

**MESTRADO**  
**GESTÃO DE SISTEMAS DE INFORMAÇÃO**

**TRABALHO FINAL DE MESTRADO**  
**DISSERTAÇÃO**

**O STATISTICAL DATA WAREHOUSE DO INE – UMA OBSERVAÇÃO  
PARTICIPATIVA**

**DUARTE LUÍS SALES VASQUES**

**OUTUBRO - 2019**

**MESTRADO**  
**GESTÃO DE SISTEMAS DE INFORMAÇÃO**

**TRABALHO FINAL DE MESTRADO**  
**DISSERTAÇÃO**

**O STATISTICAL DATA WAREHOUSE DO INE – UMA OBSERVAÇÃO  
PARTICIPATIVA**

**DUARTE LUÍS SALES VASQUES**

**ORIENTAÇÃO:**  
**PROFESSOR JOÃO PAULO VICENTE JANELA**

**OUTUBRO - 2019**

*Para a minha mãe,  
por todas as  
superações  
supostamente  
inalcançáveis;  
aqui estamos.*

## RESUMO

Vivemos na era dos dados, onde estes são equiparados ao valor do petróleo e quem detém o poder da informação detém enorme saber sobre como uma sociedade funciona em todas as suas vicissitudes. A presente dissertação aborda aquela que é, provavelmente, a base de dados mais importante que um país pode ter: a base de dados do seu instituto nacional de estatística oficial, também denominada de Statistical Data Warehouse. Pretendeu-se estudar como está desenhada a arquitetura do Statistical Data Warehouse do Instituto Nacional de Estatística português, como funciona o processo de produção de estatísticas oficiais e como é que os dois se interligam, procurando elucidar-se a dinâmica entre utilizadores, tecnologia e processos. Para tal, conduziu-se um estudo qualitativo, com forte ênfase na observação participativa, com recurso a análise documental, entrevistas com especialistas e trabalho próprio do investigador realizado no objeto de estudo. Retira-se a importância do avanço tecnológico e da digitalização na produção de estatísticas oficiais, não obstante os seus custos em termos financeiros e de recursos humanos. Denota-se, também, a importância de se envolver todo o instituto nacional de estatística em torno do novo ambiente de produção estatística, assinalando-se resultados animadores na sua adoção e desenvolvimento, sem esquecer a formação que é necessária efetuar sobre os seus utilizadores. Este estudo incide sobre um nicho de Data Warehouse, demonstrando o modo de funcionamento da sua arquitetura e como é que o modelo de produção de estatística oficial se interliga com a mesma, bem como com as áreas funcionais do instituto.

**PALAVRAS-CHAVE:** Data Warehouse; Statistical Data Warehouse; Processo Estatístico; Estatística; Dados; Metadados

## ABSTRACT

We're living in the age of data, where data are equated to the oil' value and who holds the power of information holds a huge knowledge about how a society works and all its' vicissitudes. This dissertation approach what is, probably, the most important database that a country may have: the database of its national statistics institute, also called a Statistical Data Warehouse. It was intended to study how the Statistical Data Warehouse' architecture of the Portuguese National Statistics Institute is designed, how does the official statistics production process works and how the two intertwine, trying to clarify the dynamics between users, technology and processes. For such purpose, a qualitative study was conducted, with strong emphasis no participative observation, utilizing documental analysis, interviews with experts and the own researcher' work on the studied object. The importance of a technological advance and digitalization in the production of official statistics is duly noted, notwithstanding the financial and human resources costs it carries. It's worth noting, also, the importance of engaging all the national statistics institute around the new statistics production environment, which reveals promising results in terms of its adoption and development, without forgetting the need for formation that its' users require. This dissertation approaches a niche of the Data Warehouse, demonstrating how its architecture works and how the official statistics production model connects with it, as well as with the key departments of the institute.

KEYWORDS: Data Warehouse; Statistical Data Warehouse; Statistical Process; Statistics; Data; Metadata

## ÍNDICE

Resumo .....	i
Abstract .....	ii
Índice de Figuras .....	v
Índice de Tabelas .....	v
Agradecimentos .....	vi
Glossário .....	vii
1. Introdução e Enquadramento .....	1
2. Revisão de Literatura .....	4
2.1. Data Warehouse .....	4
2.1.1. Necessidade de ter Data Warehouse .....	4
2.1.2. Adoção, Vantagens e Desvantagens de Data Warehouse .....	5
2.1.3. Arquitetura de Data Warehouse .....	7
2.2. Instituto Nacional de Estatística .....	12
2.2.1. O Processo de Produção Estatística .....	13
2.2.2. Organograma e Áreas Funcionais do INE .....	16
2.3. Statistical Data Warehouse .....	17
3. Metodologia .....	20
4. Análise de Resultados .....	22
4.1. Arquitetura do Statistical Data Warehouse .....	22
4.2. Interligação do MPPE com o Statistical Data Warehouse .....	25
4.3. Prós e Contras da Solução Implementada de Statistical Data Warehouse no INE .....	30
5. Conclusões e Investigação Futura .....	34
Referências Bibliográficas .....	37

Anexos .....	40
Anexo 1 - Organograma do Instituto Nacional de Estatística .....	40

## ÍNDICE DE FIGURAS

Figura 1 - Esquema de um Enterprise Data Warehouse Fonte: (Krishnan, 2013, p. 148) .....	9
Figura 2 - Fases e Processos do MPPE .....	14
Figura 3 - Arquitetura do Statistical Data Warehouse .....	22
Figura 4 - Organograma do Instituto Nacional de Estatística a Junho de 2019 Fonte: (INE, 2019) .....	40

## ÍNDICE DE TABELAS

Tabela 1 - Fatores de implementação/adoção das abordagens top-down e bottom-up de Data Warehouse numa organização.....	11
Tabela 2 - Prós e Contras da Solução Implementada de Statistical Data Warehouse no INE.....	33

## AGRADECIMENTOS

Primeiramente, agradeço a todos os meus colegas de faculdade, pelo apoio incalculável ao longo dos semestres deste mestrado. Por todos os trabalhos de grupo, conversas sobre as matérias lecionadas e apontamentos trocados e discutidos entre nós, um profundo agradecimento.

De seguida, teço um agradecimento ao Professor João Janela, pelo seu apoio inicial em aceitar-me como seu orientando e toda a disponibilidade que sempre revelou no processo de elaboração de tese.

Ao ISEG, faculdade que foi a minha segunda casa e uma inesgotável fonte de conhecimento ao longo da minha licenciatura e do meu mestrado. Uma casa que sei que poderei sempre visitar e ser acolhido como se fosse o primeiro dia.

Aos colegas Rogério Reis, Sónia Quaresma e Carlos Gonçalves, onde cada um contribuiu mais do que alguma vez lhes pedi, e sem os quais seria absolutamente impossível redigir a presente dissertação.

Por fim, aos meus amigos, pela paciência em me suportarem com as saudáveis dores de cabeça que esta dissertação me deu, e à minha mãe, que, por amor incondicional, deu tudo de si e foi um corrimão para que eu pudesse subir mais esta escada na minha vida.

## GLOSSÁRIO

DM - Departamentos de Matéria

DMSI - Departamento de Metodologia e Sistemas de Informação

DREM - Direção Regional de Estatística da Madeira

DRGD - Departamento de Recolha e Gestão de Dados

GSBPM - *General Statistical Business Process Model*

INE - Instituto Nacional de Estatística

MPPE - Modelo do Processo Produtivo de Estatística

ODS - *Operational Data Stores*

OLAP - *Online Analytical Processing*

OLTP - *Online Transaction Processing*

SEE - Sistema Estatístico Europeu

SEN - Sistema Estatístico Nacional

SREA - Serviço Regional de Estatística dos Açores

## 1. INTRODUÇÃO E ENQUADRAMENTO

É famosa a expressão “dados são o novo petróleo”, cunhada por Clive Humby em 2006 e tornada popular graças a uma publicação do jornal *The Economist* em 2017 (Flender, 2019). Esta expressão carrega um peso enorme graças à crescente digitalização da informação em várias vicissitudes da vida. Estima-se que, em 2017, 48% da população mundial se encontrava conectada à Internet por uma ou outra via (União Internacional de Telecomunicações, 2017). Esta mesma agência da Organização das Nações Unidas estimava que esta percentagem circundava os 35% em 2011 e os 18% em 2006 (União Internacional de Telecomunicações, 2011).

Daqui decorre a importância magnânima de se estudar os dados e a gestão dos mesmos em contexto organizacional. Saber a relação de empresas, institutos e outras organizações com a gestão dos dados, a sua utilização e de que maneiras as utilizam para desenvolverem a sua atividade é de extrema relevância nos dias que correm.

No meio de todos os tipos de organizações, existe um que faz dos dados a sua atividade principal, ou o seu *core business*: o instituto nacional de estatística de cada país.

De facto, um órgão oficial, com legitimidade estatal para produzir estatísticas com informação útil e relevante para a sociedade sobre cada aspeto do seu país, tem de ter um cuidado extremo no modo como lida com os dados que recebe para produzir estatística, como os gere e como possibilita que os técnicos produtores destas estatísticas oficiais tenham as melhores competências para executarem essa tarefa de enorme responsabilidade, com a maior fiabilidade e veracidade possíveis nas metodologias estatísticas, na recolha dos dados na produção e divulgação de produtos estatísticos oficiais.

Sendo assim, o problema central que esta dissertação pretende abordar é justamente o estudo da base de dados do Instituto Nacional de Estatística português e a forma como gerem os dados neste Data Warehouse.

Debruçando-me sobre este problema, discorri várias questões centrais de investigação, que irei procurar esclarecer ao longo da presente dissertação:

- Como está desenhada a base de dados do Instituto Nacional de Estatística em termos de arquitetura informática?
- Qual o conhecimento já existente sobre esta base de dados específica?
- De que maneira é que a base de dados do Instituto Nacional de Estatística se interliga com os processos de produção de estatística oficial?
- Qual a recetividade dos utilizadores a esta base de dados, nomeadamente dos funcionários do próprio Instituto Nacional de Estatística?
- Quais as vantagens e desvantagens, ou os prós e contras, decorrentes da solução implementada no Instituto Nacional de Estatística?

Esta área de conhecimento, embora à partida possa parecer vasta, existindo imensos autores famosos que investigaram a matéria de base de dados como William Inmon ou Ralph Kimball, é, na verdade, bastante parca em termos de trabalhos desenvolvidos, pois incide sobre a especificidade do que é uma base de dados no contexto de um instituto nacional de estatística. Como tal, a presente dissertação pretende oferecer um contributo sobre esta área de conhecimento, nomeadamente no contexto português do Instituto Nacional de Estatística, e dar a conhecer ao público em geral como funciona a dinâmica entre a base de dados estatística, a produção de estatísticas oficiais e os vários utilizadores que com ela interagem, por via de uma observação participativa que consiga analisar diretamente esta dinâmica e as suas interligações, como será melhor explicado no capítulo 3 desta dissertação.

No capítulo 2 proceder-se-á a uma revisão bibliográfica sobre o que é um Data Warehouse, porquê a necessidade das organizações a adotarem e implementarem, quais as vantagens e desvantagens decorrentes desta implementação e, por fim, quais as abordagens possíveis de implementação

de um Data Warehouse numa organização. Ademais, dar-se-á a conhecer o que é o Instituto Nacional de Estatística, qual o seu contexto na produção de estatísticas oficiais quer a nível nacional quer a nível europeu, bem como explicar todo o processo de produção de estatísticas oficiais e o que desenvolvem certos elementos-chave dentro das áreas funcionais do Instituto. Por fim, estudar-se-á qual o conhecimento existente sobre uma base de dados estatística, enaltecendo diferenças para com uma base de dados “típica” e encontrando características-chave deste tipo de base de dados.

No capítulo 3 explicarei a metodologia que adotei para desenvolver a presente dissertação e quais as vantagens da metodologia adotada.

No capítulo 4, desenvolver-se-á o conhecimento e procurar-se-á responder às questões centrais de investigação e solucionar o problema central desta dissertação, explicando como está desenhada a arquitetura da base de dados estatística, qual a dinâmica entre a base de dados, os utilizadores e os processos e indicando prós e contras da solução implementada no instituto nacional de estatística português.

Por fim, o capítulo 5 encerra a dissertação com as conclusões principais a retirar deste conhecimento desenvolvido, quais as limitações encontradas aquando a prossecução do mesmo e que pontos futuros devem ser posteriormente abordados de modo a poder avançar o conhecimento nesta área.

## 2. REVISÃO DE LITERATURA

### 2.1. *Data Warehouse*

Segundo a definição clássica de Bill Inmon (Hoffer, et al., 2016), um Data Warehouse consiste num conjunto de dados históricos, integrados e consolidados, provenientes de diversas fontes internas a uma organização, que servirá, depois, como uma base de apoio à tomada de decisão dos utilizadores de topo de uma organização.

#### 2.1.1. *Necessidade de ter Data Warehouse*

O Data Warehouse ajuda a fechar o *gap* entre sistemas operacionais e sistemas informacionais dentro de uma organização (Hoffer, et al., 2016).

Sistemas operacionais são sistemas de base de dados que lidam com operações correntes de uma organização - vendas, fichas de clientes, pesquisas simples de informação -, sendo operações de âmbito muito restrito e habitualmente executados por funcionários num nível meramente operacional ou intermédio, retirando-se uma impressão apenas ao que se passa *naquele momento* dentro da organização, exigindo-se que essa informação esteja disponível imediatamente.

Sistemas informacionais lidam com operações de âmbito muito mais largo e complexo, sendo habitualmente executados por administradores, analistas, consultores ou gestores de topo, exigindo um acesso fiável a dados históricos já previamente tratados e armazenados para melhor poder suportar o processo de tomada de decisão estratégica.

A consulta da ficha de um cliente, a encomenda de resmas de papel ao fornecedor habitual ou o número de t-shirts disponíveis em armazém são considerados exemplos de operações correntes; logo, pertencem ao domínio de sistemas operacionais. A consulta de vendas por produto e por região do país, um estudo de previsão dos horários de maior afluência a uma determinada loja ou a análise do impacto da nossa marca nas redes sociais já requer uma maior complexidade no acesso e tratamento dos dados, portanto entram no domínio de sistemas informacionais.

Este *gap* torna-se pertinente ser tratado devido à necessidade de se ter uma imagem de toda a informação presente numa organização disponível de modo pronto e acessível. Com o decorrer da atividade de uma organização, imensas quantidades de informação vão sendo introduzidas ou recolhidas para os vários sistemas de informação que a mesma possui. Esta informação acaba por não ser reunida num único ponto central, o que complica imenso quando se pretende cruzar a informação presente nos vários sistemas ao longo de todas as suas dimensões. A situação fica ainda mais complicada quando se compreende que esta informação não é uniformizada: num sistema, uma data pode estar redigida no formato “xx-yy-zzzz” e no outro sistema pode estar no formato “xx/yy/zz”.

Aqui entra o maior propósito do Data Warehouse: a consolidação e integração de dados provenientes de várias fontes, formatando-os de modo a harmonizar toda a informação presente numa organização e tornar mais eficiente o processo de tomada de decisão por parte dos vários colaboradores (Hoffer, et al., 2016).

### **2.1.2. Adoção, Vantagens e Desvantagens de Data Warehouse**

As organizações rapidamente percebem o potencial de se ter um Data Warehouse e alavancar as suas ferramentas de modo a se tornarem mais competitivos e melhorarem as suas relações com clientes, fornecedores e mesmo entre os seus colaboradores.

O mais recente inquérito da Snowflake Computing, uma empresa especializada em Data Warehouse com foco na *cloud*, a 315 *data managers* revela que 99% dos respondentes considera Data Warehouse como uma tecnologia muito importante para as suas actividades de negócio, sendo que 72% admitem estarem a aumentar o seu investimento nesta área. Contudo, 97% reconhecem fortes dificuldades em implementar um Data Warehouse nas suas organizações (Snowflake Computing, 2015).

Obter um melhor acesso aos dados, em termos de rapidez e leitura; melhorar a qualidade dos processos de *business intelligence* de modo a obter

informação analítica mais útil para a tomada de decisão dos gestores de topo de uma organização, potenciando a competitividade da mesma no mercado; ou proporcionar um melhor serviço aos clientes, fornecedores ou outros *stakeholders* através de um sistema de informação harmonizado e fiável ao longo de todos os processos e departamentos de uma organização são exemplos das vantagens que a implementação de um sistema de Data Warehouse pode fornecer a qualquer organização (Thomas, 2018).

Por outro lado, existirem fatores condicionantes ao processo de adoção da tecnologia de Data Warehouse. Por ser um projeto de enorme âmbito, existe uma grande complexidade no planeamento do mesmo, com a necessidade de se ter de envolver todos os colaboradores de uma organização.

Existe um enfoque especial no envolvimento e empenho dos gestores de topo, pois estes carregam a responsabilidade de apoiar este projeto, ajudam a definir concretamente os objetivos que o mesmo deverá ter para a organização e criam metas para que os mesmos sejam atingidos, mitigam a resistência à mudança presente naturalmente em todos os colaboradores, obtêm a assistência necessária em termos de capital financeiro e de conhecimento para a implementação do projeto e facilitam a incorporação desta tecnologia nos processos organizacionais (Grover, 1998).

Outros fatores a ter em consideração aquando a adoção de um Data Warehouse consistem na necessidade de identificar as necessidades internas da organização, para averiguar se um sistema de informação tão complexo como o Data Warehouse é realmente necessário para a organização, e, se sim, qual a melhor forma de o implementar; integrar todos os colaboradores da organização no mesmo projeto, mitigar a resistência à mudança e assegurar formação nas novas ferramentas de trabalho que eventualmente surjam com a implementação de um Data Warehouse; fazer corresponder os processos organizacionais com a arquitetura do Data Warehouse, para assegurar que não existem erros de atributos nem falhas de informação, ou seja, que existe uma interligação forte em cada etapa dos processos organizacionais com o sistema de informação de Data Warehouse implementado; por fim, pesar bem o custo-benefício de um projeto desta magnitude, considerando que o mesmo é

complexo, por vezes extremamente dispendioso financeiramente e, precisamente por consumir tantos recursos da organização que a deseje implementar, é considerado um projeto de alto risco.

### ***2.1.3. Arquitetura de Data Warehouse***

Existem duas abordagens de implementação de um sistema de Data Warehouse conhecidas na literatura. A primeira é denominada de abordagem bottom-up, desenvolvida por Ralph Kimball, e a segunda é denominada de abordagem top-down, desenvolvida por Bill Inmon (Breslin, 2004). Podemos considerar estas abordagens mais numa índole teórica, de filosofia de gestão de dados dentro duma organização, porque não existe uma abordagem necessariamente melhor que a outra. Ambas serão adequadas conforme os objetivos da organização (Araoj, 2017).

Na abordagem bottom-up, o sistema de Data Warehouse é constituído por vários pequenos data marts, com o propósito de servirem apenas para um leque escasso de objetivos, isto é, para utilizações específicas e bem-definidas para cada grupo de utilizadores dentro de uma organização. Numa organização, pode ser típico encontrar-se um data mart de vendas, outro de clientes, outro de marketing, outro do departamento financeiro. Este conjunto de data marts consiste no Data Warehouse da organização.

Um data mart é um Data Warehouse de âmbito mais limitado, obtido quer através de uma visualização filtrada da informação já carregada no Data Warehouse quer através de processos separados de extração, transformação e carregamento de dados. Na abordagem bottom-up, o data mart é obtido nesta última forma. Isto cria vários problemas em termos de estrutura do sistema de Data Warehouse, porque é necessário uniformizar as várias dimensões dentro dos vários data marts. É necessário que, por exemplo, o número de contribuinte dum cliente XPTO seja igual quer no data mart de vendas quer no data mart de clientes. Se um colaborador da organização pretender cruzar data marts, para estudar, por exemplo, que tipo de produtos é que os clientes de Lisboa adquiriram nos meses de Junho e Julho, terá de uniformizar as

dimensões das datas, da localização, do número de identificação do cliente e do número de identificação dos produtos. São quatro dimensões a uniformizar por três data marts (o data mart dos produtos, o das vendas e o dos clientes).

Porém, a curto-prazo, apenas para realizar tarefas mais rotineiras ou de âmbito limitado, esta abordagem é a mais recomendada. Tem menos custos financeiros e menos esforço e complexidade de planeamento por toda a organização, porque a implementação de um Data Warehouse via esta abordagem é apenas direcionada para áreas individuais da organização e não existe nenhum esforço unificado de implementação por todas as áreas (George, 2012).

Contudo, a longo-prazo, torna-se necessário obter flexibilidade precisamente por este motivo de cruzar e uniformizar os vários data marts provenientes das várias vicissitudes da atividade de uma organização ao longo do decorrer da sua vida (Hoffer, et al., 2016). Não se conseguindo obter esta flexibilidade, esta visão única e holística da informação, a organização não conseguirá ter as melhores ferramentas analíticas para servir de suporte à tomada de decisão. Afectando a tomada de decisão, a organização poderá perder competitividade no mercado. Nos dias que correm, de abundante acesso a informação, é cada vez mais prudente ter ferramentas robustas de suporte à tomada de decisão, porque a concorrência estará a fazer precisamente o mesmo para se tentar posicionar melhor no mercado.

Nesta abordagem, os data mart são considerados independentes, porque cada um diz respeito apenas a cada área funcional da organização, sendo cada um criado por processos ETL diferentes. Processos ETL são processos de extração, transformação e carregamento dos dados.

A outra abordagem é denominada de abordagem top-down. Nesta abordagem, o sistema de Data Warehouse consiste numa inteira visão integrada e única de toda a informação presente numa organização. Ao invés de estar dividida por áreas funcionais, esta espelhará toda a organização, através de um Enterprise Data Warehouse, que é um Data Warehouse que centraliza e harmoniza todos os dados presentes numa organização,

disponibilizando acesso aos mesmos via data marts dependentes, que são criados através de uma visualização filtrada da informação já carregada neste Enterprise Data Warehouse e em função das necessidades analíticas dos utilizadores finais, como espelhado na Figura 1.

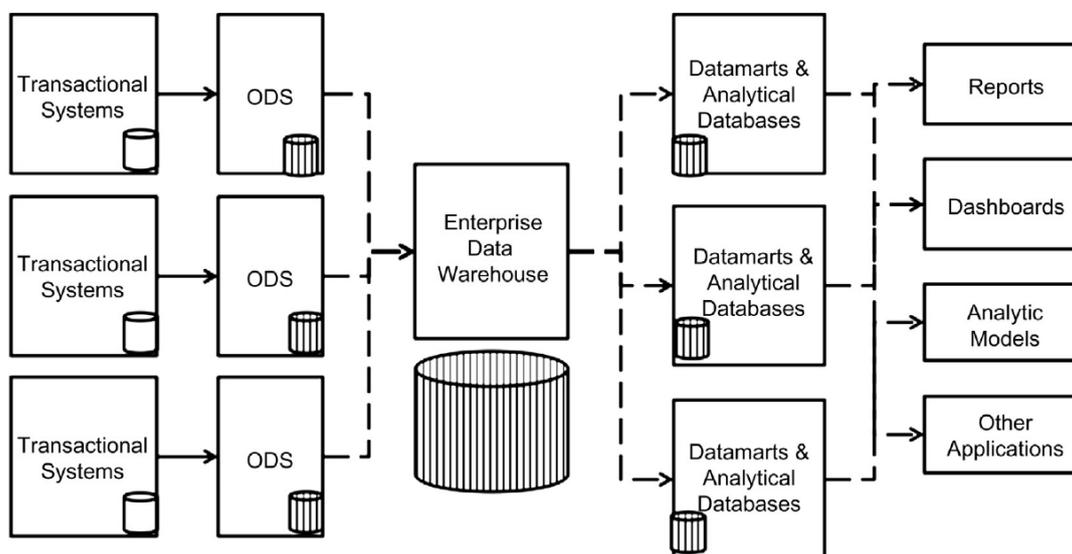


Figura 1 - Esquema de um Enterprise Data Warehouse

Fonte: (Krishnan, 2013, p. 148)

Como se observa na figura acima, denota-se a obtenção de dados na primeira coluna através dos sistemas operacionais, ou sistemas transacionais, que são gerados através das atividades rotineiras da empresa, as operações correntes mencionadas na página 4. A isto também se denomina de *Online Transaction Processing*, ou OLTP (Reddy, et al., 2010). À atividade de análise e visualização de grandes quantidades de informação para a tomada de decisão por parte de analistas e gestores de topo, que operam nas duas últimas colunas da figura acima, dá-se o nome de *Online Analytical Processing*, ou OLAP (Reddy, et al., 2010).

Na segunda coluna temos os *Operational Data Stores*, ou ODS, que são bases de dados temporária que irão agregar todos os dados operacionais provenientes do OLTP (Krishnan, 2013), onde poderão ocorrer várias atividades de limpeza e tratamento dos dados em nome da qualidade dos mesmos, de modo a estarem melhor preparados para serem armazenados no *Enterprise Data Warehouse*.

Assim, evitam-se alguns problemas resultantes da abordagem bottom-up, nomeadamente os problemas de escalabilidade que surgem com o decorrer da atividade da organização e das exigências dos utilizadores (mais atividade, mais entrada de dados, mais registos para processar), porque esta abordagem não requer tanta manutenção ao longo da sua vida comparativamente à outra abordagem, devido à enorme quantidade de planeamento especializada em criar um sistema de informação *Enterprise-wide*. Ademais, evitam-se ao máximo problemas de cruzamento e uniformização de dados entre diferentes data marts, porque esta definição é efetuada a priori, isto é, as dimensões são uniformizadas aquando a implementação do Data Warehouse e dos processos ETL, e não *on-the-fly*, ou seja, sempre que surja a necessidade de uniformização, como acontece na abordagem bottom-up. À partida, os números de identificação dos clientes, dos produtos, das localizações, os identificadores de tempo e outras chaves já estarão configurados entre si na arquitetura do Data Warehouse.

Contudo, por este tipo de abordagem exigir precisamente um planeamento e atenção afincada a cada processo dentro da organização, as implementações de Enterprise Data Warehouse são extremamente dispendiosas, quer em tempo, em recursos financeiros e em pessoas, sendo necessário, muitas vezes, contratar pessoas especializadas em gestão de dados para executar um projeto desta dimensão (Hoffer, et al., 2016, p. 422). Ademais, requer um envolvimento de toda a organização, a começar pelo compromisso dos gestores do topo até o reconhecimento das necessidades dos colaboradores, para ajudar a criar um Enterprise Data Warehouse que vá de encontro aos requisitos analíticos que a organização e os seus colaboradores têm.

Não existe uma abordagem totalmente certa nem outra totalmente errada. Ambas têm de ser devidamente estudadas e apreciadas de acordo com os objetivos da organização na sua gestão de dados e base de dados. A tabela 1 revela as principais diferenças das duas abordagens em termos de implementação e adoção, de acordo com Balasubramaniam (2015), Abramson (2004) e Rangarajan (2016).

Tabela 1 - Fatores de implementação/adoção das abordagens top-down e bottom-up de Data Warehouse numa organização

FACTORES DE IMPLEMENTAÇÃO/ADOÇÃO	ABORDAGEM TOP-DOWN	ABORDAGEM BOTTOM-UP
TEMPO DE IMPLEMENTAÇÃO	Consome imenso tempo (mais de 1 ano)	Relativamente pouco tempo
CUSTOS FINANCEIROS	Custos iniciais elevados; manutenção e projetos subsequentes de custos reduzidos	Custos iniciais reduzidos; manutenção e projetos subsequentes de custos elevados
CAPACIDADE TÉCNICA	Equipa especializada com técnicos de elevada capacidade técnica e informática; necessidade de envolver toda a organização	Equipa pouco especializada, apenas com capacidade técnica para uma implementação rápida; necessário envolver apenas partes da organização
INTEGRAÇÃO DE DADOS	Integrados ao longo da toda a organização, “Enterprise-wide”	Integrados ao longo de cada área funcional de uma organização, “business-wide”
CONSISTÊNCIA DOS DADOS	Sistema de fontes de dados pode ser diversificado e sofrer altas inconsistências	Sistema de fontes de dados tem de ser mais estável e consistente
ALTERAÇÕES NA ARQUITETURA	O desenho geral da arquitetura nunca se altera; o desenho já prevê pequenas alterações de modo a acomodar novas necessidades de informação	O desenho geral da arquitetura sofrerá imensas alterações ao longo do seu tempo conforme as necessidades imediatas de informação dos seus utilizadores

## ***2.2. Instituto Nacional de Estatística***

O Instituto Nacional de Estatística, doravante designado por INE, é o organismo oficial do Estado responsável por produzir informação estatística oficial de qualidade e relevante para toda a sociedade.

O INE enquadra-se no Sistema Estatístico Nacional, doravante designado por SEN, estabelecido pela Lei nº22/2008, de 13 de Maio, enquanto órgão central de produção e difusão de estatísticas oficiais, que assegura a supervisão e coordenação técnico-científica do SEN. O SEN também é constituído pelo Conselho Superior de Estatística, que tem o objetivo de orientar e coordenar o SEN, tendo competência para, por exemplo, definir e aprovar as linhas gerais de atividade estatística oficial ou zelar pelo princípio do segredo estatístico, previsto no Artigo 6º do mesmo diploma, tendo autoridade para realizar auditorias nos outros organismos do SEN ou outras ações de fiscalização. O Banco de Portugal é outro membro do SEN, tendo competência delegada para a elaboração das estatísticas monetárias, financeiras, cambiais e da balança de pagamentos. Para as Regiões Autónomas da Madeira e dos Açores, temos os Serviços Regionais de Estatística das Regiões Autónomas, com competência para exercer atividade estatística oficial na sua região geográfica. Os Açores têm o Serviço Regional de Estatística dos Açores (SREA) e a Madeira tem a Direção Regional de Estatística da Madeira (DREM), onde ambas irão alimentar a produção de estatística oficial feita pelo INE no plano nacional, sob orientação técnica do INE. Por fim, incluem-se outras entidades com competências delegadas para produzirem estatística oficial, como o Gabinete de Estratégia e Planeamento ou a Direção Geral de Energia e Geologia.

Ademais, sendo Portugal um membro comunitário, o INE enquadra-se no Sistema Estatístico Europeu, doravante designado por SEE. O SEE é constituído pelo Eurostat, pelos Institutos Nacionais de Estatística dos 28 Estados-Membros da União Europeia, bem como pelos dos 4 países membros da Associação Europeia de Livre Comércio (Islândia, Liechtenstein, Noruega e Suíça), e as outras autoridades estatísticas com competências delegadas destes 32 países. O Eurostat é a autoridade estatística da União,

impulsionando uma harmonização de estatísticas a nível europeu, em articulação com as várias autoridades estatísticas nacionais, de modo a assegurar que as estatísticas oriundas destes países possam ser comparadas entre si. Por sua vez, os vários INEs irão ser o interlocutor único do Eurostat no seu país, representando o seu próprio sistema estatístico no SEE.

### ***2.2.1. O Processo de Produção Estatística***

Torna-se pertinente documentar todas as atividades e tarefas necessárias para a produção de estatística oficial, por forma a termos um instrumento de comunicação interna e externa sobre o que é necessário desenvolver no ciclo de produção estatística, uma ferramenta de padronização do processo e harmonização da terminologia a utilizar e um documento de referência quer para auditorias de qualidade quer para todo o planeamento da atividade estatística.

No âmbito destes objetivos, existe um documento a nível internacional chamado GSBPM (*Generic Statistical Business Process Model*), desenvolvido pela OCDE, UNECE e Eurostat, que atualmente se encontra na versão 5.1 e que procura desenvolver um padrão na documentação dos processos de produção estatística. Se várias instituições de estatística oficiais seguirem este padrão aquando a elaboração da sua documentação, tornar-se-á mais fácil a padronização da terminologia a utilizar e a partilha de metodologias e outros conhecimentos entre os vários organismos.

O Modelo do Processo Produtivo de Estatística (doravante MPPE) desenvolvido e utilizado pelo INE encontra-se alinhado com o padrão definido por este documento GSBPM.

O padrão vem caracterizar o processo de produção estatística dividindo-o em fases, processos e subprocessos. As fases retratam as grandes etapas de uma produção estatística; os processos retratam as actividades que visam desenvolver aquelas fases; e os subprocessos são as tarefas que compõem e visam executar cada atividade.

O MPPE vai mais além que o GSBPM, na medida em que estabelece as principais tarefas para cada processo, identifica quais as unidades orgânicas dentro do INE responsáveis pela execução de cada tarefa, bem como outras que intervenham no processo, e refere documentação de apoio e documentação a produzir.

Este modelo permite que certos subprocessos se possam interligar ou complementar entre si, não sendo este ciclo de produção estatística totalmente linear. Algumas tarefas interdependem-se, e existe alguma flexibilidade na execução das mesmas.

Na Figura 2 encontram-se descritas sucintamente as fases e processos do MPPE.

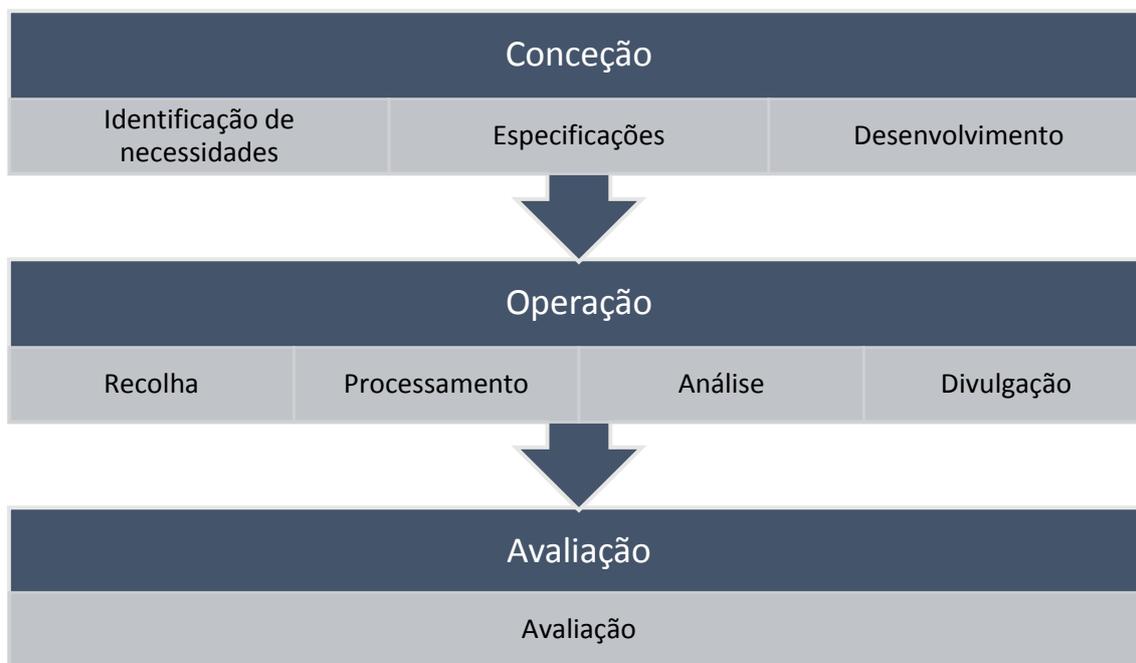


FIGURA 2 - FASES E PROCESSOS DO MPPE

A primeira grande fase de produção estatística, a conceção, inclui três processos de apoio ao desenvolvimento desta fase. O primeiro processo passa pela identificação de necessidades, onde se identificam os objetivos, conceitos gerais e fontes de dados da nova operação estatística, que surge como resposta a uma necessidade de informação, e prepara-se todo o processo produtivo. De seguida, vem a fase das especificações, onde se identificam variáveis, métodos de recolha, qual a amostra a ser inquirida,

bem como outras especificações metodológicas e informáticas necessárias à recolha, integração, tratamento e análise de dados. Por fim, esta etapa termina com o processo do desenvolvimento, onde se configura e testa todos os instrumentos que compõem o processo de produção, como os instrumentos de recolha, de análise ou de difusão, e realizam-se testes sobre o mesmo de modo a adequá-los para um ambiente operacional, onde a nova operação estatística já está lançada e pronta a ser desenvolvida. Este último processo ocorre ou aquando a criação de uma nova operação estatística ou aquando revisões na metodologia ou nas especificações da mesma.

A segunda fase de produção estatística é a operação, que inclui quatro processos. O primeiro processo é o da recolha, que abrange tarefas como a seleção da amostra de acordo com as especificidades identificadas no segundo processo da primeira fase, a pré-validação dos dados segundo regras de controlo de qualidade dos mesmos também previamente definidas, a realização de insistências para inquiridos que não tenham obedecido aos prazos legais definidos para resposta e outras tarefas de gestão de respondentes e unidades estatísticas.

O segundo processo desta fase é o processamento, onde se realiza a integração e preparação dos microdados, como, por exemplo, tarefas de classificação, codificação ou validação de certas respostas, bem como o apuramento dos dados agregados, com a sua quota-parte de cálculo estatístico já devidamente aplicada, por via do cálculo de ponderadores, dados agregados e variáveis derivadas, e finalizam-se os ficheiros de dados de modo a estarem preparados para o processo seguinte. As atividades realizadas neste processo são tipicamente programadas de modo a poderem ser executadas automaticamente com cada iteração das operações estatísticas.

O terceiro processo é a análise, onde, já com os dados devidamente validados e preparados para a análise estatística e produção de informação útil e relevante para toda a sociedade, procura-se interpretar e explicar os resultados obtidos, preparando os mesmos para o processo da difusão e garantindo sempre a confidencialidade dos mesmos.

O quarto e último processo desta fase é a divulgação, também conhecida por difusão, onde se realiza a divulgação da informação nos seus diferentes suportes e canais, desde o Portal *online* do INE até outros meios de comunicação, bem como se gere o apoio aos utilizadores desta informação e prestam-se serviços de esclarecimento dos produtos estatísticos que foram fornecidos.

A última grande etapa do MPPE é a avaliação, que inclui apenas um processo, também denominado de avaliação, onde se pretende reunir elementos e executar uma avaliação de qualidade sobre todo o processo estatístico que acabou de ocorrer, estabelecendo-se eventuais planos de ação de melhoria. Esta avaliação deve ocorrer pelo menos no fim do processo de produção estatística para cada operação estatística; isto é, após a divulgação dos resultados de um dado inquérito, deve-se proceder à avaliação da iteração da operação estatística em que se encontra no ciclo da produção estatística.

### ***2.2.2. Organograma e Áreas Funcionais do INE***

O organograma funcional do INE reparte-se em vários departamentos. No Anexo 1 encontraremos o organograma completo do INE. Contudo, importa frisar apenas três áreas funcionais, com grande importância para a produção estatística.

A primeira área é o Departamento de Recolha e Gestão de Dados, doravante DRGD. O DRGD, como o próprio nome indica, incide sobretudo sobre o processo de recolha, preocupando-se com a correta recolha dos dados junto dos respondentes, bem como a sua respetiva validação pré-processamento, isto é, assegurar que as respostas são válidas e corretas e não contêm erros que possam afetar a qualidade dos dados.

A segunda área é o Departamento de Metodologia e Sistemas de Informação, doravante DMSI. Este departamento terá como principais funções assegurar a máxima qualidade de padrão estatístico, seguindo as mais recentes e corretas metodologias de estatística, bem como controlará toda a

arquitetura dos sistemas de informação, como, por exemplo, a base de dados (foco principal desta dissertação), ou a implementação de novo *software* para os técnicos de estatística.

A terceira área funcional consiste nos Departamentos de Matéria, também designados por DM, que incluem os três primeiros departamentos que se observam no organograma do Anexo 1. Estes departamentos estão divididos por áreas temáticas de estatística, e são estes departamentos que irão, de facto, realizar toda a análise e cálculo estatísticos, tais como calcular taxas de variação ou medianas ou análises entre várias dimensões. Estes departamentos também podem executar uma validação sobre os dados que recebem, assegurando, assim, uma contínua prossecução em nome da qualidade dos mesmos, de modo a se obter a informação mais fiel possível.

Embora se possa fazer, grosso modo, a repartição estanque destes departamentos por aquelas funções, a realidade acaba por não ser tão linear. Por exemplo, para a primeira fase do MPPE, a conceção, todos os três departamentos terão de trabalhar em conjunto para definir métodos de recolha, programar variáveis no sistema de informação e definir quais os *outputs* estatísticos para cada operação estatística. Na fase da divulgação, entrará o Departamento da Difusão, precisamente responsável por redigir as publicações e divulgá-las no portal *online* e por entre órgãos de comunicação social, mas também poderão entrar os outros departamentos para darem a sua contribuição em nome da estatística oficial.

### ***2.3. Statistical Data Warehouse***

Com a digitalização dos dados e acrescente necessidade de modernização dos processos, registou-se um esforço, quer a nível nacional quer a nível internacional, de estudar e operacionalizar um Data Warehouse que servisse as necessidades adequadas de um órgão de estatística oficial, ou seja, criar um sistema de dados simultaneamente robusto o suficiente e flexível o suficiente para permitir quer a correta utilização e transformação dos dados

com qualidade quer a entrada de novos dados, novos metadados e novas fontes de dados sem perturbar todo o processo de produção estatística.

Neste âmbito, surge o conceito de Statistical Data Warehouse. De acordo com (Centre of Excellence on Data Warehousing, 2017), um Statistical Data Warehouse refere-se a um Data Warehouse construído especificamente com o objetivo de servir de apoio à produção de estatísticas nacionais e internacionais. É uma base de dados estatísticos centralizada, que visa melhorar a capacidade dos institutos de estatística oficiais de trabalhar sobre quaisquer tipos de dados de modo a produzir informação relevante, de uma maneira eficiente e possibilitando a constante reutilização dos dados já presente dentro do Statistical Data Warehouse; criar relatórios e executar *queries* e algoritmos de validação de qualidade ou de cálculo estatístico.

Diferentemente de um Data Warehouse tradicional, o mais comumente utilizado noutras organizações, um Statistical Data Warehouse desempenha um papel crucial no *core business* de um instituto de estatística oficial, que é simplesmente produzir estatísticas oficiais. Enquanto uma empresa pode utilizar o seu Data Warehouse para servir de apoio à tomada de decisão, como se se tratasse de um órgão consultivo, um Statistical Data Warehouse é mesmo uma ferramenta crucial para a produção de estatísticas oficiais, sem a qual os processos de produção estatísticas, do modo como estão configurados, não podem ser executados.

Outra diferença essencial entre um Data Warehouse tradicional e um Statistical Data Warehouse é a base de utilizadores e as suas necessidades. Para um Data Warehouse tradicional, a base de utilizadores é usualmente pequena e as suas necessidades não são exigentes. Apenas para executar *queries* básicas em termos informáticos, consultar a lista de faturas de um respetivo cliente ou ter uma *overview* rápida de informação útil da organização, o Data Warehouse tradicional acaba por ser um sistema de base de dados que não tem de oferecer um conjunto vasto e complexo de ferramentas de modo a concretizar o seu objetivo, que é servir de suporte à tomada de decisão. Ademais, quão mais complexas se tornarem as operações que se pretendem realizar dentro de um Data Warehouse tradicional, menor

fica a base de utilizadores. É rara a organização que tem técnicos de estatística a executar análises ou códigos complexos, que, por vezes, demoram horas a serem calculadas, de modo a obter resultados importantes para a tomada de decisão.

Contudo, num Statistical Data Warehouse, é precisamente isto que acontece, a toda a hora. Dados estão constantemente a entrar e sofrem regras de formatação, validação e outras tarefas de tratamento no processo ETL; executam-se algoritmos, códigos e cálculos estatísticos sobre uma enorme quantidade de dados de modo a gerar *outputs* para a produção estatística; e é necessário ter sempre a informação disponível na hora quer para os técnicos de estatística quer para quem consulte os dados de fora, no portal oficial online dos institutos de estatística.

Logo, um Statistical Data Warehouse tem de ser um sistema de base de dados robusto, capaz de suportar esta quantidade enorme de códigos a serem executados ao mesmo tempo proveniente de vários utilizadores dentro do mesmo instituto de estatística. Ademais, o modelo de dados tem de estar muito bem construído, de modo a possibilitar a análise entre vários domínios de dados, como, por exemplo, poder-se analisar o número de desempregados em cada distrito de Portugal e cruzar estas duas dimensões, e tantas mais quantas se desejar, sem qualquer erro de ligação de dados.

Isto é possível apenas com uma gestão apropriada dos metadados. Os metadados são, de acordo com (Kimball, et al., 2008, p. 117), o ADN de um Data Warehouse, funcionando como uma “enciclopédia” (Kimball & Ross, 2002, p. 14) que vai definir todos os elementos do mesmo e como é que trabalham em conjunto. No fundo, descreve a ligação entre os vários conceitos, elementos, as várias classificações, definições e variáveis que são construídos no modelo de dados. Isto é imperativo para que se possam executar análises que cruzem várias variáveis, ou várias dimensões de dados, e se possa obter um *output* que, de outra maneira, não seria possível obter.

Assim também se satisfazem outras necessidades dos utilizadores. Uma é a possibilidade de se poder construir uma série histórica de dados. Tendo os

vários dados um «carimbo» temporal, se se tiver os metadados corretamente alinhados ao longo dos períodos definidos, é possível executar uma análise temporal e calcular, por exemplo, índices, tendências ou taxas de variação. Outra necessidade que fica igualmente satisfeita é a de informar os utilizadores finais - isto é, investigadores, estudantes, órgãos de comunicação social ou qualquer pessoa que consulte as publicações de estatística ou o portal oficial do instituto de estatística - sobre o que é que cada elemento significa. Qual a informação por detrás de cada variável, qual o significado de cada conceito ou qual a hierarquia de uma determinada classificação são questões que ficam satisfeitas com uma boa gestão dos metadados.

Assim, a gestão de metadados estatísticos tem de ser correta, fiável, consistente e estandardizada (Goossens, 2012, pp. 9-10), de modo a assegurar uma manutenção adequada de um Statistical Data Warehouse e garantir a execução dos processos estatísticos e a produção de informação estatística oficial.

No desenvolvimento da presente tese, estudar-se-á quais as tarefas são executadas dentro de um Statistical Data Warehouse ao longo de cada fase da produção estatística, dando-se a conhecer qual a arquitetura de Statistical Data Warehouse implementada no INE. Depois, procurar-se-á perceber qual o relacionamento entre um Statistical Data Warehouse e o Modelo de Processos de Produção Estatística, bem como qual o *software* utilizado e se as necessidades dos utilizadores são devidamente cumpridas para executar o processo estatístico. Por fim, analisar-se-á os prós e contras da solução implementada no INE, bem como qual foi a aderência dos utilizadores e quais os pontos a melhorar, se existirem.

### 3. METODOLOGIA

Por forma a recolher dados para esta dissertação, foi utilizada uma abordagem qualitativa, caracterizada por se basear na “perceção e compreensão humana” (Stake, 2010, p. 11) e por ser uma metodologia “interpretativa, baseada em experiências, situacional e humanística” (Stake,

2010, p. 31), nomeadamente na forma de uma observação participante. Como definiu Denzin (1978, p. 198), uma observação participante “combina simultaneamente a análise de documentos, a entrevista a sujeitos e informantes, a participação e observação direta e a introspeção”.

Para a presente dissertação, o investigador pretendeu estudar o contexto e a dinâmica do objeto de estudo - que é o Statistical Data Warehouse do INE -, não querendo apenas analisá-la individualmente ou componente a componente, separada de todos os sujeitos que a envolvem e que com ela interagem rotineiramente a vários níveis (Correia, 2009).

Para tal, o investigador decidiu proceder a uma forte análise documental, pesquisando várias publicações relevantes para o objeto de estudo, que é muito específico e sobre o qual existe pouca investigação desenvolvida - e a que existe é desenvolvida pelas próprias organizações, os INEs, que as utilizam.

Após recolhidos os primeiros documentos e uma primeira análise ter sido efetuada, estes documentos foram posteriormente analisados em entrevista com uma pessoa especialista nesta matéria, técnica superior especialista em estatística e estimada colega de trabalho, de modo a captar todo o seu conhecimento especializado sobre esta matéria e dar a conhecer toda uma realidade específica sobre o objeto de estudo aplicado ao contexto em questão: ao INE de Portugal e a toda a dinâmica do processo de produção estatística, com as suas fases e os utilizadores em cada fase.

Ademais, como já foi dado a entender no parágrafo anterior, o investigador também se encontra empregue no INE, pelo que uma grande dose de observação e participação direta foi utilizada, de modo a recolher uma quantidade de dados sobre o modo de funcionamento do objeto em estudo. Como este objeto de estudo insere-se num nicho da área de Data Warehouse, porque só se verifica a sua existência nos vários órgãos de estatística oficial que existem em cada país, este contacto direto e diário com o objeto de estudo e todo o contexto envolvente permitiu-me fomentar uma sensibilidade para esta matéria que, outrora, não a poderia ter obtido de modo algum.

## 4. ANÁLISE DE RESULTADOS

### 4.1. Arquitetura do Statistical Data Warehouse

A arquitetura do Statistical Data Warehouse é desenvolvida através de uma abordagem bottom-up, onde existem quatro camadas de gestão de dados, como exemplificado na figura 3.

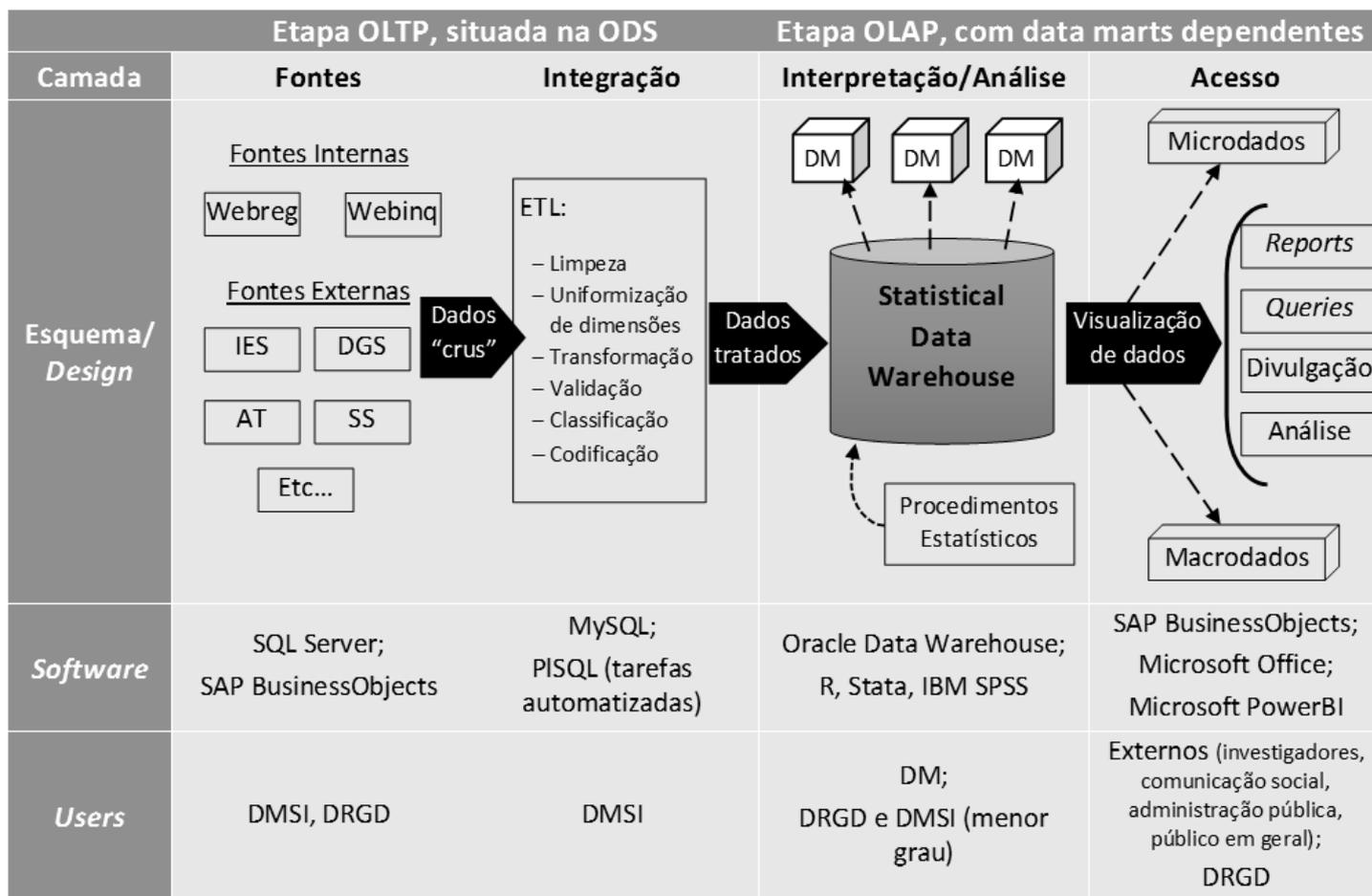


FIGURA 3 - ARQUITETURA DO STATISTICAL DATA WAREHOUSE

Esta arquitetura pretende estabelecer um fluxo de dados desde a recolha dos dados até à produção de matéria estatística oficial. É necessário, portanto, desenhar todo o sistema do Statistical Data Warehouse como integrando-se num modelo de produção estatística, de uma forma holística, envolvendo todos os diferentes departamentos e processos de produção estatística na criação do dito sistema.

Na camada das fontes, é feito um primeiro contacto com os dados por via da recolha e armazenamento dos mesmos. Os dados provêm de duas fontes: fonte interna, se for recolhido por inquérito feito pelo INE, independentemente do modo de recolha (pessoalmente, por telefone, registo interno, sendo feito no portal Webreg, ou registo *online* por iniciativa dos respondentes, sendo feito no portal Webinq); ou fonte externa, se for recolhido de fontes administrativas, resultantes de protocolos já previamente estabelecidos com outras entidades, como a Autoridade Tributária (AT) ou os certificados de óbitos da Direção Geral de Saúde (DGS). Nesta camada, existe, portanto, uma forte interação entre vários utilizadores e a atividade de recolha de dados, obrigando a alguma transformação dos mesmos, para posterior carregamento no Statistical Data Warehouse. Como tal, acabam por serem utilizadas várias ferramentas de registo e transformação. Para além de serem utilizadas ferramentas desenvolvidas internamente para registo de dados aquando a resposta a inquéritos por parte dos respondentes, tal como o portal Webinq ou a ferramenta de registo interno Webreg, também são utilizadas ferramentas de pré-validação de dados que serve como preparação para a fase de integração, como o SQL Server, e ferramentas de visualização dos mesmos, como o SAP BusinessObjects. À semelhança de uma típica arquitetura de Data Warehouse, esta camada encaixa-se na definição de um OLTP, por se tratar de recolher e processar dados que são objeto de tratamento rotineiramente, e é a primeira porta dos dados antes destes serem introduzidos no Statistical Data Warehouse, sendo, também, o primeiro ponto de contacto dos utilizadores com estes mesmos dados.

Na camada da integração, o objetivo é fornecer *outputs* de dados já tratados para a análise e produção estatística, existindo, portanto, uma enorme atividade de tratamento de dados por parte dos técnicos de estatística. Esta camada corresponde ao típico processo ETL que ocorre num Data Warehouse «comercial», onde são definidas várias tarefas de limpeza, standardização, uniformização de dimensões e outras validações. No caso do Statistical Data Warehouse, adicionalmente, são definidas e executadas várias tarefas de classificação, codificação, validação ou edição de dados, de modo a garantir a máxima qualidade dos dados para futura análise estatística dos mesmos. Estas tarefas podem ser executadas automática ou manualmente, conforme a especificidade de cada inquérito e variável a ser tratada. Esta camada interliga-se com a definição de ODS. Estes dados ficam estacionados na ODS, uma base de dados temporária criada especificamente para tarefas de um nível operacional e para utilizadores com pouca especialização na gestão de dados, antes de serem definitivamente transferidos para o Data Warehouse, onde encontraremos as duas próximas camadas. As principais ferramentas utilizadas nesta camada são MySQL e Python com SQL, ou PISQL (para automatizar as regras de validação e outras de tratamento de dados).

A terceira camada, de interpretação e análise de dados, consiste em manipulação dos dados para o desenvolvimento de conceitos estatísticos, tais como índices, séries ajustadas à sazonalidade, rácios, intervalos de confiança, quartis, imputações de valores ou taxas de variação, entre outros. Esta manipulação ocorrerá somente por utilizadores especialistas em estatística e em ferramentas de análise estatística, e todos estes conceitos estatísticos devem ser criados e devidamente validados antes de serem executados para produção estatística, onde sairá o *output* final de dados prontos para a divulgação da informação estatística. A este *output* também se dá o nome de dados agregados, pois já são dados devidamente validados, analisados e calculados, que estão prontos para o processo final da operação estatística. Daqui surgirá a bifurcação entre microdados e macrodados. Os microdados são os dados recolhidos e devidamente integrados na ODS, que também são transferidos para o Statistical Data Warehouse; os macrodados são os

microdados processados por utilizadores especialistas em estatística no Statistical Data Warehouse após os cálculos estatísticos terem sido concebidos, validados e executados por estes mesmos utilizadores. Várias ferramentas são utilizadas nesta camada, tal como R, Stata ou IBM SPSS, de modo a realizar estes cálculos estatísticos sobre os microdados presentes no Statistical Data Warehouse. Para visualizar os microdados e executar os procedimentos estatísticos sobre os mesmos, temos o Oracle Data Warehouse. No INE, grande parte dos utilizadores operam sobre esta camada. Ao contrário de um Data Warehouse tradicional, onde a maior parte dos utilizadores organizacionais operam apenas sobre a última camada, que iremos elucidar de seguida, no INE existe uma grande especialização dos funcionários em termos estatísticos e informáticos, pelo que a grande parte executa as suas tarefas precisamente nesta camada, directamente sobre o Statistical Data Warehouse.

Por fim, a última camada, de acesso aos dados, consiste em fornecer uma imagem consistente, válida e flexível aos dados já previamente recolhidos, integrados, validados e analisados, quer os microdados quer os macrodados. Aqui sucedem várias tarefas de apresentação e divulgação de informação estatística, que pode ser pesquisada ou filtrada de várias maneiras, conforme o utilizador final pretenda. Serão executadas várias *queries* ou produção de relatórios, conforme as necessidades dos utilizadores. Para visualização e edição interna de informação, é utilizado Microsoft PowerBI e Microsoft Office, bem como SAP BusinessObjects. Para utilizadores exteriores, tem-se o Portal *Online* do INE, com destaque para a visualização das publicações, que contêm macrodados, e possibilidade de visualização dos microdados para as várias áreas temáticas de estatística sobre que o INE opera.

#### **4.2. Interligação do MPPE com o Statistical Data Warehouse**

Na primeira fase da produção estatística, a fase da conceção, entra em jogo uma das principais utilidades de um Statistical Data Warehouse: a possibilidade de se reutilizar dados e metadados já disponíveis dentro da base

de dados. Ao definir-se uma nova operação estatística, ou ao atualizar-se uma operação estatística já existente, deve-se, primeiro, investigar que conceitos e definições já estão presentes dentro da base de dados e reutilizá-las. Por exemplo, para operações estatísticas cuja população inquirida são as empresas, não se irá construir o conceito do NIF para cada operação. Constrói-se uma vez, fica corretamente definido na gestão de metadados, e, para cada nova operação estatística, reutiliza-se este conceito já previamente definido, não existindo necessidade de o recriar.

Para o primeiro processo desta fase, pretende-se juntar a identificação de necessidades de informação e as exigências dos utilizadores com a informação já disponível. Existe, assim, um reaproveitamento da informação já criada, de modo a reduzir a carga sobre os técnicos de informática e de estatística para a definição de objetos dentro do Statistical Data Warehouse. Ademais, ao debruçar-se sobre a informação já existente no Statistical Data Warehouse, é feita uma análise sobre a mesma, existindo, assim, a possibilidade de atualizar e melhorar certos conceitos, definições e variáveis para melhor servir as necessidades de informação das operações estatísticas. Como tal, existirá a necessidade de se consultar o Statistical Data Warehouse, estando este processo conectado com a terceira camada do Statistical Data Warehouse, a interpretação/análise. Grande parte dos técnicos especialistas em estatística do INE trabalham sobre esta camada, e, ao cruzar-se as necessidades de informação com a informação já presente no Statistical Data Warehouse, pretender-se-á aceder à mesma para cruzar esta informação.

Continuando nesta primeira fase, chegamos ao processo das especificações. As definições das especificações da amostra a ser recolhida no âmbito de uma operação estatística vão entrar na primeira camada do Statistical Data Warehouse, a das fontes. As outras especificações metodológicas, como as regras e codificação, validação, tratamento e outros processos ETL ligar-se-ão com a segunda camada, a da integração. Subjacente a todo este processo, e a toda arquitetura do Statistical Data Warehouse, está uma forte gestão de metadados. Esta fase da conceção - mas, sobretudo este processo das especificações - terá uma carga muito forte de criação e

validação de metadados. Aliás, a definição de conceitos estatísticos e outros *outputs* incidirá sobre a terceira camada, de modo a definir, no Statistical Data Warehouse, onde encaixarão os vários conceitos e variáveis provenientes das respostas a uma operação estatística.

Por fim, no processo de desenvolvimento, testar-se-ão todos os instrumentos e componentes do processo de produção estatística já previamente identificados e especificados. Portanto, teremos uma utilização de todas as camadas de Statistical Data Warehouse, de todas as componentes a serem implementadas em cada processo da operacionalização estatística, be como uma validação dos metadados definidos, de modo a assegurar a consistência e correta implementação no Statistical Data Warehouse. Mais uma vez, temos a importância de uma correta gestão dos metadados, para assegurar a consistência da produção estatística e, também, para assegurar que todos os técnicos especialistas em estatística trabalham sobre uma mesma base, uma mesma “linguagem” sobre o que é que cada conceito estatístico realmente significa para a operação estatística em concreto.

Chegando à segunda fase da produção estatística, a operação, muito do que aqui se descreverá já vem, de certo modo, descrito na Figura 3 da página 22. É aqui onde ocorre grande parte da atividade desenvolvida pelo INE e outros elementos do SEN, e onde a grande parte dos utilizadores gasta o seu tempo e recursos para as tarefas que pretendam executar, sobretudo os técnicos especialistas em estatística do INE.

O primeiro processo da operação é a recolha. Como nos indica a Figura 3, o processo de recolha será a primeira porta de entrada dos dados para o Statistical Data Warehouse. Todas as atividades de gestão de respondentes, registo de informação ou pré-validação das respostas inquiridas por parte do DRGD ocorrerão na camada das fontes. Os TSEE podem visualizar os dados e cruzar informação no SAP BusinessObjects, ou, inclusive, extraírem a informação de que necessitam precisamente nesse *software* e trabalhá-la de forma independente no seu próprio computador de serviço. Como também se observa, temos a possibilidade de obtermos dados provenientes de duas fontes. As fontes internas são as respostas recolhidas directamente dos

respondentes, utilizando instrumentos desenvolvidos internamente, como o Webreg ou o Webinq, e possibilitando de imediato uma pré-integração dos dados precisamente por os respondentes estarem a responder numa plataforma desenvolvida pelo INE e de acordo com os conceitos estatísticos já especificados na fase anterior. As fontes externas resultam de protocolos estabelecidos entre o INE e outras entidades por forma a satisfazer necessidades de informação para certas operações estatísticas e, de igual forma, reduzir a carga estatística sobre os respondentes, seguindo o princípio de que se é possível obter a mesma informação de outras fontes externas que o respondente já teve de as fornecer, não se deve solicitar ao respondente a mesma informação duas vezes. Como as fontes externas resultam de protocolos, existe, neste processo e nesta camada, uma integração entre os conceitos provenientes dessas entidades externas e os conceitos definidos pelo INE, para que se possa desenvolver a atividade estatística continuando a utilizar os mesmos metadados definidos para cada operação estatística, facilitando o trabalho do técnico na produção de estatística oficial.

O segundo processo nesta fase estatística denomina-se de processamento, e é executada na camada da integração. O conjunto de atividades desenvolvidas neste processo corresponde a um típico processo ETL, frequente em qualquer arquitetura de Data Warehouse. O fundo, pretende-se tratar de cada registo de dados, proveniente das diferentes fontes internas ou externas e que já sofreram uma pré-validação por parte do DRGD, assimilar estes dados sob os mesmos conceitos e variáveis estatísticos, que já estão definidos nos metadados, e prepará-los para o processo seguinte, um processo de análise estatística em que os dados têm de ter a melhor qualidade e fiabilidade possíveis para assegurar a máxima veracidade e qualidade da estatística oficial. A maioria das actividades aqui executada acaba por ser programada por técnicos especialistas em estatística e com competências em Python e SQL, sendo posteriormente executadas automaticamente, com cada carregamento de dados “crus”, ou não-estruturadas, da camada das fontes para a camada da integração, donde sairão dados “tratados” ou estruturados.

O terceiro processo é o de análise, e ocorre na camada de interpretação e análise de dados. A maior parte dos utilizadores internos do INE, os técnicos especialistas em estatística, trabalha sobre esta camada durante a maior parte do seu tempo. Com os dados todos centralizados no Statistical Data Warehouse, cada grupo de técnicos ter-lhes-á atribuído um conjunto de operações estatísticas com que trabalhar. Assim sendo, de acordo com as exigências de cada grupo de técnicos, teremos vários data marts, que conterão os dados necessários para os técnicos desenvolverem as suas atividades de produção, validação, análise e interpretação estatística, bem como preparação desta informação estatística nova para a divulgação da mesma, com a ajuda do Departamento da Difusão. Maioritariamente, são os Departamentos de Matéria que aqui trabalham, onde cada departamento terá o seu serviço que, por sua vez, incidirá sobre cada área temática de estatística, como nos demonstra o Anexo 1. Contudo, também se poderá contar com a colaboração do DRGD e do DMSI, para validação contínua dos dados ou para solicitar algum tipo de ajuda informática que seja necessária. Aqui é necessário um grande conhecimento de estatística e informática para que se possa trabalhar sobre esta camada. Logo, daqui parte uma grande diferença do Statistical Data Warehouse para o Data Warehouse tradicional, em que a maioria dos utilizadores internos trabalha, de facto, sobre esta camada, de modo a produzirem *outputs* para o próximo processo estatístico. Como o *core business* do INE são os dados, é natural que grande parte do trabalho desenvolvido pelos técnicos do INE incida precisamente sobre a sua base de dados, o Statistical Data Warehouse (que é construída com recurso ao *software* Oracle Data Warehouse).

Por fim, o processo da divulgação incidirá sobre a camada do acesso aos dados. Daqui teremos os *outputs* produzidos pelos técnicos especialistas em estatística na camada anterior prontos para divulgação ou visualização, quer por utilizadores externos, desde a Administração Pública até a comunicação social, passando pelas universidades, seus alunos e investigadores, ou mesmo o público em geral porque a informação estatística pretende ser útil e relevante para toda a sociedade, quer por utilizadores internos que queiram

procurar informação ou desenvolvê-la melhor e cruzar um determinado *output* produzido com outras dimensões ou variáveis por forma a obter a informação que pretende. Os produtos estatísticos são produzidos nesta camada e nesta fase pelos técnicos do Departamento da Difusão, utilizando toda a informação que foi produzida nas camadas anteriores e executando as suas próprias tarefas de validação e qualidade dos dados de modo a assegurar que estes produtos têm, de facto, a qualidade suficiente para serem divulgados pelo INE, isto é, para que o INE possa “dar a cara” pelos resultados que divulga e assegurar a sua máxima fiabilidade e relevância para a sociedade.

Chegando à última fase e último processo do MPPE, a avaliação será executada sobre todas as camadas do Statistical Data Warehouse. Todos os *outputs* resultantes do processo da avaliação serão armazenados no Statistical Data Warehouse e mantidos como um registo histórico para a constante melhoria do processo de produção estatística. Estes outputs podem vir na forma de relatórios de execução ou relatórios de qualidade, ficheiros de registo (*log files*) e *feedback* por parte de qualquer utilizador ou respondente. Desta avaliação poderá resultar novas propostas de melhoria da operação estatística, resultando, por exemplo, na reformulação de conceitos estatísticos ou na atualização da amostra a ser inquirida, tendo tal várias implicações na gestão de metadados e nos processos informáticos já previamente programados no Statistical Data Warehouse, obrigando, novamente, a uma dinâmica entre o DRGD, os DM e o DMSI, de modo a reconfigurar estes aspetos da operação estatística.

#### ***4.3. Prós e Contras da Solução Implementada de Statistical Data Warehouse no INE***

Agora observaremos alguns prós e contras da solução de Statistical Data Warehouse implementada no INE. Um quadro-resumo destas observações pode ser encontrado na Tabela 2.

Começando pelos pontos positivos da solução implementada, denota-se a importância de uma das principais características de um Statistical Data Warehouse: a sua capacidade de manter registos históricos de metadados. Com qualquer alteração ou nova versão criada aos metadados, sejam elas os conceitos, as variáveis, documentos, etc., as versões anteriores são mantidas em arquivo, funcionando como uma espécie de registo de todo o trabalho que foi desenvolvido e do porquê se estar a criar ou atualizar os metadados. A ferramenta desenvolvida pelo INE português é o Sistema de Metainformação<sup>1</sup>, ou SMI, consultável pelo endereço da nota de rodapé desta página, onde se pode consultar todos os metadados utilizados nas operações estatísticas e as suas versões históricas.

Outra vantagem que decorre de uma característica essencial de um Statistical Data Warehouse é a capacidade de se reutilizarem dados e metadados, já previamente existentes no Statistical Data Warehouse e na gestão de metadados, que possibilita uma conceção estatística mais fácil e eficiente. A correta gestão de metadados também possibilita um cruzamento de dimensões ao longo de várias áreas temáticas de estatística, como a possibilidade de cruzar população desempregada com distritos, por exemplo, para se saber quais os distritos com maior taxa de desemprego.

Esta solução apoia a standardização de métodos e conceitos, levando todo o INE a falar sobre um mesmo “dicionário” e a operar sobre a mesma plataforma informática e concetual. De igual forma, também apoia o avanço tecnológico na infraestrutura informática e leva à partilha do conhecimento a nível nacional e europeu, como é atestado, aliás, graças à iniciativa do Eurostat em criar uma *task-force* que precisamente se debruça e partilha as experiências com Statistical Data Warehouse em cada INE de cada país europeu, denominado “*Centre of Excellence for Data Warehouse*”, projeto

---

<sup>1</sup> Sistema de Metainformação consultável em: <http://smi.ine.pt/>

integral do “*Collaboration in Research and Methodology for Official Statistics*” também desenvolvido pelo Eurostat.

Por fim, o Statistical Data Warehouse do INE permite o acesso fácil a registos e publicações históricas, bem como uma fácil consulta e manipulação de informação estatística oficial por parte dos utilizadores externos via o portal *online* do INE.

Contudo, uma solução de implementação assim comporta imensos custos financeiros, custos de recursos humanos em termos de se ter uma equipa especializada para a implementação dum sistema de informação assim e em termos da formação que é necessária dar aos técnicos especialistas em estatística (a maior parte dos funcionários do INE) para o novo ambiente de produção estatística, e custos de tempo para todo o planeamento, desenho e implementação de um Statistical Data Warehouse ao longo de todo o INE.

Por se tratar de um sistema de informação que serve como base e como plataforma para todas as fases do processo de produção estatística, existe uma necessidade de unificar todas as áreas funcionais do INE numa mesma direção, para que trabalhem em conjunto e possam definir todos os requisitos informáticos para implementarem devidamente a produção estatística no Statistical Data Warehouse. Isto aliado à necessidade de formação intensiva para todos os funcionários de modo a se adaptarem às novas ferramentas e metodologias de produção estatística, encontramos a principal barreira, que é a resistência à mudança por parte dos utilizadores.

Curiosamente, deve-se notar algo particular aquando a implementação do Statistical Data Warehouse no INE português: existiu uma receptividade positiva deste novo ambiente de produção estatística por boa parte dos técnicos. As formações lecionadas eram frequentadas com interesse e participação por parte dos mesmos, e existiu um constante *feedback* entre os especialistas de Statistical Data Warehouse e os utilizadores, o que possibilitou uma implementação mais eficiente deste sistema e com constante melhoria do mesmo.

Outra desvantagem a assinalar é algo que se prende com qualquer sistema de informação implementado em qualquer organização, e é ainda mais crítico no INE: a dependência total do processo estatístico em toda a infraestrutura informática, o que pode dificultar ou atrasar certas etapas da produção estatística. Uma falha num componente do sistema faz parar o “esquema de montagem” da produção estatística, o que, se não for tratada atempadamente, e como não existe ambiente alternativo para a produção estatística, prejudica severamente o *core business* do INE.

Por fim, e algo que infelizmente se assinala no INE embora se prevejam algumas mudanças neste âmbito, um ponto negativo da solução implementada é a falta de investimento contínuo na mesma. Se não existir investimento, por exemplo, na aquisição de *software* atualizado, ou na melhoria de alguns componentes do Statistical Data Warehouse, poder-se-á não conseguir responder às exigências dos utilizadores nem às necessidades de informação que apareçam no âmbito do planeamento da estatística oficial. É de assinalar, porém, o investimento de formação que o INE tem para com os seus técnicos especialistas em estatística. São realizadas imensas ações de formação em *software* como Microsoft Access ou R, bem como outras temáticas não relevantes para este campo, e nota-se um investimento nas pessoas por parte do INE. A infraestrutura, por outro lado, carece de uma melhoria, e, embora existam vontades de mudança neste aspeto, os custos que tal acarreta e a implementação destas melhorias na infraestrutura são obstáculos a esta mudança.

Tabela 2 - Prós e Contras da Solução Implementada de Statistical Data Warehouse no INE

Pontos Fortes	Pontos Fracos
Manter registos históricos dos metadados	Imensos custos financeiros, de recursos humanos e de tempo
Reutilização de dados e metadados	Resistência à mudança por parte das pessoas (embora em menor grau)
Estandarização de métodos e conceitos	Necessidade de dar formação intensiva

	aos utilizadores
Partilha do conhecimento	Dependência total do processo estatístico
Acesso fácil a publicações históricas, com consulta e manipulação de dados	Alguma falta de investimento na arquitetura

## 5. CONCLUSÕES E INVESTIGAÇÃO FUTURA

A presente dissertação permitiu avançar conhecimento na área do Statistical Data Warehouse, uma área que carece de alguma pesquisa, nomeadamente no que toca ao caso particular português. Pôde-se analisar que todo o processo de produção estatística foi efetivamente moldado para poder acomodar o novo ambiente informático de produção estatística e vice-versa, isto é, a dinâmica entre os processos e a tecnologia foi positiva, agilizando as várias tarefas dentro de cada processo de produção estatística e configurando-as dentro do sistema de informação concebido a pensar, precisamente, no *core business* do INE: os dados e a sua manipulação estatística.

Foram desenvolvidas várias soluções internas como forma de registo e integração dos dados, adquiriu-se *software* aquando as necessidades informáticas assim o exigiam, como para a construção do Statistical Data Warehouse em si com recurso às ferramentas da Oracle, e as ferramentas de visualização e produção de *queries* e relatórios também foram adquiridas externamente.

Ademais, existiu uma forte aposta na correta gestão dos metadados, para que o INE pudesse falar sobre o mesmo “dicionário” ao longo de todo o processo de produção estatística, mantendo registos históricos sobre as alterações efetuadas sobre os mesmos, por forma a registar as melhorias realizadas nos processos, a nível de conceitos, variáveis e outros elementos de metadados.

Denotou-se, também, uma relação positiva dos utilizadores, nomeadamente dos técnicos especialistas em estatística - os funcionários do INE em grande parte -, com este novo ambiente de produção estatística. A resistência à mudança efetivamente existiu, mas foi positivamente colmatada com várias ações de formação e interação constante entre os técnicos especialistas em estatística e o DMSI.

Em suma, a dinâmica entre utilizadores, processos e tecnologia observou-se como positiva, embora não se possa ocultar vários pontos que carecem de melhoria, nomeadamente a nível do investimento nas ferramentas (*software*) e no desenvolvimento da arquitetura de Statistical Data Warehouse.

Este estudo enfrentou fortes limitações sobretudo pela falta de documentação que pudesse suportar a minha própria pesquisa. Os documentos encontrados eram escassos e desatualizados em certos aspetos. Contudo, isto obrigou-me, e ainda bem que assim o foi, a pensar com recurso à minha própria experiência enquanto funcionário do INE, trazendo o meu contributo para este trabalho de investigação, adicionando-lhe, também, conversas valiosas que tive com técnicos especialistas na matéria.

Vários pontos merecerão reflexão futura, ainda por mais quando este tópico se insere numa área de âmbito largo como o é a área de dados e gestão de dados. A lidação do INE com *big data*, que possibilitará recolher vastas quantidades de dados potencialmente úteis para o desenvolvimento de produtos estatísticos, mas cuja integração dos mesmos não se enquadra na atual arquitetura desenhada de Statistical Data Warehouse e torna-se necessário projetar outras arquiteturas, nomeadamente Real-Time Data Warehouse; o objetivo destacado do INE em se tornar uma base nacional de armazenamento integrado de dados, pretendendo tornar-se um “banco” de dados único, completo e acessível a todo o público; a integração de ferramentas de inteligência artificial para uma produção mais eficiente e completa de informação estatística, sobre as quais existem pouquíssimas iniciativas a se registarem mas que também merecem a sua anotação; e, por fim, a matéria de recursos humanos e de especialização de competências, no contexto de uma Administração Pública com uma população empregada

francamente envelhecida e que necessita de introduzir força de trabalho jovem e com um maior saber nas áreas de Tecnologias de Informação e Comunicação.

## REFERÊNCIAS BIBLIOGRÁFICAS

Abramson, I., 2004. *Data Warehouse: The Choice of Inmon versus Kimball*. [Online]

Disponível em: <https://www.ismll.uni-hildesheim.de/lehre/bi-10s/script/Inmon-vs-Kimball.pdf>

[Acedido em 6 Outubro 2019].

Araoj, A., 2017. *Bill Inmon vs. Ralph Kimball - Data Analytics and BI WORLD*. [Online]

Disponível em: <https://arjunjune.wordpress.com/2017/03/20/bill-inmon-vs-ralph-kimball/>

[Acedido em 11 Setembro 2019].

Balasubramaniam, A. K., 2015. Methodological Standpoint: Inmon Vs Kimball. 19 Fevereiro, p. 12.

Breslin, M., 2004. Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models. *Business Intelligence Journal*, 1(9), pp. 6-20.

Centre of Excellence on Data Warehousing, 2017. *Overall Handbook to Set Up a S-DWH*, s.l.: Comissão Europeia.

Correia, M. d. C. B., 2009. A Observação Participante Enquanto Técnica de Investigação. *Pensar Enfermagem*, pp. 30-36.

Denzin, N. K., 1978. *The Research Act: A Theoretical Introduction to Sociological Methods*. Estados Unidos da América: McGraw-Hill.

Flender, S., 2019. *Data is not the new oil - Towards Data Science*. [Online] Disponível em: <https://towardsdatascience.com/data-is-not-the-new-oil-bdb31f61bc2d>

[Acedido em 11 Outubro 2019].

George, S., 2012. *Inmon or Kimball: Which approach is suitable for your data warehouse?* [Online]

Disponível em: <https://www.computerweekly.com/tip/Inmon-or-Kimball->

Which-approach-is-suitable-for-your-data-warehouse

[Acedido em 11 Setembro 2019].

Goossens, H., 2012. *The Statistical Data Warehouse: A Central Data Hub, Integrating New Data Sources and Statistical Output*. Genebra, Comissão Económica das Nações Unidas para a Europa, pp. 1-11.

Grover, R., 1998. *Identification of factors affecting the implementation of data warehousing*. Dissertação ed. Auburn: Auburn University.

Hoffer, J. A., Topi, H. & Ramesh, V., 2016. *Modern Database Management*. 12<sup>a</sup> ed. s.l.:Pearson Education.

Hwang, H.-G., Ku, C.-Y., Yen, D. C. & Cheng, C.-C., 2004. Critical factors influencing the adoption of data warehouse technology: a study of the banking industry in Taiwan. *Decision Support Systems*, 1(37), pp. 1-21.

Kimball, R. & Ross, M., 2002. *The Data Warehouse Toolkit*. Indianapolis: John Wiley and Sons, Inc..

Kimball, R. et al., 2008. *The Data Warehouse Lifecycle Toolkit*. Indianapolis: Wiley Publishing, Inc..

Krishnan, K., 2013. *Data Warehousing in the Age of Big Data*. Waltham: Morgan Kaufmann.

Panoply, 2018. *The Difference Between a Database and a Data Warehouse*. [Online]  
Disponível em: <https://panoply.io/data-warehouse-guide/the-difference-between-a-database-and-a-data-warehouse/>

Rangarajan, S., 2016. *Data Warehouse Design - Inmon versus Kimball*. [Online]  
Disponível em: <https://tdan.com/data-warehouse-design-inmon-versus-kimball/20300>  
[Acedido em 6 Outubro 2019].

Reddy, G. S., Srinivasu, R., Rao, M. P. C. & Rikkula, S. R., 2010. Data Warehousing, Data Mining, OLAP and OLTP Technologies are essential

elements to support decision-making process in industries. *International Journal on Computer Science and Engineering*, 2(9), pp. 2865-2873.

Salou, G., 2008. THE ECB Statistical Data Warehouse: improving data accessibility for all users. *IFC Bulletin no 28*, Agosto, pp. 313-317.

Shahid, M. B. et al., 2016. Application of Data Warehouse in Real Life: State-of-the-art Survey from User Preferences' Perspective. *International Journal of Advanced Computer Science and Applications*, 7(4), pp. 415-426.

Snowflake Computing, 2015. *Growing Investment in Data Warehouse & Big Data*. [Online] Disponível em: <https://www.snowflake.com/news/new-research-reveals-growing-investment-in-data-warehousing-and-the-limitations-of-current-and-new-big-data-technologies/> [Acedido em 3 Setembro 2019].

Stake, R. E., 2010. *Qualitative Research: Studying How Things Work*. Nova Iorque: The Guilford Press.

Thomas, B. E., 2018. *The data warehouse in 2018*. [Online] Disponível em: <https://www.cio.com/article/3245386/the-data-warehouse-in-2018.html> [Acedido em 1 Setembro 2019].

União Internacional de Telecomunicações, 2011. *ICT Facts and Figures*, Genebra: International Telecommunication Union.

União Internacional de Telecomunicações, 2017. *ICT Facts and Figures*, Genebra: International Telecommunication Union.

# ANEXOS

## Anexo 1 - Organograma do Instituto Nacional de Estatística

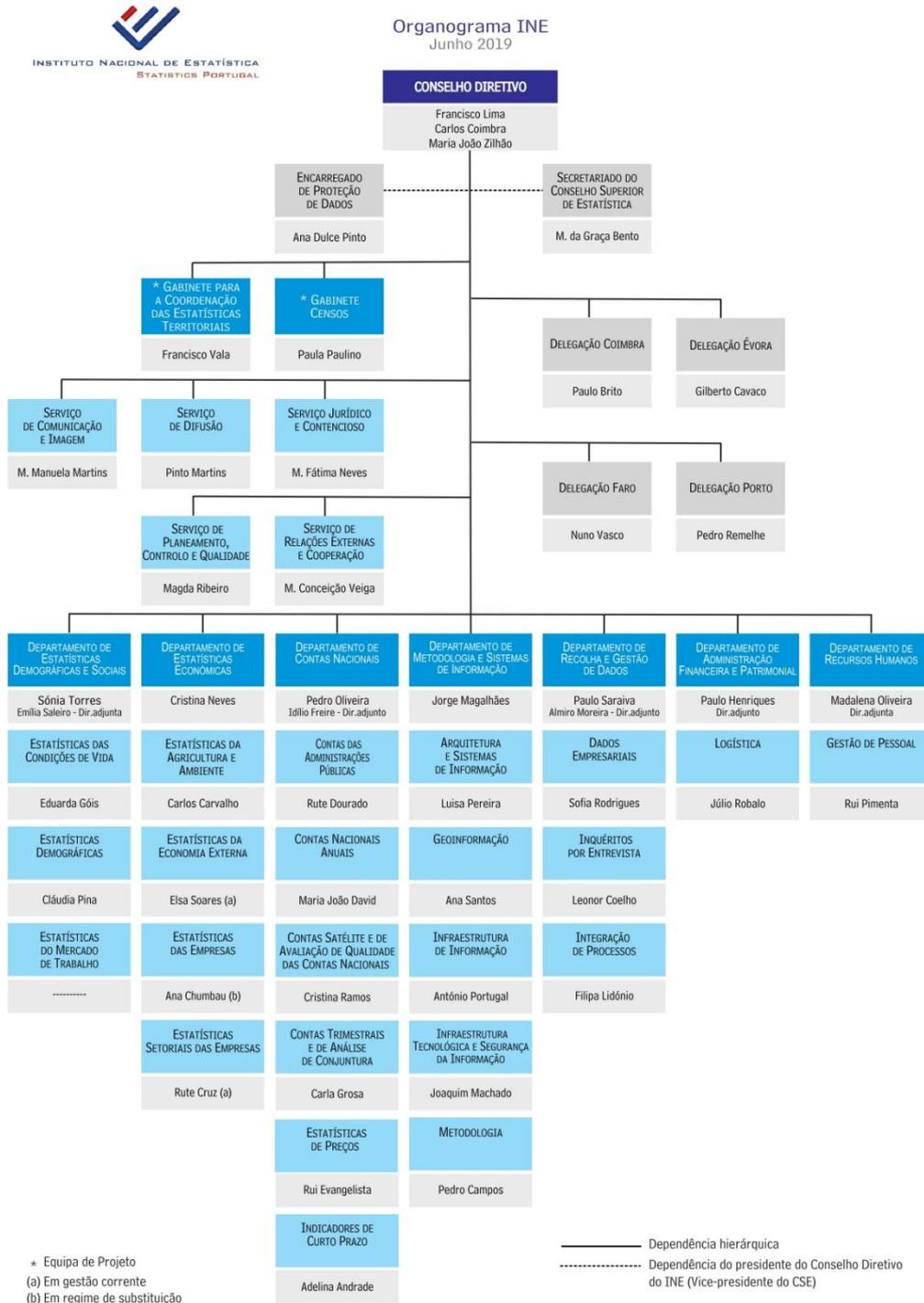


FIGURA 4 - ORGANOGAMA DO INSTITUTO NACIONAL DE ESTATÍSTICA A JUNHO DE 2019

FONTE: (INE, 2019)