estruturado, nos capítulos anteriores foi elaborada a estrutura conceptual do trabalho e iniciado o ciclo de pesquisa

através da recolha dos dados, consubstanciados nos estudos relevantes no que diz respeito à qualidade da informação e à sua relação com a qualidade dos sistemas de informação, e da sua análise comparativa.

A revisão da literatura efectuada permitiu aferir que os estudos desenvolvidos sobre a temática em estudo têm sido construídos sobre abordagens intuitivas e abordagens teóricas à qualidade dos dados. A abordagem intuitiva, através da selecção de atributos de qualidade para um estudo em particular, de acordo com a experiência e intuição do investigador sobre aqueles que lhe parecem ser os mais importantes atributos para avaliar a qualidade da informação, tem a vantagem de ser adaptável a qualquer realidade objecto de estudo, uma vez que os atributos são seleccionados caso a caso. A abordagem teórica à qualidade da informação tem o seu enfoque na forma como os dados se podem tornar deficientes durante o seu processamento. A grande vantagem deste tipo de abordagem é que consegue fornecer um conjunto completo de atributos de qualidade dos dados que são intrínsecos a qualquer produto de informação. O exemplo mais conhecido na literatura sobre este tipo de abordagem é o de Wand e Wang (1996).

Contudo, estes dois tipos de abordagem têm a desvantagem de se focarem apenas no produto em si, nas suas características de desenvolvimento, e não olharem para as suas características de utilização (Wang, Strong e Guarascio, 1996).

Foi na década de 90 que surgiram os primeiros estudos que já incluíram atributos como *acessibilidade* e *interpretabilidade*, revelando uma preocupação com as características de utilização do produto de informação (Wang, Reddy e Kon, 1995; Wang, Kon e Madnick, 1993) e apontando o caminho para os modelos de avaliação da qualidade da informação orientados para os utilizadores.

No capítulo seguinte pretende-se reflectir sobre os resultados da análise dos dados e induzir um modelo de avaliação da qualidade dos dados que possa ser aplicado em auditoria de sistemas de informação, que passa necessariamente pela avaliação da qualidade dos dados.

CAPITULO 4 – Modelo de avaliação de um sistema de informação

4.1. O processo de auditoria e as metodologias de avaliação da qualidade da informação

A qualidade da informação está directamente ligada à qualidade do sistema de informação. De acordo com o modelo de sucesso de um sistema de informação de DeLone e McLean (1992) a base de um sistema de informação de sucesso é a qualidade da informação processada e a qualidade técnica do sistema.

A auditoria constitui uma actividade de exame rigoroso aos procedimentos/processos de uma organização por forma a garantir que estes são adequados para a prossecução eficaz dos seus objectivos. Por seu lado, a auditoria informática ou de sistemas pretende garantir que as tecnologias de informação implementam correctamente todos os processos da organização de forma fiável para suportar o processo de tomada de decisão e ajudar a desenvolver o trabalho operacional da organização.

Assim, do ponto de vista de auditoria, o maior enfoque dá-se na fase de avaliação do sistema de informação que suporta os processos organizacionais para que as respectivas acções de melhoramento possam ser desencadeadas de forma coerente e eficaz.

De acordo com as metodologias de avaliação da qualidade dos dados revistas por Batini et al (2009), existem seis delas⁶ que são especificamente vocacionadas para a actividade de auditoria, uma vez que se focalizam exactamente na fase de avaliação (Quadro 4.1).

Dentro da fase de avaliação todas as metodologias analisadas incluem o procedimento de análise dos dados a avaliar, cinco delas identificam as áreas críticas para seleccionar os dados mais críticos que devem ser avaliados. Ainda cinco metodologias aplicam métricas para medir a qualidade dos dados seleccionados. Quanto às dimensões a avaliar, as metodologias AIMQ, CIHI e QAFD incluem dimensões fixas estando a aplicação da metodologia sempre sujeita à avaliação daquelas dimensões, enquanto as

⁶ AIQM – A methodology for Information Quality Assessment (Lee et al., 2002)
CIHI – Canadian Institute for Health Information Methodology (Long and Seko, 2005)
DQA – Data Quality Assessment (Pipino et al., 2002)
IQM – Information Quality Measurement (Eppler and Münzenmaier, 2002)
AMEQ – Activity-based Measuring and Evaluating of product information quality (AMEQ) Methodology (Su and Jin, 2004)
QAFD – Methodology for the Quality Assessment of Financial Data (De Amicis and Batini, 2004)

metodologias DQA, AMEQ e IQM permitem que sejam seleccionadas as dimensões a avaliar de acordo com os dados e com os processos.

QUADRO 4.1 - ESPECIFICAÇÕES DAS METODOLOGIAS DE AUDITORIA

| | | AIMQ | CIHI | DQA | AMEQ | QAFD | IQM |
|--|--|-----------|-------|-----------|-----------|-------|-----------|
| Fase de avaliação | Análise de dados | X | X | X | X | X | X |
| | Análise de requisitos de qualidade dos dados | | | | | X | |
| | Identificação de áreas críticas | X | X | X | X | X | |
| | Modelagem de processos | | | | X | | |
| | Medição da qualidade | X | | X | X | X | X |
| | Dimensões | Fixas | Fixas | Maleáveis | Maleáveis | Fixas | Maleáveis |
| Fase de melhoramento | Identificação das causas dos erros | | | X | X | | |
| | Monitorização dos melhoramentos | | | | X | | |
| Tipos de dados | Estruturados | X | X | X | X | X | X |
| | Semi-estruturados | Adaptável | X | | Adaptável | | X |
| Tipos de sistemas de informação | Monolíticos | X | X | X | X | X | |
| | Distribuídos | Adaptável | X | Adaptável | | | |
| | WEB | | | | | | X |

Fonte: Adaptado de Batini et al (2009)

Existem ainda duas metodologias que efectuem alguns dos processos de melhoramento: a metodologia DQA identifica as causas dos erros e a metodologia AMEQ, para além desta identificação, efectua ainda a monitorização dos melhoramentos. Todas as restantes metodologias não efectuem qualquer procedimento da fase de melhoramento, sendo utilizadas apenas para avaliar os dados.

Quanto ao tipo de dados e ao tipo de sistema de informação a avaliar, também existem algumas diferenças nas metodologias analisadas, sendo que a DQA e a QAFD apenas avaliam dados estruturados e as restantes podem avaliar dados estruturados e semi-estruturados. As metodologias AIMQ, CIHI e DQA podem avaliar dados de sistemas monolíticos ou distribuídos, enquanto a AMEQ e a QAFD apenas avaliam dados de sistemas monolíticos e a IQM se destina a avaliar sistemas WEB.

Existem dois tipos de auditoria: interna e externa. Quanto à primeira é efectuada dentro da própria organização por auditores internos, cujo objectivo consiste na avaliação dos processos organizacionais de modo a identificar os problemas existentes e apontar medidas para a sua mitigação. Quanto à auditoria externa, esta é efectuada por auditores

de fora da organização e pretende ir um pouco mais além emitindo recomendações para que se possam melhorar os processos. A auditoria inclui as fases de planeamento, execução e relato, sendo que este último inclui a avaliação e as recomendações de melhoramento e para que o auditor possa recomendar é necessário conhecer as causas dos erros.

Por este motivo, optou-se por seleccionar neste trabalho de investigação uma metodologia para avaliar a qualidade dos dados que seja o mais abrangente possível para poder ser utilizada em qualquer auditoria informática e que inclua a identificação da causa dos erros como um dos seus procedimentos. Assim, e apesar da metodologia AMEQ abranger esta fase, foi seleccionada a metodologia DQA (Data Quality Assessment) uma vez que esta pode ser utilizada tanto para dados estruturados como semi-estruturados, e ainda, não apenas em sistemas monolíticos mas também em sistemas distribuídos. Por outro lado, a metodologia AMEQ foi desenvolvida para avaliar dados de produtos mecânicos e por esse motivo dá um enfoque especial ao procedimento de modelagem de processos.

A metodologia DQA será utilizada como ponto de partida para a avaliação da qualidade dos dados e será adaptada aos propósitos desta investigação.

A definição mais consensual de qualidade dos dados é aquela que, considerando o utilizador como consumidor da informação, admite que esta apenas terá qualidade se servir para aquilo que os seus consumidores pretendem (*fitness-for-use*). Neste sentido, quando se avalia a qualidade dos dados deve-se estar preparado, por um lado para perceber as percepções subjectivas dos utilizadores e, por outro, para efectuar uma avaliação objectiva através de *queries* às bases de dados (Pipino et al, 2002).

A avaliação objectiva deverá ser suportada por métricas que permitam avaliar de forma inequívoca as dimensões da qualidade dos dados seleccionadas. Já a avaliação subjectiva poderá ser efectuada através de questionários aos consumidores dos dados ou outros utilizadores com interesse nos dados.

Existem ainda dois tipos de métricas: as independentes dos processos e as dependentes dos processos. As primeiras reflectem o estado dos dados sem ser necessário conhecer o contexto do sistema e podem por isso ser aplicadas a qualquer base de dados, as segundas já dependem do contexto aplicacional, sendo necessário conhecer as regras de

negócio envolvidas, as disposições regulamentares e as regras aplicadas à base de dados, para poder efectuar a avaliação.

A metodologia DQA identifica os princípios gerais de medição da qualidade comuns nas pesquisas anteriores, dividindo as métricas (dependentes e independentes dos processos) destinadas a efectuar a avaliação objectiva em três classes: rácio simples (*simple ratio*), valor mínimo ou valor máximo (*min or max value*) e média aritmética.

Rácio simples – mede o rácio entre os resultados desejados e o total dos resultados. Pode ser também aplicado pela excepção medindo o rácio entre os resultados não desejados sobre o total dos resultados subtraído de 1. O resultado será um valor entre 0 e 1, sendo que 1 é o valor mais desejado e 0 o menos desejado.

Dimensões que podem ser avaliadas por esta métrica: *exactidão/precisão; exaustividade/completude; consistência; representação e relevância.*

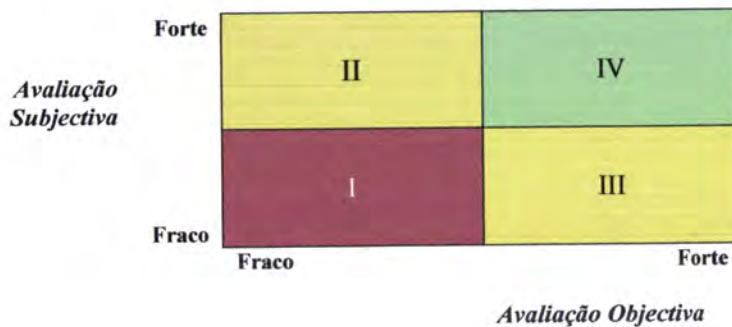
Valor mínimo ou valor máximo – métrica utilizada para avaliar as dimensões que requerem a agregação de vários indicadores de qualidade (variáveis). Nesta métrica o valor da avaliação da qualidade de uma dada dimensão corresponde ao menor ou ao maior dos valores obtidos nas várias variáveis avaliadas para a dimensão. Por exemplo, para avaliar a credibilidade pode-se ter três variáveis; avaliação dos utilizadores relativamente à credibilidade que estes sentem relativamente aos dados; comparação com standards; e experiências anteriores. Cada uma destas variáveis é avaliada com um valor, normalmente entre 0 e 1, sendo considerado para efeitos de avaliação da dimensão o menor ou o maior consoante o critério de mínimo ou máximo valor.

Dimensões que podem ser avaliadas com esta métrica: *credibilidade; quantidade de informação; disponibilidade e acessibilidade.*

Média aritmética – Constitui uma alternativa à métrica do valor mínimo ou máximo, considerando como o valor final de avaliação da dimensão a média aritmética entre os valores atribuídos às variáveis.

Depois de efectuadas a análise dos dados a avaliar, a identificação das áreas críticas e as avaliações subjectivas e objectivas das dimensões seleccionadas, a metodologia DQA propõe uma matriz de classificação (figura 4.1):

FIGURA 4.1 – METODOLOGIA DQA - MATRIZ DE CLASSIFICAÇÃO DA QUALIDADE



Fonte: Pipino et al, 2002

O resultado da comparaç o entre a avaliaç o subjectiva e a avaliaç o objectiva ir  cair num dos quatro quadrantes da matriz, sendo que o objectivo ser  que caia no quadrante IV para que esteja garantido um sistema de informaç o com qualidade. Se o resultado da an lise indicar o quadrante I, existe de facto um problema com a informaç o que tem influ ncia no modo como o utilizador v  o sistema de informaç o, tratando-se de um problema que dever  ser solucionado. Por outro lado, se o resultado da an lise indicar um dos quadrantes II ou III, revela que n o existe concord ncia entre aquilo que   a percepç o do utilizador e os dados f sicos da base de dados, implicando a necess ria investigaç o.

4.2. Desenho do modelo de avaliaç o

Associando os modelos da qualidade da informaç o j  analisados, mais uma vez, ao modelo de sucesso de um sistema de informaç o desenvolvido por DeLone e McLean (1992),   f cil distinguir claramente as dimens es que est o intimamente ligadas   qualidade da informaç o e as que est o ligadas   qualidade t cnica do sistema (figura 4.2), apontadas por estes como os antecedentes chave de um sistema de informaç o de sucesso, ou seja, de qualidade.

FIGURA 4.2 - INTEGRAÇÃO DOS MODELOS DE QUALIDADE DA INFORMAÇÃO COM O MODELO DE SUCESSO DE UM SISTEMA DE INFORMAÇÃO DE DELONE E MCLEAN (1992)

| | Qualidade Intrínseca | Qualidade Contextual | Qualidade de Acessibilidade | Qualidade de Representação |
|---------------------------------------|--|---|-----------------------------|---|
| Ballou e Pazer (1985) | Exactidão/Precisão Exaustividade/ /Compleitude Consistência | Actualidade | | |
| Wang, Reddy e Kon (1995) | Exactidão/Precisão Exaustividade/ /Compleitude Consistência Credibilidade da fonte | Relevância Actualidade | Disponibilidade | Sintaxe Semântica |
| Wand e Wang (1996) | Correcção (<i>fiabilidade</i>) Não ambiguidade (<i>Exactidão/Precisão</i>) | Exaustividade/ /Compleitude | | Significado |
| Wang e Strong (1996) | Exactidão/Precisão Objectividade Credibilidade Reputação | Relevância Valor-Acrescentado Oportunidade Exaustividade/ /Compleitude Quantidade de informação | Acessibilidade Segurança | Interpretabilidade Facilidade de entendimento Representação concisa Representação consistente |
| Bovee, Srivastava e Mak (2003) | Exactidão/Precisão Exaustividade/ /Compleitude Consistência Existência | Relevância | Acessibilidade | Interpretabilidade |
| | Qualidade da informação | | Qualidade do sistema | |

Fonte: Elaboração Própria

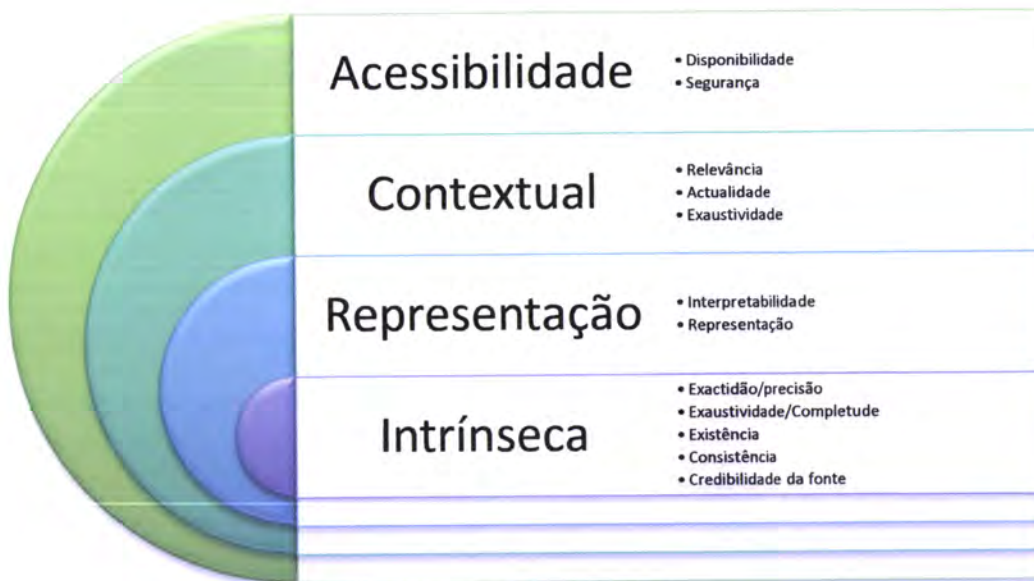
Assim, é necessário que estejam presentes com sucesso as quatro dimensões da qualidade para que o sistema de informação possa ser um sistema de sucesso para a organização. Em modo inverso, basta que uma das dimensões não ofereça qualidade, para que se altere a percepção de satisfação do utilizador e consequentemente diminua a qualidade do sistema.

Então, para avaliar um sistema de informação poderá não ser necessário avaliar todas as dimensões da qualidade dos dados, uma vez que, logo que uma delas apresente deficiências já não será necessário avaliar as outras para poder concluir que o sistema apresenta deficiências e, logo, não poderá ser um sistema de sucesso.

Deste modo, a hipótese que se propõe para avaliar a qualidade de um sistema de informação consiste numa avaliação por níveis, que correspondem a cada uma das quatro dimensões da qualidade, sendo que, basta que uma das dimensões revele fraca qualidade para que o sistema no seu todo não tenha qualidade. Apenas se passará para o nível de avaliação seguinte se o anterior passar todos os testes de qualidade.

Para avaliar a qualidade de um sistema de informação deverá então partir-se do nível mais baixo, que está mais intimamente ligado ao coração do sistema – a base de dados, que é representado pela dimensão intrínseca e ir abrindo a avaliação do sistema para os níveis seguintes: dimensão de representação, dimensão contextual e dimensão de acessibilidade (Figura 4.3).

FIGURA 4.3 – MODELO DE AVALIAÇÃO DE UM SISTEMA DE INFORMAÇÃO



Fonte: Elaboração Própria

O modelo apresentado na figura anterior sustenta a ideia de que a maioria dos sistemas de informação de insucesso apresenta problemas no nível mais baixo, intrínseco, pois todos estes problemas reflectem de alguma forma a fraca qualidade aos outros níveis (representação, contextual e de acessibilidade), não sendo pois necessário passar à avaliação dos níveis superiores. Isto só acontecerá se efectivamente todos os atributos

intrínsecos (exactidão/precisão, exaustividade/completude, existência, consistência e credibilidade na fonte) forem avaliados positivamente.

Por outro lado, existem problemas que embora sejam identificados nos níveis das dimensões contextual, de representação e de acessibilidade, eles revelam-se igualmente no nível da dimensão intrínseca. Por exemplo, o atributo *relevância* na dimensão contextual poderá revelar-se também no atributo *consistência* da dimensão intrínseca. De facto, um factor de relevância é a actualização dos dados, sobretudo aqueles que são fundamentais para aplicação das regras de negócio. Se estes não forem actuais a informação deixa de ser relevante e poderá conduzir a inconsistência dos dados uma vez que as regras de negócio são aplicadas com base em dados incorrectos no momento porque estes não estão actuais.

Outro exemplo é o atributo *interpretabilidade* na dimensão de representação, que poderá revelar-se no atributo *exactidão/precisão* da dimensão intrínseca, uma vez que se não houver uma interpretação correcta dos dados, pode conduzir à introdução de outros dados incorrectos.

De igual forma, problemas no atributo *segurança* da dimensão acessibilidade poderão revelar-se no atributo credibilidade da fonte da dimensão intrínseca, dado que, se houver indicação dos utilizadores que inseriram ou alteraram registos, poderá saber-se por exemplo, se alguém sem permissões alterou determinado registo.

Seguidamente apresenta-se uma breve descrição de todos os atributos incluídos no modelo apresentado, de acordo com o que é mais comumente aceite na literatura (Wang, Reddy e Kon, 1995; Wang, 1996; Redman, 1996; Wang e Strong, 1996; Pipino et al, 2002; Bovee, Srivastava e Mak, 2003):

Dimensão intrínseca

Está intimamente ligada às questões mais físicas dos dados e que revelam a sua fiabilidade e credibilidade perante os utilizadores.

Exactidão/Precisão

Não existe uma definição consensual entre os vários autores para o atributo *exactidão* (accuracy). As definições mais comuns são, por um lado, a que considera a exactidão

como sendo o modo como os dados estão correctos, fiáveis e livres de erro (Wang, 1996); e, por outro lado a que considera o modo como os dados no sistema de informação se aproximam dos do sistema real (Redman, 1996).

Em termos gerais a exactidão mede a distância entre um valor apresentado num campo de uma tabela da base de dados (v) e o valor que é considerado correcto (v'), sendo que existem dois tipos de exactidão: sintáctica e semântica.

A primeira refere-se a valores errados, ou seja, v é diferente de v' e v está efectivamente incorrecto. A segunda refere-se a valores incorrectos, ou seja, ainda que v possa ser um valor correcto no sistema real não está correcto no contexto do registo que se está a analisar.

A métrica mais utilizada para aferir da *exactidão* é o rácio simples: Número de valores correctos / número total de valores.

Exaustividade/Compleitude

A definição de exaustividade ou completude é mais consensual do que a exactidão, sendo considerada a medida em que os dados existem (não estão em falta) e são suficientes para o fim a que se destinam (Wang, 1996).

Podem ser identificados três tipos de exaustividade/completude: Exaustividade do esquema; exaustividade dos campos e exaustividade da população. A exaustividade/completude do esquema consiste na medida em que não faltam entidades nem atributos ao esquema da base de dados; a exaustividade/completude dos campos é verificado pela medida em que não faltam valores num campo ou numa tabela; e a exaustividade/completude da população consiste na quantidade de valores em falta mediante uma determinada população (Pipino et al, 2002).

O rácio simples é a métrica mais utilizada para este atributo: Número de valores não vazios / número total de valores.

Existência

Toda a informação que existir no sistema de informação tem que existir no sistema real exactamente no mesmo contexto em que é inserida no SI, ou seja, o atributo existência verifica se existem dados falsos ou redundantes na base de dados. Neste caso, ao

contrário de dados incorrectos, que não passam o teste da exactidão, tratam-se de dados fictícios, que possuem uma correspondência no sistema real mas não no mesmo contexto em que estão no sistema de informação (Bovee, Srivastava e Mak, 2003).

A informação que não passa o teste da existência é deliberadamente falseada para simular um estado real que poderia efectivamente ter ocorrido, mas que na realidade não ocorreu.

Outro problema de existência é a redundância de registos (dois registos para a mesma entidade) e que leva à ambiguidade uma vez que um deles não deverá existir. Na presença de uma situação destas torna-se difícil saber qual a entidade “verdadeira”.

Métrica do rácio simples: Número de registos distintos / Número total de registos numa tabela e Número de valores falsos/número total de valores.

Consistência

A verificação da consistência procura capturar violações de regras definida num conjunto de dados. Existem duas categorias de inconsistências: intra-relação e inter-relação.

As primeiras têm a ver com as relações entre atributos simples numa única tabela, enquanto as segundas já verificam as relações entre mais do que uma tabela. Por exemplo, verifica-se uma inconsistência intra-relação se uma pessoa com menos de 16 anos tiver na base de dados, simultaneamente, o estado civil de *casado* ou se tiver uma data de nascimento superior a uma data de falecimento. A inconsistência intra-relação existe uma vez que em qualquer dos dois casos é difícil saber qual das informações está correcta. Uma inconsistência inter-relação poderá ser uma pessoa ter associado um estado civil que não existe na tabela de estados civis. Neste caso, se houver uma regra de integridade referencial estabelecida, para além de inconsistência de dados verifica-se uma quebra de integridade referencial.

A métrica mais utilizada para este atributo é o rácio simples: número de valores consistentes/número total de valores.

Credibilidade da fonte

A credibilidade da fonte pretende incluir as questões apontadas por Wang, Reddy e Kon (1995) (inclusão de *tags* de qualidade na própria base de dados), e outras adicionais como sejam a existência de históricos de alterações de registos, que permitam saber a evolução de determinado registo no tempo, e a existência de trilhos de auditoria que permitam a todo o tempo efectuar o *roll-back* dos registos para verificar a precisão do seu processamento. Este atributo é muito importante do ponto de vista do auditor externo porque permite, para além de atestar a veracidade dos dados e a correcção do seu processamento, despistar quaisquer manipulações indevidas dos dados.

Inclui métricas de valor mínimo e valor máximo onde serão incluídas variáveis como a existência de *tags* de qualidade e a existência de trilhos de auditoria, entre outras.

Dimensão de representação

Tem a ver com a forma como os dados são mostrados e revelados ao utilizador, ou seja, está intimamente ligado às questões de interface.

Interpretabilidade

A interpretabilidade corresponde à forma como a informação é interpretável para o utilizador. Está mais ligada à qualidade do sistema, uma vez que se este não fornecer informação facilmente interpretável pode levar o utilizador a efectuar juízos incorrectos, levando-o inclusivamente a introduzir informação incorrecta.

A informação inteligível é aquela que é capaz de ser percebida pelo utilizador e a informação com significado é aquela que fornece ao utilizador uma sensação de atribuição de significado ou valor (Bovee, Srivastava e Mak, 2003).

Assim, para ser interpretável a informação tem que ser apresentada ao utilizador de forma inteligível e com significado.

Representação

A representação de um sistema de informação deve ser concisa e consistente, e tem a ver com o aspecto como a informação é apresentada ao utilizador (formato dos dados)

(Wang e Strong, 1996). A informação deve ser apresentada de modo compacto, representar correctamente as regras de negócio, e ainda apresentar sempre o mesmo tipo de formato, para ser mais facilmente interpretável pelo utilizador.

Dimensão contextual

Revela a qualidade dos dados face ao contexto aplicacional em que se inserem.

Relevância

A informação relevante é aquela que traz valor acrescentado para as tarefas diárias dos utilizadores (Strong, Lee e Wang, 1997a). A informação irrelevante não tem utilidade para o utilizador.

Assim, o sistema de informação deve ser capaz de tratar e produzir toda a informação que tenha capacidade de gerar valor para o utilizador.

Actualidade

A actualidade da informação contribui para a sua relevância e conseqüente utilidade. Está directamente ligada com o valor da informação no tempo e com a sua volatilidade (velocidade com que a informação se altera no sistema real).

O atributo de qualidade da informação actualidade pode ser medido através da diferença entre a data em que a informação foi registada na base de dados e a data em que esta se alterou no sistema real.

Exaustividade

A exaustividade ao nível da dimensão contextual refere-se à permissão por parte do sistema de informação da existência de todos os estados do sistema real. Este atributo também está ligado à relevância, uma vez que uma informação só será relevante se for completa e actual para o fim a que se destina, só assim podendo ser útil.

Um sistema de informação deve assim, permitir que seja inserida toda a informação necessária para que seja completa para o fim a que se destina.

Dimensão de acessibilidade

Está ligada à questão de que os dados deverão estar acessíveis a quem deles necessita.

Disponibilidade

O atributo disponibilidade tem qualidade se a informação estiver suficientemente actualizada relativamente ao fim a que se destina. Não importa apenas a informação estar actual, ela tem que ser disponibilizada ao utilizador na altura certa para ser útil.

Se a informação não estiver disponível na altura certa, todos os outros atributos de qualidade são irrelevantes (Bovee, Srivastava e Mak, 2003).

Segurança

A segurança no acesso, que assegura a privacidade e a confidencialidade, deve ser garantida mas de forma adequada para que não sejam criadas barreiras à acessibilidade. Por outro lado, o sistema de informação tem que oferecer segurança de modo a garantir a credibilidade dos dados.

Apenas para os atributos de qualidade intrínseca é possível encontrar métricas objectivas que permitam aferir objectivamente da qualidade dos dados. Para as restantes dimensões torna-se difícil definir métricas uma vez que apenas podem ser avaliadas de acordo com juízos subjectivos dos utilizadores.

Oliveira et al (2005), desenvolveram uma taxonomia de erros passíveis de se encontrarem nos dados, onde são reflectidos todos os problemas de qualidade intrínseca, que poderá ser utilizada para avaliar o nível mais básico do modelo de avaliação da qualidade dos dados proposto (Quadro 4.2).

QUADRO 4.2 - TAXONOMIA DOS PROBLEMAS DE QUALIDADE DOS DADOS

| 1. Uma única base de dados | | |
|--|--|---|
| 1.1. Uma única tabela | | |
| 1.1.1. Análise de um valor simples | a) Valor em falta (NULL) | Inexistência de um valor num atributo de inserção obrigatória. |
| | b) Violação de sintaxe | Existe uma discrepância entre a sintaxe (formatação) apresentada pelo atributo e aquela que este deveria ter (por exemplo em relação a campos de data). |
| | c) Valor desactualizado | O valor do atributo não está actualizado, não correspondendo à situação real. |
| | d) Violação de intervalo | Existe uma violação do valor do atributo relativamente ao intervalo de valores válidos. |
| | e) Erro de introdução | O valor apresentado está escrito de forma errada, normalmente por erro na introdução de dados. |
| | f) Valor inadequado no contexto do atributo | O valor introduzido no atributo não corresponde ao que deveria lá estar (por exemplo, escrever o nome na morada ou vice versa) |
| | g) Valor introduzido para além do contexto do atributo | São introduzidos valores em excesso no atributo, embora estes estejam dentro do contexto (por exemplo quando se insere o código postal na morada e existe um campo específico para o efeito). |
| | h) Valor sem significado | O valor introduzido não tem a ver nem com o conteúdo do campo onde está nem com nenhum dos outros. |
| | i) Valor impreciso ou com significado duvidoso | As abreviaturas são um caso de valor impreciso pelo facto de permitirem várias interpretações. |
| | j) Violação de regras de domínio | O valor introduzido viola as regras do domínio em que se insere a tabela. |
| 1.1.2. Análise de todos os valores de um campo | a) Violação de valor único | Existem dois registos com o mesmo valor, quando a regra existente é de que este deve ser único. |
| | b) Existência idêntica | Utilização arbitrária de valores diferentes mas com o mesmo significado. |
| | c) Violação de regras de domínio | Violação de uma regra que envolve os valores do campo num todo. |
| 1.1.3. Análise de todos os valores de um registo | a) Registo não completo | Um grande número de campos de inserção obrigatória no registo não está preenchido. |
| | b) Inconsistência entre valores | Existe uma violação de valores que se relacionam entre si (por exemplo a data de casamento ser anterior à data de nascimento). |
| | c) Violação de regras de domínio | Violação de regras que envolvem dois ou mais atributos do registo, e que são inerentes ao domínio. |
| 1.1.4. Análise de todos os valores da tabela | a) Redundância relativamente a uma entidade | A mesma entidade é representada por mais do que uma forma em vários registos. Na prática existe repetição de registos, sendo a entidade representada de forma diferente (por exemplo num registo o nome está por extenso e no outro está abreviado) |
| | b) Inconsistência relativamente a uma entidade | A mesma entidade é representada em vários registos mas em cada um deles os dados contêm divergências, conduzindo à falta de confiança na entidade. |

| 1.2. Várias tabelas relacionadas | |
|---|---|
| 1.2.1. Violação de integridade referencial | Quando o valor de um campo que é chave estrangeira de outro contém um valor que não existe na chave primária daquele (por exemplo, quando é associado a um registo de moradas um código postal que não existe na tabela de códigos postais possíveis). |
| 1.2.2. Referência desactualizada | Apesar da integridade referencial ser respeitada, o valor da chave estrangeira não foi actualizado e não corresponde à situação real. |
| 1.2.3. Inconsistência de sintaxe | Dependendo da relação, existem várias sintaxes de representação para o mesmo tipo de atributo (por exemplo nos atributos do tipo data existem várias representações: dd/mm/aaaa; aaaa/mm/dd) |
| 1.2.4. Inconsistência entre valores de atributos relacionados | Existência de inconsistências entre os valores de atributos de várias tabelas que se relacionam entre si (por exemplo um campo com o valor total ser diferente a soma dos campos dos valores das parcelas). |
| 1.2.5. Circularidade entre registos | Corresponde a inconsistências de valores de atributos que se relacionam todos entre si de forma circular. |
| 1.2.6. Violação de regras de domínio | Violação de regras inerentes ao domínio que envolvem os relacionamentos de várias tabelas. |
| 2. Várias bases de dados | |
| 2.1. Inconsistência de sintaxe | Sintaxes diferentes para o mesmo atributo nas várias bases de dados. |
| 2.2. Unidades de medida diferentes | Unidades de medida diferentes utilizadas em atributos de várias bases de dados que se relacionam. |
| 2.3. Inconsistência de representação | Diferentes conjuntos de valores, do mesmo tipo ou não, são utilizados em atributos relacionados de bases de dados diferentes, para representarem as mesmas situações (por exemplo numa base de dados ser utilizado <i>T</i> ou <i>F</i> para <i>true</i> ou <i>false</i> e noutra ser utilizado o <i>1</i> ou <i>0</i> para representar exactamente o mesmo). |
| 2.4. Diferentes níveis de agregação | O nível de detalhe apresentado em diferentes bases de dados para relações equivalentes não é igual. |
| 2.5. Existência sinónima | Utilização de valores sintacticamente diferentes com o mesmo significado semântico em atributos relacionados em bases de dados diferentes. |
| 2.6. Existência antónima | Utilização de valores sintacticamente iguais mas com diferentes significados semânticos em atributos relacionados em bases de dados diferentes. |
| 2.7. Redundância relativamente a uma entidade | A mesma entidade é representada por uma representação igual ou equivalente em diferentes bases de dados. |
| 2.8. Inconsistência relativamente a uma entidade | Existem inconsistências e contradições num ou mais atributos de uma entidade, representada em mais do que uma base de dados. |
| 2.9. Violação de regras de domínio | Relações equivalentes pertencentes a diferentes bases de dados respeitam individualmente uma regra inerente ao domínio, mas violam-na quando consideradas num todo. |

Fonte: Adaptado de Oliveira et al (2005)

A taxonomia apresentada poderá ser validada através da execução de *queries* às bases de dados que indicarão a presença ou não dos erros nos dados. Estes podem ser posteriormente medidos através das métricas de qualidade de propostas, dando a

indicação do índice de qualidade dos dados e validando a necessidade de continuar a verificação dos outros níveis do modelo.

4.3. Discussão do modelo

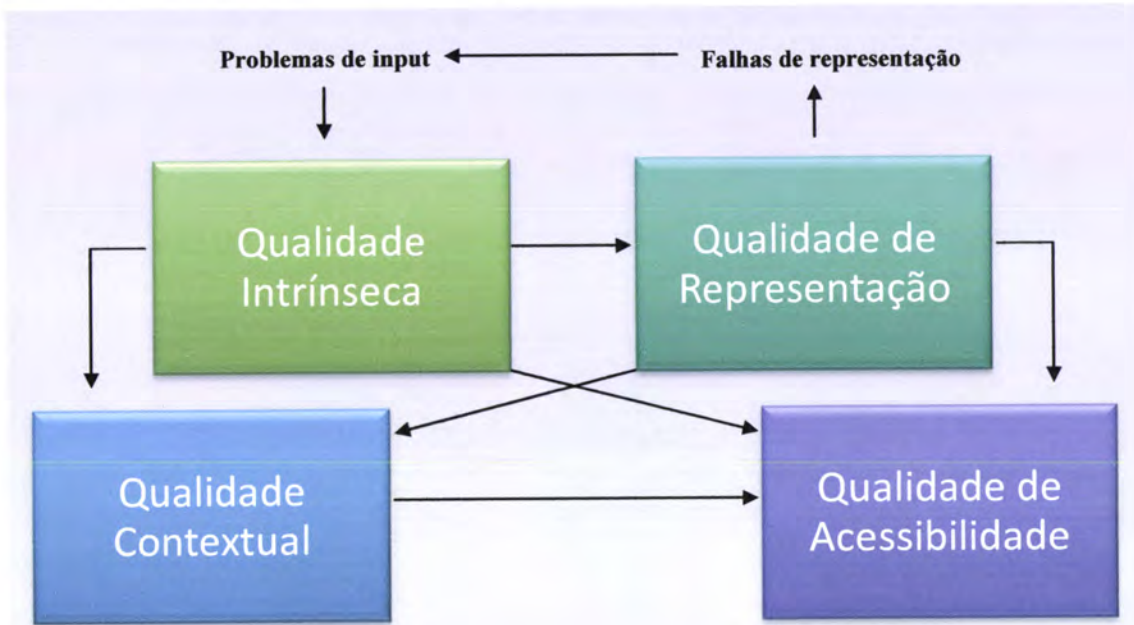
O modelo apresentado inclui as quatro categorias de qualidade dos dados que actualmente são comumente aceites pela comunidade científica, uma vez que é possível organizar todos os atributos de qualidade dentro destas categorias. Os atributos de qualidade foram seleccionados tendo em conta toda a revisão da literatura e incluindo a preocupação da avaliação do sistema de informação do ponto de vista do auditor.

Por este motivo, são importantes os atributos existência e credibilidade da fonte na dimensão intrínseca, uma vez que estes dão pistas muito importantes sobre a segurança e a credibilidade de todo o sistema de informação. Não é só necessário que os dados tenham qualidade mas também que o utilizador possa confiar no sistema e na informação que este lhes proporciona, sob todos os pontos de vista de qualidade.

Tomando como referência o estudo realizado por Wang e Strong (1996), no que diz respeito aos seus resultados relativamente aos problemas que afectam a qualidade dos dados e a forma como esta é percebida pelos utilizadores e é capaz de os influenciar, a explicação do modelo de qualidade dos dados proposto pode ser realizada através da figura 4.4.

A falta de qualidade intrínseca nos dados tem maioritariamente origem em problemas relacionados com o seu *input*. Os problemas de *input*, por sua vez, podem ter origem em falhas na representação (múltiplas fontes para os mesmos dados, fraca objectividade na representação originando fraca interpretabilidade e consequentemente permitindo múltiplos julgamentos para os mesmos dados).

FIGURA 4.4 - ESPECIFICAÇÃO DO MODELO DE QUALIDADE DOS DADOS



Fonte: Elaboração Própria

Desta forma pode-se afirmar que a dimensão de representação poderá ter uma influência forte e directa na qualidade intrínseca. A introdução pelo utilizador de dados incorrectos, a falta de processos correctamente desenhados que indiquem ao utilizador aquilo que está errado e o que efectivamente podem inserir repercutem-se no nível da qualidade de representação, uma vez que os processos e os dados não são correctamente representados ou estes não são correctamente interpretados pelo utilizador.

Refira-se como exemplo que alguns problemas de exactidão e de exaustividade poderão ter origem em falhas na representação, nomeadamente quando o sistema não disponibiliza informação concisa e consistente que indique correctamente ao utilizador o que este deve inserir.

Assim, ao avaliarmos a qualidade intrínseca poderemos simultaneamente aferir da qualidade de representação.

A qualidade contextual também é afectada por problemas de *input*, pelo que, avaliando a qualidade intrínseca que apenas é afectada por este tipo de problemas, pode-se concluir que, se existem problemas deste género, provavelmente estes irão afectar a qualidade contextual.

Por outro lado os problemas de armazenamento também afectam a qualidade contextual na forma como esta é percebida pelo utilizador, uma vez que podem provocar representações inconsistentes dos dados. Assim, na verdade trata-se de um problema de representação que é percebido pelo utilizador como um problema de qualidade contextual (vd. Figura 2.5).

Assim, ao avaliarmos a qualidade intrínseca e a qualidade de representação, estamos a avaliar a qualidade contextual.

Por sua vez, a má representação dos dados dá origem a problemas na sua interpretabilidade (nos *outputs*) e a má estrutura de armazenamento permite que a representação dos dados não seja adequada. Contudo, para o utilizador a má representação num sistema de informação origina problemas que, para ele, são problemas de acessibilidade (vd. Figura 2.5). Assim, a fraca representação dos dados poderá estar na origem de alguns problemas de acessibilidade.

Pode ainda verificar-se que alguns problemas de armazenamento podem condicionar a informação que é disponibilizada ao utilizador, criando-lhe aquilo que para si são problemas de acessibilidade mas que na realidade são problemas de ordem contextual, uma vez que não são proporcionados aos utilizadores os dados relevantes na altura certa (actuais), ou seja, estes não são oportunos.

Assim, se avaliarmos a qualidade de representação e contextual poderemos aferir da qualidade de acessibilidade.

De acordo com o modelo proposto, se forem encontrados problemas de qualidade ao nível da dimensão intrínseca, existem necessariamente problemas nos outros níveis, logo, o sucesso do sistema de informação estará em causa. Se, por outro lado, não forem encontrados problemas na dimensão intrínseca, então deverão ser verificadas as restantes dimensões de acordo com a avaliação subjectiva dos utilizadores. O inverso não se aplica, ou seja, a existência de deficiências ao nível da qualidade de acessibilidade, por exemplo, não implica necessariamente a existência de problemas de qualidade intrínseca.

Este modelo permite avaliar objectivamente a qualidade da informação e a qualidade do sistema num sistema de informação, um dos vectores da matriz proposta pela metodologia DQA (Pipino et al, 2002).

Do ponto de vista de auditoria, para chegar à conclusão de que o SI não tem de facto qualidade basta fazer esta avaliação, uma vez que ainda que o utilizador não tenha percepção de que existem de facto problemas no sistema, o auditor saberá que este não oferece qualidade. Interessará efectuar a avaliação subjectiva (o segundo vector da matriz de Pipino et al, 2002) para verificar em que medida é que a falta de qualidade do sistema de informação é percebida pelo utilizador. Este tipo de avaliação terá um interesse mais elevado do ponto de vista de auditoria interna, para que se possam efectuar as necessárias adaptações ao sistema, uma vez que, de acordo com os problemas encontrados podem ser adoptadas formas de mitigação.

Na avaliação subjectiva deverão ser verificados os aspectos que se relacionam com a utilização da informação e que se prendam com qualidade de representação, de acessibilidade e contextual do sistema de informação.

4.4. Aplicação Real

Nesta fase do trabalho e com o objectivo de tornar mais claro tudo o que anteriormente foi dito e de validar o modelo proposto, efectuou-se uma avaliação de um sistema de informação real, actualmente em produção, com recurso à metodologia DQA e de acordo com o modelo de avaliação de dados proposto.

Descrição da organização

A organização em causa pertence à Administração Pública portuguesa e, por motivos de confidencialidade, o seu nome não poderá ser divulgado nesta dissertação.

O sistema de informação avaliado gere os “clientes” da organização, tendo em conta que serve de base a todo o negócio organizacional. Este negócio passa pela atribuição de “prémios” aos clientes, de acordo com determinadas regras, sendo que existe limite de prémios por cada cliente e que estes necessitam de reunir determinadas condições para que os referidos prémios possam efectivamente ser-lhes atribuídos. Neste sentido, o sistema de informação de gestão de clientes é o “coração” de todo o negócio operacional, uma vez que, se este não tiver qualidade, a organização corre o risco de

atribuir os prémios de forma incorrecta, aumentando os seus custos e, conseqüentemente, diminuindo o sucesso organizacional. De facto, para esta organização a manutenção de um sistema de informação de gestão de clientes com elevada qualidade é fundamental e constitui uma vantagem competitiva.

Avaliação

Para avaliar o sistema de informação foram efectuados *queries* à base de dados, recorrendo à taxonomia apresentada no quadro 4.2, que permitiram identificar os problemas de qualidade dos dados, nomeadamente para a dimensão intrínseca, de forma objectiva. Posteriormente, foram efectuadas entrevistas a alguns utilizadores para perceber qual a sua percepção relativamente à utilização do referido sistema, constituindo esta uma avaliação subjectiva.

Apesar de não estarem definidas métricas objectivas para avaliar as dimensões de representação, contextual e de acessibilidade, estas foram avaliadas de acordo com a experiência profissional do auditor.

Os resultados obtidos foram colocados na matriz de classificação de Pipino et al (2002), de acordo com a metodologia DQA desenvolvida pelos autores, permitindo concluir acerca da qualidade do sistema de informação do ponto de vista do auditor e do ponto de vista de utilização.

Resultados

Os resultados da avaliação efectuada encontram-se espelhados no quadro 4.3 e resultam na seguinte matriz de classificação DQA:

FIGURA 4.5. RESULTADOS OBTIDOS NA AVALIAÇÃO - EXEMPLO TEÓRICO

| | | | |
|-----------------------------|--------------|--|------------------------|
| <i>Avaliação Subjectiva</i> | Forte | Exaustividade/Compleitude Representação Relevância Exaustividade contextual Segurança | Credibilidade na fonte |
| | Fraco | Exactidão/Precisão Existência Consistência Interpretabilidade Actualidade Disponibilidade | |
| | | Fraco | Forte |
| | | <i>Avaliação Objectiva</i> | |

QUADRO 4.3 – RESULTADOS OBTIDOS NA AVALIAÇÃO – EXEMPLO TEÓRICO

| Dimensão da Qualidade | Avaliação Objectiva | Avaliação Subjectiva | Observações |
|-----------------------------------|---------------------|----------------------|---|
| Qualidade intrínseca | | | |
| Exactidão/Precisão | Fraco | Fraco | Este atributo foi avaliado objectivamente através da análise dos valores das tabelas e resultou numa avaliação de fraca qualidade. Verificou-se, por exemplo, a existência de pessoas com o mesmo número de contribuinte, o que revela não só um problema de exactidão dos dados, como também pode originar problemas de consistência nos dados. Os utilizadores demonstraram igualmente a percepção da falta de exactidão dos dados. |
| Exaustividade/Compleitude | Fraco | Forte | A exaustividade foi avaliada através da verificação da existência de campos sem nenhum valor preenchido, quando estes deverão ser de preenchimento obrigatório. Mais uma vez refira-se o caso de existirem pessoas sem número de contribuinte associado. Contrariamente ao esperado, os utilizadores manifestaram acreditar que o sistema dispõe de toda a informação que estes necessitam. |
| Existência | Fraco | Fraco | A existência foi avaliada objectivamente de fraco uma vez que foram encontrados registos em duplicado, redundantes para o mesmo cliente, com a particularidade destes registos conterem em alguns casos valores diferentes associados aos campos. Os utilizadores manifestaram igualmente a percepção de que este atributo de qualidade da informação é fraco. |
| Consistência | Fraco | Fraco | O atributo consistência pode ser afectado quando existem problemas nos atributos descritos anteriormente, nomeadamente no que diz respeito à inconsistência entre registos. Para além disso, foram encontradas inconsistências dentro do mesmo registo como, por exemplo, a data de nascimento ser superior a outras datas. Refira-se também a existência de clientes com idades inferiores a 16 anos e que têm uma profissão associada. Esta situação de inconsistência, dado que legalmente ninguém pode trabalhar antes desta idade, acarreta possivelmente um problema de exactidão, uma vez que um dos dois campos contém informação não exacta. Os utilizadores manifestaram a mesma opinião, percebendo que o atributo consistência é fraco no sistema de informação em causa. |
| Credibilidade da fonte | Forte | Forte | A credibilidade da fonte foi avaliada através da detecção da presença de campos nas tabelas que indicam o utilizador e a data de alteração dos registos, bem como a existência de tabelas que registam o histórico das alterações aos dados. Foi também verificado se todos os utilizadores que tiveram intervenção nos registos estavam devidamente autorizados. Os utilizadores manifestaram igualmente a mesma percepção. |
| Qualidade de representação | | | |
| Interpretabilidade | Fraco | Fraco | Objectivamente o atributo interpretabilidade foi avaliado de fraco dado que se a informação que é mostrada ao utilizador não tem qualidade, este terá dificuldade em interpretá-la, de forma a poder utilizá-la correctamente. Um exemplo desta situação é por exemplo o caso da redundância de registos. Se o utilizador pretender visualizar os dados de um determinado cliente, e se esse cliente estiver registado em duplicado no sistema mas com alguns dados diferentes, o utilizador terá dificuldades em perceber qual dos registos contém a informação correcta. |

| | | | |
|------------------------------------|-------|-------|---|
| | | | Os utilizadores revelaram a mesma avaliação pelo facto de já se terem deparado com algumas questões desta natureza. |
| Representação | Fraco | Forte | A representação é necessariamente avaliada de fraco, desde logo porque não consegue representar correctamente as regras de negócio, permitindo a existência de problemas intrínsecos nos dados apenas passíveis de acontecer porque não existem validações de recolha correctamente implementadas. De igual forma isto conduz também a que a representação da informação para o utilizador possa ser incorrecta e por isso mesmo fracamente interpretável. O utilizador não consegue aperceber-se destes problemas, considerando talvez que o sistema tem um problema de acessibilidade, não conseguindo aceder aos dados que necessita. |
| Qualidade contextual | | | |
| Relevância | Fraco | Forte | Uma vez que existem problemas de exaustividade dos dados, o atributo relevância não poderá ser avaliado objectivamente de forte, uma vez que nem toda a informação relevante se encontra no sistema. Contudo, neste caso o utilizador não atribui a mesma avaliação, considerando que o sistema inclui toda a informação relevante. |
| Actualidade | Fraco | Fraco | Este sistema mostrou um nível fraco de actualidade, uma vez que já foram detectados alguns problemas originados pelo facto dos dados não serem actualizados de forma célere, causando até custos adicionais para a organização. Refira-se a título de exemplo o caso de um cliente que deixa em determinada data de reunir as condições para usufruir dos prémios que a organização dá. Se os dados não forem de imediato actualizados no sistema, a organização continuará a emitir prémios para o cliente quando este já não tem direito a recebê-los. Esta é uma situação também conhecida dos utilizadores que classificam este atributo igualmente de fraco. |
| Exaustividade | Fraco | Forte | A exaustividade ao nível contextual implica que todos os estados do sistema real estejam representados no sistema de informação. No sistema avaliado, embora todos os estados estejam de certa forma representados no sistema real, eles não estão correctamente representados, permitindo falhas nos dados. Por este motivo este atributo recebeu uma avaliação objectiva de fraco. Contudo, para o utilizador |
| Qualidade de acessibilidade | | | |
| Disponibilidade | Fraco | Fraco | Se o sistema não consegue representar correctamente os dados e tem igualmente um problema de actualização, a disponibilidade dos dados de forma correcta e oportuna será colocada em causa. Isto é demonstrado também na avaliação dos utilizadores que avaliam este atributo de fraco. |
| Segurança | Fraco | Forte | O atributo segurança foi avaliado objectivamente através de uma análise cuidada à gestão dos utilizadores do sistema de informação. A avaliação de fraco prende-se com o facto de ter sido detectado um número elevado de utilizadores (62% do total de colaboradores da organização) que podem alterar os dados e, destes, 12,5% têm privilégios de administração dos dados, ou seja, têm acesso total ao sistema. Contudo, para os utilizadores esta situação não é detectada, considerando o sistema com uma forte qualidade em segurança. Os problemas de segurança poderão, no entanto, reflectir-se nos dados, ainda que para o utilizador apenas seja perceptível na avaliação de outros atributos. |

Fonte: Elaboração Própria

Como resultado global da avaliação do sistema de informação de gestão de clientes é notória a fraca qualidade demonstrada. No entanto, apesar de uma avaliação objectiva permitir aferir este resultado, para os utilizadores isso nem sempre é visível. Esta situação poderá ter várias origens que importa verificar do ponto de vista de auditoria interna. No entanto, ainda que os utilizadores não se apercebam dos problemas, eles existem na realidade e o produto de informação fornecido por este sistema poderá estar a influenciar negativamente os resultados organizacionais.

Este exemplo teórico permite verificar que de facto a falta de qualidade intrínseca influencia necessariamente as restantes dimensões. No caso de esta avaliação ser efectuada no âmbito de uma auditoria externa, bastaria avaliar a qualidade intrínseca para aferir que o sistema não tem qualidade, uma vez que necessariamente vai reflectir a má qualidade nas outras dimensões. Do ponto de vista de auditoria interna, é importante verificar todas as dimensões de qualidade e perceber quais as diferenças entre a avaliação objectiva e a avaliação subjectiva.

CONCLUSÕES E TRABALHO FUTURO

Numa sociedade onde o sucesso das organizações está fortemente ligado com a informação com que estas operam, a grande vantagem competitiva não está na disponibilidade da informação por si só, mas sim na qualidade desta. Nesta medida torna-se vital a avaliação da qualidade da informação para que se possam tomar decisões eficazes com base nela.

Um sistema de informação de sucesso é aquele que consegue produzir informação com qualidade, capaz de ser percebida e utilizada pelos seus consumidores, transformando-a num factor de sucesso organizacional. Contudo, nem sempre é fácil avaliar a qualidade da informação e a qualidade de um sistema de informação.

Este trabalho propôs-se responder à questão se *será possível avaliar a qualidade de um sistema de informação pela avaliação da qualidade dos seus dados*. Para tal foi necessário, através da revisão bibliográfica, identificar o significado de qualidade dos dados em sistemas de informação, perceber a influência da qualidade dos dados nos órgãos de decisão e conseqüentemente nas políticas da organização e determinar o modo como a qualidade dos dados pode influenciar o desempenho de um sistema de informação.

Com o trabalho de investigação realizado pretendeu-se encontrar uma explicação teórica para o fenómeno da qualidade dos sistemas de informação e da sua ligação com a qualidade da informação, e encontrar uma ferramenta que permita ao auditor avaliar a qualidade do sistema de informação no seu todo através da avaliação da qualidade da informação que este é capaz de tratar e produzir. Pela análise objectiva das bases de dados, o coração do sistema de informação, é possível verificar a existência de problemas de qualidade aos outros níveis da qualidade do sistema.

Com base na análise dos dados recolhidos de estudos efectuados anteriormente, e após uma reflexão crítica foi possível apresentar um modelo de avaliação de um sistema de informação que contém as quatro categorias de qualidade: intrínseca, de representação, contextual e de acessibilidade. A utilização do modelo implica começar pela avaliação

objectiva da qualidade intrínseca dos dados e pressupõe que ao estarmos a avaliar esta categoria estamos simultaneamente a avaliar dimensões das outras categorias. Apenas quando não são encontrados problemas nesta categoria se passará a avaliar a categoria de qualidade de representação. Quando, por sua vez, esta não apresentar problemas então poderá passar-se à qualidade contextual e, nas mesmas circunstâncias à qualidade de acessibilidade. Estas categorias, ao contrário da intrínseca deverão ser analisadas com recurso não só a métricas objectivas como também à avaliação subjectiva dos utilizadores.

A qualidade intrínseca e a qualidade contextual estão directamente ligadas à qualidade da informação, enquanto a qualidade de representação e a qualidade de acessibilidade estão relacionadas com a qualidade do sistema. Desta forma, poder-se-á afirmar que se não estiver garantida a qualidade intrínseca da informação o sistema não terá qualidade, e não será necessário passar para a avaliação das outras categorias para concluir que estas também apresentam problemas uma vez que estes se reflectem nos dados.

Foi apresentado um exemplo teórico que procura mostrar, ainda que teoricamente, a aplicabilidade do modelo conjuntamente com a metodologia de avaliação de dados DQA (Data Quality Assessment), recorrendo à taxonomia de problemas de qualidade dos dados apresentada por Oliveira et al (2005) para a avaliação objectiva da qualidade intrínseca.

Contudo, ainda existem algumas áreas que se relacionam com o modelo de avaliação de sistemas de informação proposto nesta investigação e que carecem de maior investigação que poderá ser realizada em trabalhos futuros, nomeadamente:

- ✓ Perceber as medidas reais de afectação de umas dimensões de qualidade para as outras. Embora este trabalho tenha permitido concluir que existe efectivamente uma influência da qualidade nos níveis mais baixos para a qualidade nos níveis superiores, seria interessante perceber em concreto a quantidade de afectação;
- ✓ Desenvolver formas de avaliar objectivamente as dimensões de representação, contextual e de acessibilidade, de modo a que o auditor não fique dependente unicamente da avaliação subjectiva dos utilizadores para efectuar uma avaliação destas dimensões.

O trabalho do auditor informático, ao nível da qualidade dos sistemas de informação, é muitas vezes influenciado pela falta de um reconhecimento geral sobre aquilo que são

as boas práticas nesta matéria e de ferramentas que permitam garantir que essas boas práticas existem, são reconhecidas e estão implementadas nos sistemas de informação.

Este trabalho pretendeu ser mais um contributo para enquadrar teoricamente aquilo que muitas vezes é realizado de forma empírica e contribuir com uma metodologia que tem como base muitos trabalhos de reconhecido valor, que têm sido efectuados em investigação na área da qualidade da informação.

BIBLIOGRAFIA

- [Ballou e Pazer, 1985] Ballou, Donald P.; Pazer, Harold L. – Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems. *Management Science*, Vol. 31, nº 2 (Feb. 1985), p. 150 a 162.
- [Batini et al, 2009] Batini, C.; Cappiello, C.; Francalanci, C.; Maurino, A. – Methodologies for Data Quality Assessment and Improvement. *ACM Computing Surveys*, Vol. 41, nº 3, Article 16 (Jul. 2009), p. 16 a 52.
- [Bovee, Srivastava e Mak, 2003] Bovee, M.; Srivastava, Rajendra P.; Mak, B. - A Conceptual Framework and Belief-Function Approach to Assessing Overall Information Quality. *International Journal of Intelligent Systems*. Vol. 18, nº 1 (Jan. 2003), p. 51 a 74.
- [Carrol e Swatman, 2000] Carroll, J.M.; Swatman, P.A. – Structured-Case: a Methodological Framework for Building Theory in Information Systems Research. *European Journal of Information Systems*. Vol. 9 (2000), p.235 a p.242.
- [Choo, 2003] Choo, Chun Wei - *Gestão da Informação para a Organização Inteligente*. Lisboa: Editorial Caminho, 2003.
- [DeLone e McLean, 1992] DeLone, William H.; McLean, Ephraim R. - Information Systems Success: The quest for the dependent variable. *Information Systems Research*. Vol. 3, nº 1 (1992) p.60 a 95.
- [Deming, 1986] Deming, E.W: *Out of the Crisis*; Center for Advanced Engineering Study, MIT; Cambridge, MA; 1986.
- [Esteves, Ramos e Carvalho, 2007] Esteves, José; Ramos, Isabel; Carvalho, João – Use of grounded theory in information systems area: an exploratory analysis; *European Conference on Research Methodology for Business Management*. (2007), p.129 a 136.

- [Glaser e Strauss, 1967] Glaser, B.; Strauss, A.: *The Discovery of Grounded Theory*; Aldine Press; Chicago; 1967.
- [Laudon e Laudon, 2009] Laudon, Kenneth C.; Laudon, Jane P. - *Essencial of Management Information System*. Eighth Edition. Upper Saddle River, New Jersey: Prentice-Hall, 2009.
- [Mason, 1978] Mason, R. O. – Measuring Information Systems Output: A Communication Systems Approach. *Information & Management*. Vol. 1, nº 5 (1978). P. 219 a 234.
- [Moresi, 2000] Moresi, Eduardo Amadeu Dutra – Delineando o Valor do Sistema de Informação de uma Organização. *Ciência da Informação*. Vol. 29, nº 1 (Jan./Abr. 2000), p.14 a 24.
- [Nelson, Todd e Wixom, 2005] Nelson, R. Ryan; Todd, Peter A.; Wixom, Barbara H. – Antecedents of Information and System Quality: An Empirical Examination Within the Context of Data Warehousing. *Journal of Management Information Systems*. Vol. 21, nº 4 (2005), p.199 a 235.
- [Oliveira e Amaral, 1999] Oliveira, João Nuno; Amaral, Luís Alfredo - *O papel da qualidade da informação nos sistemas de informação*. 1999. Comunicação proferida na “Conferência Especializada em Sistemas e Tecnologias de Informação”, que decorreu em Outubro de 1999 nas instalações da Universidade Católica Portuguesa, em Lisboa. Acessível na Universidade do Minho, Braga, Portugal (<http://repositorium.sdum.uminho.pt> – acedido em Dezembro/2008).
- [Oliveira et al, 2005] Oliveira, Paulo; Rodrigues, Fátima; Henriques, Pedro; Galhardas, Helena – A Taxonomy of Data Quality Problems, *Proceedings of the international workshop on data and information quality*, p. 219 a 233, Porto, Junho 2005.
- [Orr, 1998] Orr, Ken – Data Quality and Systems. *Communications of the ACM*. Vol. 41, nº 2 (Fev. 1998), p. 66 a 71.
- [Pipino et al, 2002] Pipino, Leo L.; Lee, Yang W.; Wang, Richard Y. – Data Quality Assessment, *Communications of the ACM*, Vol. 45, nº 4 (April 2002), p. 211 a 218.

- [Redman, 2005] Redman, Thomas C.; Ph. D. Navesink Consulting Group – A comprehensive approach to Data Quality Governance, *Data Quality Governance, White Paper – Final* (2005).
- [Redman, 1998] Redman, Thomas C. – The Impact of Poor Data Quality on the Typical Enterprise. *Communications of the ACM*, vol. 41, nº 2 (Feb. 1998), p.79 a 82.
- [Redman, 1996] Redman, Thomas C. – *Data Quality for the Information Age*. 1ª edição. Artech House Publishers, 1996.
- [Scannapieco et al, 2005] Scannapieco, M., Missier, P., Batini, C. – Data Quality at a Glance, *Datenbank-Spektrum*, Vol. 14 (2005), p. 6-14.
- [Scannapieco e Catarci, 2002] Scannapieco, M., Catarci, T. - Data Quality under the Computer Science perspective, *Journal of "Archivi & Computer"*, Vol. 2 (2002).
- [Seddon, 1997] Seddon, Peter B. - A Respecification and Extension of the DeLone and McLean Model of IS Success. *Information Systems Research*. Vol. 8, nº 3 (1997), p.240 a 253.
- [Shannon e Weaver, 1949] Shannon, C. E.; Weaver, W. – *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1949.
- [Strauss e Corbin, 1990] Strauss, A.; Corbin, J.: *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*; SAGE Publications; Newbury Park; CA; 1990.
- [Strong, Lee e Wang 1997a] Strong, Diane M.; Lee, Yang W.; Wang, Richard Y. – 10 Potholes in the Road to Information Quality. *Computer Society*; Vol. 30, nº 8 (Aug 1997a), p. 38 a 46.
- [Strong, Lee e Wang, 1997b] Strong, Diane M.; Lee, Yang W.; Wang, Richard Y. – Data Quality in Context. *Communications of the ACM*; Vol. 40, nº 5 (May 1997b), p. 103 a 110.

- [Symon e Cassel, 1998] Symon, G.; Cassel, C.: *Qualitative Methods and Analysis in Organizational Research. A Practical Guide*; p. 29 a p. 47; SAGE Publications; London; Thousand Oaks; New Delhi; 1998.
- [Wand e Wang, 1996] Wand, Y.; Wang, Richard Y. – Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM*, Vol. 39, nº 11 (Nov. 1996), p. 86 a 95.
- [Wang, 1998] Wang, Richard Y. – A product perspective on total data quality management. *Communications of the ACM*, Vol. 41, nº 2 (Feb. 1998), p. 58 a 65.
- [Wang, Kon e Madnick, 1993] Wang, Richard Y.; Kon, H. B.; Madnick, S. E. – Data quality requirements analysis and modeling. *Proceedings of the 9th International Conference on Data Engineering*, pp. 670 a 677. Vienna: IEEE Computer Society Press (Apr. 1993).
- [Wang, Reddy e Kon, 1995] Wang, Richard Y.; Reddy, M. P.; Kon, Henry B. – Toward quality data: An attribute-based approach. *Journal of Decision Support Systems*, nº 13 (1995), p. 349 a 372.
- [Wang e Strong, 1996] Wang, Richard Y.; Strong, D. – Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, Vol. 12, nº 4 (1996), p. 5 a 34.

