



Lisbon School
of Economics
& Management
Universidade de Lisboa

MESTRADO

GESTÃO DE SISTEMAS DE INFORMAÇÃO

TRABALHO FINAL DE MESTRADO

TRABALHO DE PROJETO

ANÁLISE E MELHORIA DE UMA SOLUÇÃO *BUSINESS INTELLIGENCE* DE APOIO À TOMADA DE DECISÃO SOBRE OS ODS DE UMA INSTITUIÇÃO DE ENSINO SUPERIOR

MARIA BEATRIZ SILVEIRA CRUZ

OUTUBRO – 2022



Lisbon School
of Economics
& Management
Universidade de Lisboa

MESTRADO EM GESTÃO DE SISTEMAS DE INFORMAÇÃO

TRABALHO FINAL DE MESTRADO TRABALHO DE PROJETO

**ANÁLISE E MELHORIA DE UMA SOLUÇÃO *BUSINESS INTELLIGENCE* DE APOIO À TOMADA DE DECISÃO
SOBRE OS ODS DE UMA INSTITUIÇÃO DE ENSINO
SUPERIOR**

MARIA BEATRIZ SILVEIRA CRUZ

ORIENTAÇÃO:

PROFESSORA DOUTORA WINNIE NG PICOTO

OUTUBRO – 2022

Agradecimentos

Apesar do esforço pessoal dedicado durante a concretização deste trabalho final de mestrado, não estaria concluído sem o auxílio de um conjunto de pilares que conseguiram tornar o caminho menos penoso e um objetivo mais fácil de alcançar, pelo que, não posso deixar de aqui as referir.

Em primeiro lugar, expresso um agradecimento especial a minha orientadora, Professora Winnie Ng Picoto, que além de aconselhar e apoiar, desafiou-me na aprendizagem de uma ferramenta tecnológica não lecionadas durante o meu percurso académico. Também gostaria de deixar uma palavra de obrigada ao Professor Doutor Jesualdo Fernandes e a ex-aluna Maria Constança De Sá Portalete que em conjunto com a Professora Winnie Ng Picoto transmitiram conhecimentos base para a execução do projeto proposto.

Um eterno obrigada a Iara Chande, pela paciência e disponibilidade durante o desenvolvimento prático deste projeto, demonstrado assim a excelente pessoa e profissional que é.

Aos meus amigos com quem partilho histórias que vão ficar eternizadas na minha memória e irei sempre recordar com saudade. Em especial a Maria José Gomes, Sofia Guerreiro, Bárbara Pereira por todo o tempo alocado em debates existências.

Por fim, mas não menos importante, quero agradecer à minha família, que além de ser o meu principal pilar na vida, ajuda-me a ultrapassar os momentos de maior indecisão e apoia incondicionalmente a alcançar os meus objetivos.

Resumo

A Organização das Nações Unidas (ONU) na sequência da conferência, sobre o Desenvolvimento Sustentável, realizada em 2012, lançou 17 Objetivos de Desenvolvimento Sustentável (ODS), que definem as prioridades e aspirações globais para erradicar a pobreza e criar uma vida com dignidade e oportunidades para todos, dentro dos limites do planeta até 2030.

Embora a implementação desta iniciativa seja em grande medida responsabilidade dos governos de cada país, as Instituições de Ensino Superior (IES), sendo órgãos privilegiados na disseminação de conhecimento representam um papel crucial sobre os ODS. As IES podem contribuir para os ODS em quatro vertentes diferentes, que são a investigação, o ensino, o campus e a sociedade. Embora as quatro atividades sejam frequentemente abordadas separadamente, na realidade elas estão intimamente interligadas.

No âmbito deste projeto, desenvolvido em parceria com o Instituto Superior de Economia e Gestão da Universidade de Lisboa (ISEG) foi elaborada uma solução de *Business Intelligence* (BI), que consegue mapear de forma clara e correta o impacto das atividades de ensino e de investigação realizadas pelo ISEG entre 2011 e 2021.

Assim, utilizando as diretrizes dos desenvolvimentos realizados e após avaliar as possíveis fontes de dados, verificou-se a necessidade de adicionar novas dimensões (“DimPeople” e “DimDepartment”) ao modelo dimensional existente, estudar novos requisitos, bem como atualizar e modificar as informações analisadas e respetivas fontes de dados.

Apesar das dificuldades sentidas durante o processo de extração de dados, este desenvolvimento permitiu consolidar temáticas lecionadas no mestrado, adquirir novos conhecimentos e ainda possibilitou elaborar um conjunto de *dashboard* para que o ISEG possa tomar decisões tendo por base a análise fornecida.

Palavras-Chave: Objetivos de Desenvolvimento Sustentável; Instituições de Ensino Superior; *Business Intelligence*; Ensino; Investigação; Modelo Dimensional

Abstract

Following the United Nations (UN) conference on Sustainable Development, held in 2021, 17 Sustainable Development Goals (SDGs) were launched, which define global priorities and aspirations to eradicate poverty and create a life with dignity and opportunities for all, within the limits of our planet by 2030.

Although each country's government is largely responsible for the implementation of this initiative, Higher Education Institutions (HEI), which are privileged bodies in the dissemination of knowledge, play a crucial role on the SDGs. HEI can contribute to the SDGs in four different ways, which are research, education, campus and society. Even though the four activities are often discussed separately, in reality they are closely intertwined.

In the scope of this project, developed in partnership with the Higher Institute of Economics and Management of the University of Lisbon (ISEG), a Business Intelligence (BI) solution was developed that can clearly and correctly map the impact of the teaching and research activities carried out by ISEG between 2011 and 2021.

Thus, using the guidelines of the developments carried out and after evaluating the possible data sources, it was necessary to add new dimensions ("DimPeople" and "DimDepartment") to the existing dimensional model, study new requirements, as well as update and modify the analyzed information and respective data sources.

Despite the difficulties experienced during the data extraction process, this development allowed to consolidate themes taught in the master's degree, to acquire new knowledge and also to elaborate a set of dashboards so that ISEG can make decisions based on the analysis provided.

Key-Words: Sustainable Development Goals; Higher Education Institutions; Education; Research; Dimensional Model

Índice

Índice de Figuras	v
Índice de Tabelas	vi
Lista de Abreviaturas.....	vii
1. Introdução.....	1
1.1. Enquadramento e Motivação	1
1.2. Objetivos do Projeto	2
1.3. Relevância dos objetivos	3
2. Revisão da Literatura	3
2.1. Objetivos de Desenvolvimento Sustentável	4
2.2. O ODS e as Instituições de Ensino Superior	6
2.3. <i>Business Intelligence</i>	8
2.3.1 Arquitetura de uma Solução de <i>Business Intelligence</i>	9
2.3.1.1. Identificação das fontes de dados (Data Collection)	9
2.3.1.2. Processo ETL.....	10
2.3.1.3. Armazenamento dos dados na Data Warehouse	10
2.3.1.4. Análise dos Dados da Data Warehouse	12
2.3.1.5. Apresentação dos Dados	13
2.4. Outros projetos desenvolvidos por IES	14
3. Metodologia.....	15
3.1. Enquadramento do Projeto.....	17
3.2. Identificação das Fontes de Dados.....	18
3.3. Conjunto de <i>Keywords</i>	18
3.4. Ambiente de Desenvolvimento.....	19
4. Desenvolvimento do Projeto	21
4.1. Tratamento de Dados	21
4.1.1. Alteração dos Parâmetros Analisados em <i>Python</i>	22

4.1.2. Divisão de <i>Outputs</i> em Formato Excel em <i>Python</i>	22
4.1.3. Unir <i>outputs</i> no <i>Power Query</i>	23
4.1.4. Substituição de Valores	23
4.2. Algoritmo de Classificação dos ODS	24
4.3. Modelo Dimensional.....	24
4.4. Visualização dos Dados	26
4.4.1. Análise e Discussão dos Novos <i>Dashboards</i>	28
4.4.2. Análise e Discussão dos <i>Dashboards</i> Atualizados.....	35
5. Conclusão e Propostas de Investigação Futuras	36
Referências Bibliográficas.....	39
Anexos.....	45

Índice de Figuras

Figura 1 - Processo de <i>Business Intelligence</i>	9
Figura 2 - Fluxograma do Projeto	16
Figura 3 - Renomeação dos parâmetros dos dados sobre as teses desenvolvidas nos mestrados entre 2011 e 2020	22
Figura 4 - Renomeação dos parâmetros dos dados sobre as teses desenvolvidas nos mestrados entre 2011 e 2021	22
Figura 5 - Algoritmo de divisão dos <i>outputs</i>	23
Figura 6 - Processo de criação das tabelas em ambiente <i>Microsoft Power Query</i>	23
Figura 7 - Formulário “Substituir Valores”	24
Figura 8 - Modelo dimensional de dados	26
Figura 9 - <i>Dashboard Research Contributions of 2021</i>	28
Figura 10 - Alternativa do <i>Dashboard Research Contributions of 2021</i>	29
Figura 11 - <i>Dashboard 2021 Review of Master and Phd Thesis</i>	30
Figura 12 - <i>Dashboard Master and Thesis Contributions from 2019 to 2021</i>	31
Figura 13 - <i>Dashboard 2021 Review Articles - Scopus</i>	32
Figura 14 - <i>Dashboard Articles from 2019 to 2021</i>	32
Figura 15 - <i>Dashboard Course Units</i>	33

Figura 16 - <i>Dashboard Course Units Contributions Between 2019 and 2021</i>	34
Figura 17 - <i>Dashboard Department</i>	34
Figura 19 - Versão 1 do Modelo dimensional	45
Figura 20 - Versão 2 do Modelo dimensional	45
Figura 21 - Versão 3 do Modelo dimensional	46
Figura 22 - Versão 4 do Modelo dimensional	46

Índice de Tabelas

Tabela I - Objetivo de Desenvolvimento Sustentável	4
Tabela II - As cinco áreas em que atuam os ODS	6
Tabela III - Atividades das IES sobre os ODS	8
Tabela IV - Valores das colocações no ensino superior	36

Lista de Abreviaturas

AGNU - Assembleia Geral das Nações Unidas

BI - Business Intelligence

DAX - Data Analysis Expressions

DM - Data Mining

DW - Data Warehouse

ERP - Enterprise Resource Planning

ETL - Extract, Transform and Load

IES - Instituições de Ensino Superior

ISEG - Instituto Superior de Economia e Gestão da Universidade de Lisboa

NLTK - Natural Language ToolKit

ODS - Objetivos de Desenvolvimento Sustentável

OLAP - Online Analytical Processing

ONU - Organização Nações Unidas

OWL - Web Ontology Language

SDGs - Sustainable Development Goals

SGBD - Sistemas de Gestão de Bases de Dados

UC - Unidade Curricular

UCs - Unidades Curriculares

UN- United Nations

XML - Extensible Markup Language

1. Introdução

1.1. Enquadramento e Motivação

No atual contexto global, o lento crescimento económico, a desigualdade social, as alterações climáticas, instabilidades políticas, entre outras problemáticas, estão a criar desafios sem precedentes para a comunidade internacional. Neste sentido, não é viável continuar com os mesmos padrões de produção ou consumo é necessária a mudança, para um desenvolvimento sustentável inclusivo e de longo prazo. O Desenvolvimento Sustentável, é a capacidade de satisfazer as necessidades da geração atual, sem comprometer as necessidades das futuras gerações (UNESCO, 2012). As problemáticas anteriormente identificadas, por serem globais dificilmente serão ultrapassadas através de projetos locais ou pequenas campanhas de sensibilização, daí o papel das grandes organizações como a Organização das Nações Unidas serem consideradas fundamentais.

Neste sentido, a ONU criou em 2015 a Agenda 2030 para o desenvolvimento sustentável, que apela à responsabilidade dos atores, públicos e privados, bem como de todos os Estados-membro da ONU, para alcançar os 17 Objetivos de Desenvolvimento Sustentável e as respetivas 169 metas.

O processo de construção desta agenda durou cerca de três anos e procurou ser o mais abrangente possível, tendo envolvido os diferentes estados e peritos internacionais, como também os representantes da sociedade civil, do sector privado e da comunidade científica. Para integrar os diferentes interesses e preocupações, a agenda além dos temas sociais como a erradicação da pobreza e a igualdade de género, inclui outros tópicos, nomeadamente, os padrões sustentáveis de produção e de consumo, o crescimento económico inclusivo, a justiça social e segurança (República Portuguesa, 2017).

Apesar das diversas ações desenvolvidas, considera-se que os objetivos definidos ainda não estão a ser atingidos à velocidade desejada (Mohammed, 2020), uma vez que é necessário os indivíduos modificarem e adaptarem a sua forma de pensar e de agir (Rieckmann et al, 2017).

O setor do ensino é um dos poucos setores que conseguem promover e contribuir para todos os ODS (Junior et al., 2019). Inserido neste setor são identificadas as Instituições do Ensino Superior, que são entidades privilegiadas na produção e na comunicação de

conhecimento, através do ensino e da investigação. Estas desempenham um papel preponderante na promoção do conceito de sustentabilidade e a atuar como impulsionador da mudança em direção a um mundo mais sustentável (Haupt et al., 2015).

O Instituto Superior de Economia e Gestão, como Instituição do Ensino Superior, manifestou disponibilidade e interesse em adaptar a sua visão e estratégia com vista a proporcionar a produção de soluções sob o desenvolvimento sustentável. Neste contexto, Marcelino (2020), apresentou uma solução de *Business Intelligence* capaz de classificar e mapear as atividades investigação e de ensino do ISEG, que contribuem para os ODS, e, ao mesmo tempo, apresentar esses resultados sob a forma de *dashboards* (Marcelino, 2020). No ano letivo seguinte, Portalete (2021) para garantir a qualidade da solução de *Business Intelligence* apresentada por Marcelino (2020), procedeu ao melhoramento do algoritmo de classificação dos SDG de forma a reduzir os “falsos positivos” resultantes da classificação anterior (Portalete, 2021).

1.2. Objetivos do Projeto

Tendo por base os desenvolvimentos realizados por Marcelino (2020) e Portalete (2021), o presente projeto tem como objetivo principal **classificar e comunicar de que formas as atividades de ensino e de investigação realizadas pelo ISEG entre 2011 e 2021 estão a contribuir para a concretização dos ODS**. Assim, para a concretização deste objetivo, foram definidos os seguintes objetivos secundários:

- a) Análise e atualização das fontes de dados;
- b) Realização do tratamento de dados de acordo com o modelo já definido;
- c) Modificação do algoritmo de classificação de artigos, teses e programas de unidades curriculares (UCs) por ODS, para que fosse possível a análise de um maior volume e novos dados;
- d) Adição das dimensões “DimPeople” e “DimDepartment” ao modelo dimensional de dados;
- e) Atualização de *dashboards* para a visualização da informação;
- f) Desenvolvimento de novos *dashboards*;
- g) Identificação de oportunidades de melhoria para trabalhos futuros.

Embora todos os objetivos secundários sejam pertinentes para alcançar o objetivo principal destaco o tópico d) (Adição das dimensões “DimPeople” e “DimDepartment” ao modelo dimensional de dados) e f) (Desenvolvimento de novos *dashboards*). Estes

dois objetivos secundários têm especial atenção, uma vez que vão dar resposta a uma das recomendações propostas por Marcelino (2020): “Integração de novas fontes de dados, de forma a contemplar todas as atividades da instituição e a possibilidade de criar novas análises e conhecimentos relevantes para a instituição, como acrescentar a dimensão “autores” juntamente com a respetiva filiação, adicionar novos indicadores, KPI’s, entre outras;” (Marcelino, 2020, p.34)

1.3. Relevância dos objetivos

Perante a possibilidade de qualquer IES, nomeadamente o ISEG, incluir e adaptar a sua estratégia e visão para os ODS é importante que a mesma se questione, sobre os possíveis benefícios, que podem retirar dessa adaptação. Além disso é indispensável, que a instituição reconheça o seu papel enquanto promotora de inovação e influenciadora principal sobre os futuros agentes de mudança das atitudes da sociedade, os alunos (ISEG, 2021).

Tanto o presente projeto como os que foram desenvolvidos por Marcelino (2020) e Portalete (2021), têm como principais benefícios permitir a monitorização das atividades académicas da instituição ao longo dos anos, de forma simples, intuitiva e dinâmica, bem como reforçar o posicionamento do ISEG enquanto universidade que contribui ativamente para a implementação dos ODS.

Além dos benefícios anteriormente referidos o ISEG, através deste último desenvolvimento, até a data, irá ter acesso a um conjunto de dados atualizados que surgiram do cruzamento de novas fontes de dados e assim continuar a definir os próximos passos a tomar em relação à estratégia da instituição.

2. Revisão da Literatura

No decorrer deste capítulo serão discutidos e abordados tópicos análogos ao tema do projeto. Deste modo, vão ser desenvolvidos os temas relacionados com a ferramenta de BI, os objetivos de desenvolvimento sustentável, a relação entre os ODS com as instituições de ensino superior e projetos desenvolvidos por IES partindo do conhecimento e descobertas já realizadas anteriormente, tanto por Marcelino (2020) e Portalete (2021), bem como por outros autores experientes em pesquisas semelhantes.

2.1. Objetivos de Desenvolvimento Sustentável

A 25 de setembro de 2015, a Assembleia Geral das Nações Unidas (AGNU) adotou por unanimidade na cimeira de chefes de estado e de governo a iniciativa denominada de “Transforma o nosso mundo: a Agenda 2030 para o Desenvolvimento Sustentável”. Esta iniciativa é constituída por 17 Objetivos de Desenvolvimento Sustentável (tabela I), desdobrados em 169 metas e entrou em vigor a 1 de janeiro de 2016 (Centro de Informação Regional das Nações Unidas para a Europa Ocidental, 2018). A agenda das nações unidas para 2030, assinalada por mais de 190 países, constitui um plano de ação centrado nas pessoas, no planeta, na prosperidade, na paz e nas parcerias (5P), tendo como finalidade a erradicação de problemáticas, nomeadamente, a pobreza e o desenvolvimento sustentável (República Portuguesa, 2017).

Ban Ki-Moon, antigo secretário-geral da Organização Nações Unidas, define os ODS como sendo uma visão comum para a humanidade e um contrato social entre os líderes mundiais e a sociedade. Acrescenta ainda, que estes correspondem a uma lista de medidas a serem realizadas em nome da sociedade e no planeta (Centro de Informação Regional das Nações Unidas para a Europa Ocidental, 2018). Assim, os 17 ODS são considerados como um conjunto de prioridades que pretendem criar linhas orientadoras aos países, para solucionar problemáticas já identificadas (SDSN Australia/Pacific, 2017).

Tabela I - Objetivo de Desenvolvimento Sustentável

ODS	Objetivo
ODS 1 Erradicar a Pobreza	Erradicar a pobreza em todas as suas formas, em todos os lugares.
ODS 2 Erradicar a Fome	Erradicar a fome, alcançar a segurança alimentar, melhorar a nutrição e promover a agricultura sustentável.
ODS 3 Saúde de Qualidade	Garantir o acesso à saúde de qualidade e promover o bem-estar para todos, em todas as idades.
ODS 4 Educação de Qualidade	Garantir o acesso à educação inclusiva, de qualidade e equitativa, e promover oportunidades de aprendizagem ao longo da vida para todos.
ODS 5 Igualdade de Género	Alcançar a igualdade de género e empoderar todas as mulheres e raparigas.
ODS 6 Água Potável e Saneamento	Garantir a disponibilidade e a gestão sustentável da água potável e do saneamento para todos.
ODS 7 Energias Renováveis e Acessíveis	Garantir o acesso a fontes de energia fiáveis, sustentáveis e modernas para todos.

ODS 8 Trabalho Digno e Crescimento Económico	Promover o crescimento económico inclusivo e sustentável, o emprego pleno e produtivo e o trabalho digno para todos.
ODS 9 Indústria, Inovação e Infraestruturas	Construir infraestruturas resilientes, promover a industrialização inclusiva e sustentável e fomentar a inovação.
ODS 10 Reduzir as Desigualdades	Reduzir as desigualdades no interior dos países e entre países.
ODS 11 Cidades e Comunidades Sustentáveis	Tornar as cidades e as comunidades mais inclusivas, seguras, resilientes e sustentáveis.
ODS 12 Produção e Consumo Sustentáveis	Garantir padrões de consumo e de produção sustentáveis.
ODS 13 Ação Climático	Adotar medidas urgentes para combater as alterações climáticas e os seus impactos.
ODS 14 Proteger a Vida Marinha	Conservar e usar de forma sustentável os oceanos, mares e os recursos marinhos para o desenvolvimento sustentável.
ODS 15 Proteger a Vida terrestre	Proteger, restaurar e promover o uso sustentável dos ecossistemas terrestres, gerir de forma sustentável as florestas, combater a desertificação, travar e reverter a degradação dos solos e travar a perda de biodiversidade.
ODS 16 Paz, Justiça e Instituições Eficazes	Promover sociedades pacíficas e inclusivas para o desenvolvimento sustentável, proporcionar o acesso à justiça para todos e construir instituições eficazes, responsáveis e inclusivas a todos os níveis.
ODS 17 Parcerias para a Implementação dos Objetivos	Reforçar os meios de implementação e revitalizar a parceria global, para o desenvolvimento sustentável.

(Fonte: Centro de Informação Regional das Nações Unidas para a Europa Ocidental, 2018)

Os ODS abordam três dimensões do desenvolvimento sustentável, nomeadamente, a dimensão social, económica e ambiental (Centro de Informação Regional das Nações Unidas para a Europa Ocidental, 2018). Estes atuam, ainda, sobre cinco áreas que são identificadas como os 5P (tabela II).

Tabela II - As cinco áreas em que atuam os ODS

Área	Objetivo	ODS
Pessoas	Erradicação da pobreza e da fome, respeito da dignidade e da igualdade.	1, 2, 3, 4, 5 e 6
Planeta	Produção e consumo sustentáveis, combater as mudanças climáticas e gestão de recursos naturais.	7, 8, 9 e 10
Prosperidade	Garantir que todas as pessoas usufruem de prosperidade, saúde e progresso económico e social.	11, 12, 13, 14 e 15
Paz	Promover sociedades pacíficas, justas e inclusivas que sejam livres de medo e violência.	16
Parcerias	Integração transversal, interconexão e mobilização conjunta em prol dos mais vulneráveis.	17

(Fonte: SDSN Australia/Pacific, 2017; República Portuguesa, 2017)

A implementação da iniciativa, “Transforma o nosso mundo: a Agenda 2030 para o Desenvolvimento Sustentável”, é responsabilidade de qualquer indivíduo que esteja inserido num sector (SDSN Australia/Pacific, 2017). No entanto, em grande medida, as responsabilidades pertencem aos governos de cada país, uma vez que é necessário criar parcerias e uma maior expressão da solidariedade internacional evitando, que ninguém seja deixado para trás. Durante o processo de implementação é crítico, cada país, realizar regularmente a avaliação dos progressos alcançados. A avaliação utiliza um conjunto de 230 indicadores globais, cujos resultados serão compilados num relatório global anual, o “The Sustainable Development Goals Report”. Este relatório demonstra em que ponto o mundo se encontra na prossecução dos objetivos, recorrendo a dados estatísticos, e sublinhando as principais lacunas e os mais prementes desafios que o mundo enfrenta (República Portuguesa, 2017).

2.2. O ODS e as Instituições de Ensino Superior

Existe uma relação entre os ODS e as IES, na medida em que contribuem para o desenvolvimento de ambos.

As Instituições de Ensino Superior assumem, atualmente, uma posição importante no crescimento económico e desenvolvimento regional, com um contributo ao nível da melhoria da qualificação do capital humano, inovação nacional e transferência de conhecimento, que se reflete conseqüentemente no aumento da qualidade de vida da população.

Assim, a maioria das IES procura ou deveria procurar desenvolver estratégias no sentido de estimular a colaboração, a inovação e transferência de conhecimento (Cabrito et al., 2019).

Atualmente, as universidades são elementos cruciais para alcançar os ODS (Junior et al., 2019), na medida em que proporcionam o conhecimento, ideias e as soluções que sustentam a implementação dos ODS; formam os atuais e futuros executores e responsáveis de implementar os ODS; servem de modelo sobre como apoiar, adotar e implementar os ODS nas políticas de gestão e na cultura; por fim desenvolvem lideranças intersectoriais que orientam os ODS. Por sua vez, as IES irão beneficiar ao estar envolvidas em iniciativas deste género, na medida em que a procura de formação, sobre os ODS irá aumentar; proporcionam uma definição globalmente aceite e compreendida do conceito de universidade responsável; oferecem uma estrutura para demonstrar o impacto causado; criam fontes de financiamento; apoiam a colaboração como novos parceiros internos e externos (SDSN Australia/Pacific, 2017).

Segundo o SDSN Australia/Pacific (2017) o compromisso das instituições, para com os ODS refletem-se nas suas atividades de aprendizagem e ensino (Educação); na produção de investigação (Investigação); na interação com a sociedade (Interação com a sociedade); nas operações e na gestão da instituição (Operações no campus), como se pode verificar na tabela III.

Embora as quatro atividades sejam frequentemente abordadas separadamente, na realidade elas estão intimamente interligadas.

Tabela III - Atividades das IES sobre os ODS

Investigação	Incentiva e promove os ODS como um tópico de pesquisa, interdisciplinar e transdisciplinar; Apoia e planeja soluções inovadoras de desenvolvimento sustentável; Apoia ativamente a implementação nacional e local dos ODS capacitando os países em desenvolvimento a realizar e usar estudos, sobre os mesmos.
Educação	Disponibiliza uma educação/formação focada na sustentabilidade, providenciando aos alunos <i>skills</i> e motivação necessária para encarar os desafios dos ODS.
Operações no campus	Incorpora os ODS nas suas estruturas e operações identificando e direcionando qualquer “ <i>key gap</i> ”, que possa existir na resposta da universidade aos ODS ou integra os ODS nos relatórios organizacionais das universidades.
Interação com a sociedade	Fortalece o envolvimento e a participação do público na abordagem dos ODS, através de palestras, workshops ou plataformas digitais; Desempenha um papel de liderança no desenvolvimento de políticas e na defesa do desenvolvimento sustentável; Demonstra a importância e compromisso do setor universitário na implementação dos ODS.

(Fonte: SDSN Australia/Pacific, 2017)

Por outro lado, as IES beneficiarão ao apoiar e adotar medidas que ajudem a alcançar os ODS, pois são classificadas como universidades responsáveis e comprometidas, têm acesso a novos fundos de financiamento, nomeadamente, órgãos governamentais, entidades bancárias, entre outros, implementam e disponibilizam aos alunos um ensino em parceria com os ODS, estabelecem novas parcerias e colaborações externas e internas, uma vez que a Agenda fornece uma estrutura comum partilhada por organizações de diferentes setores, que estão interligados e disponibilizam uma estrutura que permite demonstrar o impacto e a pertinência que as IES têm na contribuição para o bem-estar local e global (SDSN Australia/Pacific, 2017).

2.3. *Business Intelligence*

Entre as décadas de 70 e 90 foram investigados métodos de processamento de dados, realizada a modelagem e o desenvolvimento de software, para grandes volumes de dados. Atualmente, as organizações procuram rapidez e prontidão perante os riscos e oportunidades (Caldeira & Lopes, 2018). Assim, diante o mercado competitivo *Business Intelligence* ganha papel de destaque, uma vez que permite transformar grandes quantidades de dados (Daniel, 2020).

A *Business Intelligence*, é um conceito que integra múltiplas fontes de informação e que atua como uma ferramenta estratégica diferenciadora das empresas (Barbieri, 2001), uma vez que possibilita fornecer uma visão sistémica do negócio. Assim, BI permite distribuir dados de forma concreta, transforma-os em informações útil para a tomada de decisão, possibilita a visualização da informação de diversas formas, cruzamento de dados proveniente de diversas fontes e análise do desempenho empresarial (Angeloni & Reis, 2006).

Para desenvolver uma solução de *Business Intelligence* composta por cinco componentes distintos, podem ser adotadas diversas ferramentas nomeadamente, o *Microsoft Power BI*, disponibilizado pela *Microsoft* (Microsoft, 2022).

2.3.1 Arquitetura de uma Solução de *Business Intelligence*

Mediante a revisão de literatura verifica-se, que existem diversas estruturas de BI, que diferem nas suas camadas, componentes, processos e relacionamentos (Ong et al., 2011). Uma das estruturas mais utilizadas no BI apresenta cinco componentes, que colaboram entre si: Identificação das fontes de dados (*Data Collection*), Processo de *Extract, Transform and Load* (ETL) (*Data Integration*), Armazenamento dos dados na *Data Warehouse* (DW) (*Data Repository*), Análise dos dados na *Data Warehouse* (*Analytics*) e Apresentação dos dados (*Data Presentation*) (Azeroual & Theel, 2018).

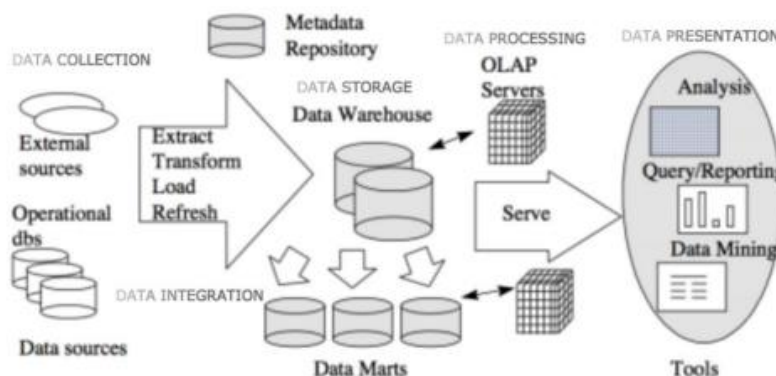


Figura 1 - *Processo de Business Intelligence*

(Fonte:Azeroual & Theel (2018), p.32)

2.3.1.1. Identificação das Fontes de Dados (*Data Collection*)

O primeiro componente do processo de BI consiste na identificação das diferentes fontes de dados internas ou externas. Numa organização, existem grandes quantidade de fontes

de dados relevantes, que por sua vez irão estar localizados em diferentes Sistemas de Gestão de Bases de Dados (SGBD) (Ong et al., 2011).

Os dados que serão mais tarde recolhidos das fontes identificadas podem ser: **estruturados** por exemplo dados provenientes de um *Enterprise Resource Planning* (ERP), histórico de compras de um cliente, transações realizadas por um cartão de crédito, documentos de Excel, entre outros; **não estruturados**, nomeadamente, vídeos, imagens, textos, áudios, entre outros; **semiestruturados** que são ficheiros *Extensible Markup Language* (XML), *Web Ontology Language* (OWL), o conteúdo de um e-mail, entre outros (Coelho, 2011).

2.3.1.2. Processo ETL

Após a identificação das fontes de dados, dá-se o processo de ETL responsável pela extração, transformação e carregamento de dados (Niu et al., 2009), de uma ou mais fontes, para serem integrados no fluxo de dados e repositórios, que estão na arquitetura do sistema de BI utilizado. Devido a complexidade do processo ETL é comum lidar com alguns problemas operacionais. Assim são implementadas algumas ferramentas, nomeadamente, *SQL Server Integration Services*, *SAP BusinessObjects Data Integrator*, *Oracle Data Integrator*, entre outras.

No processo ETL primeiramente é realizado a extração dos dados das diferentes fontes, para o ambiente da *data warehouse* (Ong et al., 2011).

Após a extração dos dados, os mesmos sofrem transformações, com base num conjunto de regras de negócio, de forma a melhorar a qualidade dos mesmos. Estas modificações, eliminam possíveis inconsistências, permitindo que só a informação relevante seja guardada no DW (Watson, 2009).

Por fim, após serem realizadas todas as tarefas que compõem a fase da transformação, os dados estão prontos para serem carregados no DW (Silva, 2016). Esta fase é crucial, uma vez que a qualidade dos componentes seguintes dependem deste processo (Niu et al., 2009).

2.3.1.3. Armazenamento dos Dados na Data Warehouse

Nesta fase após os dados serem transformados, limpos, integrados e consolidados são redirecionados e armazenados em plataformas próprias da arquitetura de BI de forma a serem depois analisados.

A *Data Warehouse*, consiste num repositório que armazena dados de diferentes fontes, internas ou externas, possibilitando desenvolver análises e consultas para apoiar a tomada de decisão numa organização (Ong et al., 2011).

Uma DW não é nenhum tipo de aplicação que se possa comprar e instalar nos computadores da organização. Na realidade, a sua implementação exige a integração de vários processos (Moura, 2017).

Com a implementação cada vez mais frequente deste tipo de repositório surgiram dois tipos de ferramentas capazes de analisar os dados existentes num DW, nomeadamente, as ferramentas de *Data Mining* (DM) e/ou as ferramentas On-Line Analytical Processing (OLAP) (Reddy & Schroeder, 1999).

As ferramentas de DM representam um papel crucial na criação de algoritmos de exploração de dados, relações entre os dados armazenados e ajudando a identificar padrões. Relativamente as ferramentas OLAP, estas disponibilizam novas formas de realizar uma análise multidimensional e, ainda, possibilitam criar hierarquias entre dimensões. Estas hierarquias permitem uma análise detalhada e a manipulação rápida dos dados, ao longo dos diferentes níveis de granularidade nas dimensões num DW.

Além das ferramentas mencionadas, é importante referir os modelos mais utilizados na modelagem dimensional, que torna a DW mais compreensível perante o utilizador, com uma navegação e um desempenho de resposta eficiente para ajudar na tomada de decisão (Kimball & Ross, 2013). O *star schema* (modelo em estrela) e *snowflake schema* (modelo de floco de neve) são os modelos dimensionais mais utilizados nos sistemas de gestão de base de dados relacionais (Niu et al., 2009).

Os modelos são compostos por tabelas de facto que estão associadas a tabelas de dimensão, através de chaves primárias que só deverão conter valores únicos não nulos e chaves estrangeiras. Tanto na tabela de facto como na de dimensão, ambas têm uma chave primária simples (um único atributo) ou composta (combinação de vários atributos) que identifica de forma única cada registo da tabela. (Moody & Kortink, 2000).

O modelo em estrela ou *star schema*, proposto inicialmente por Ralph Kimball é o esquema mais simples que permite a construção de um modelo dimensional sem que exista a perda de dados (Marques, 2015). Genericamente, no modelo em estrela é composto por uma tabela central (tabela de facto) e a ela estão ligadas algumas tabelas de

dimensão, formando assim uma estrela (Moody & Kortink, 2000). No entanto, apesar da sua simples estrutura poderá ser necessário o aumento do número de atributos, diminuindo a complexidade relacional e aumentando a complexidade de cada tabela (Marques, 2015). Relativamente ao esquema em floco de neve ou *snowflake schema*, genericamente pode ser descrito como um esquema em estrela com todas as hierarquias explicitamente visíveis, isto é cada tabela de dimensão pode conter várias hierarquias independentes. Esta estrutura ao ser implementada, o espaço de armazenamento poderá ser mais reduzido. No entanto é considerada uma estrutura mais complexa.

As grandes diferenças entre a estrutura em estrela e a estrutura em floco de neve deve-se ao facto de que a primeira não exige a normalização das tabelas dimensionais, enquanto a segunda exige (Kimball & Ross, 2013). Perante o leque diversificado de arquiteturas diferentes o ideal será implementar uma estrutura, que resulte do compromisso e equilíbrio entre o volume de dados e os requisitos de cada organização.

2.3.1.4. Análise dos Dados da Data Warehouse

Após identificar as fontes de dados, realizar o processo ETL e armazenamento dos dados na *data warehouse*, é necessário implementar algo que permita às empresas tirar partido da informação existente. A DW permite a utilização de diversos métodos de análise, que podem ser aplicadas graças à sua estrutura relacional multidimensional, assim como a uniformização e integração dos dados.

Um dos tipos de análise mais preditivo, que surgiu com o constante crescimento e amadurecimento dos sistemas de informação é *Data Mining* (Gomes, 2011). Esta técnica representa uma solução para a problemática de obter informação, a partir de grandes volumes de dados, colocando à disposição dos analistas de dados várias técnicas, que lhes permitem extrair dados relevantes, identificar padrões de comportamento relevantes para o negócio ou área em questão, ajudando assim na tomada de decisão. Essencialmente, as técnicas de *Data Mining* podem ser agrupadas em quatro classes distintas, estas são: algoritmos de *clustering*, algoritmos de classificação, algoritmos de regressão, e regras de associação. (Gomes, 2011)

Outro tipo de análise muito comum em BI são os cubos OLAP que são uma tecnologia que organiza os dados, sobre visões multidimensionais de um modelo de relação de dados, onde podem efetuar algumas operações como de corte (recorte de plano ou sub-cubo),

rotação (mudar a visão de plano), perfuração (visão detalhada) entre outras (Castro, 2016).

2.3.1.5. Apresentação dos Dados

A última fase do processo de BI corresponde a apresentação dos dados, durante esta etapa são necessárias ferramentas ou tecnologias, que permitam a análise dos dados, sob a forma de relatórios, gráficos, infográficos, *dashboards*, entre outros. Nesta fase as ferramentas devem permitir a interpretação imediata e intuitiva dos dados, dependendo do tipo de informação e do utilizador final. As organizações ao implementarem estas tecnologias irão conseguir lidar com o *information overload*. Este fenómeno decorre quando órgãos de gestão são bombardeados frequentemente com informação pouco organizada e desprovida de uma boa apresentação (Castro, 2016).

Os *dashboards* são um exemplo de uma ferramenta de BI, que permitem planear objetivos, estratégias a seguir, monitorizar atividades e, ainda, comunicar informações a *stakeholders* importantes (Silva, 2021).

Durante o desenvolvimento de *dashboard* é crucial haver um alinhamento com o negócio para refletir de forma correta e eficaz o propósito do mesmo (Few, 2006).

Além deste alinhamento o *dashboard* deve apresentar um conjunto de características, inerentes ao acrónimo SMART: (1) “*Synergetic*” que possibilita ao utilizador desenvolver sinergias relativamente às informação apresentada; (2) “*Monitor KPI’s*”, expõe os KPI’s determinantes, para a tomada de decisão; (3) “*Accurate*”, disponibiliza informação precisa de modo a transmitir confiança aos indivíduos que visualizam o dashboard; (4) “*Responsive*”, responde a limites predefinidas, através da criação de alertas, com o intuito de avisar o utilizador para assuntos relevantes; (5) “*Timely*”, apresenta informações o mais atuais possíveis para garantir uma tomada de decisão eficaz (Malik, 2005).

Além destas características é crucial um *dashboard* utilizar um design eficaz, que irá permitir a comunicação de aspetos chave aos utilizadores criando um acesso fácil à informação de suporte da organização (Peters & Brath, 2004). O design de um *dashboard* considera dois tipos de características: visuais e funcionais. As características visuais relacionam-se com os princípios da visualização de dados, isto é, quão eficazmente e eficientemente a informação é apresentada ao utilizador final. As características

funcionais estão indiretamente relacionadas com a visualização, mas descrevem o que se consegue fazer com o *dashboard* (Yigitbasioglu & Velcu, 2012).

Michael Peters e Richard Brath (2004) sugere que para alcançar a eficácia e eficiência da visualização é necessário o uso adequado dos visuais às métricas pretendidas, pois alguns gráficos podem causar o enviesamento da decisão ao causar ilusões visuais. Outras recomendações partilhadas incluem o cuidado no excessivo uso de cores para realçar informação pertinente, uma vez que, esta pode distrair o utilizador e ainda maximizar o rácio *data-ink* que mede a proporção entre a tinta ou pixels usados para apresentar dados e a total quantidade de tinta ou pixels usados no *dashboard* inteiro (Yigitbasioglu & Velcu, 2012).

Apesar das recomendações propostas, Few (2006) identifica um conjunto de erros comuns no desenho de *dashboards* que devem ser evitados. No total foram apontados treze erros, nomeadamente: a informação excede a dimensão do ecrã; contexto inadequado; excesso de detalhes; uso de métricas inadequadas; elementos visuais pouco ilustrativos; uso desadequado do mesmo visual ilustrativo; má projeção de visuais ilustrativos; exibir dados quantitativos de forma pouco precisa; organização inadequada de conteúdos; destaque ineficaz de dados relevantes; preencher o ecrã com elementos decorativos desnecessários; uso excessivo ou indevido de cores; projetar um display visual pouco atrativo.

2.4.Outros Projetos Desenvolvidos por IES

Uma das primeiras Instituições de Ensino Superior portuguesas a mapear os Objetivos de Desenvolvimento Sustentável foi a Universidade de Évora, reforçando assim o seu alinhamento com as políticas traçadas na Agenda 2030 das Nações Unidas. Neste processo primeiramente foi necessário reconhecer os contributos, que uma determinada unidade curricular (UC) pode dar para os ODS, onde se verificou que existem UCs específicas sobre ODS e, por outro, UCs que, através dos seus resultados de aprendizagem, podem contribuir para alguns ODS. Assim, se os objetivos de uma determinada UC estiverem alinhados com um ou mais ODS ou se uma UC fornecer aos alunos os conhecimentos para dar resposta aos desafios relacionados com os ODS, os mesmos serão indicados via online na respetiva página web da disciplina. Este projeto adota uma metodologia ágil, que tem envolvido a diversos elementos da comunidade

académica e que coloca a Universidade de Évora a par de outras universidades de referência (Universidade de Évora, 2020).

O projeto “SDGs Labs – Making the SDGs our Business” é outra atividade desenvolvida por outras Instituições de Ensino Superior, que resulta da aliança coordenada pela *Vienna University for Economics and Business* e constituída por outros sete parceiros: Universidade de *Vechta*, *Wiesenhof*, Universidade de Trás-os-Montes e Alto Douro, *Terra Institute*, *Regia-Douro Park*, *ISEKI-Food Association* e *CEIFAcop*.

O *SDGs Labs* tem como objetivo desenvolver colaborações entre indivíduos envolvidos no agronegócio e na produção de bens alimentares, com o objetivo de incorporação holística dos ODS nas práticas empresariais modernas do setor (Regia-Douro Park, 2022).

Para que o objetivo do projeto seja alcançado as IES, as organizações não governamentais e incubadoras de start-ups irão traduzir os ODS abstratos em soluções práticas de negócios, para os atores de toda a cadeia de abastecimento do agronegócio e do setor de produção de alimentos, bem como desenvolver métodos, ferramentas e ambientes de aprendizagem para colaborações futuras (SDGS- Labs, 2022).

3. Metodologia

O ISEG além de ser a primeira escola portuguesa de economia e gestão, tem vindo a implementar medidas estratégicas que permitem garantir a sua qualidade. Uma das medidas adotados foi o desenvolvimento de projetos e/ou estudos com uma abordagem mais sustentável.

Neste âmbito, o presente projeto pretende influenciar a comunidade académica, especialmente os alunos procurando consciencializá-los, para o papel de todos enquanto agentes de transformação da economia, para que esta se torne o mais verde possível e alinhadas com os SDG's (ISEG, 2021).

Tendo por base os desenvolvimentos realizados por Marcelino (2020) e Pontalete (2021), o seguinte projeto, que embora corresponda à continuação do trabalho desenvolvido exigiu primeiramente a definição dos objetivos a serem alcançados, assim como planear as principais atividades a concretizar. A Figura 2 apresenta o fluxograma das fases, que foram realizadas no presente projeto.

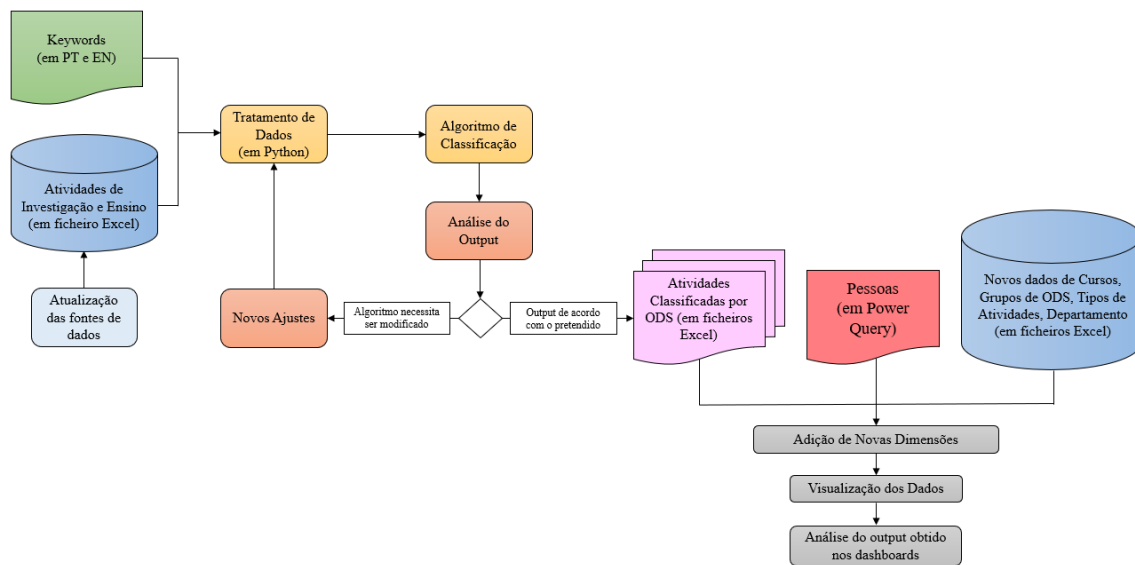


Figura 2 - Fluxograma do Projeto

(Fonte: Adaptado de Portalete, 2021)

No ano letivo 2021/2022, o ISEG, atravessou um período de mudança da plataforma de gestão académica. Durante esse processo os dados do Aquila foram migrados, para a atual plataforma implementada, o Fénix. Neste sentido, foi necessária a atualização das fontes de dados, de forma a que estas fossem ao encontro das necessidades da instituição, e permitissem, a extração dos ficheiros. Embora tenha sido efetuada esta alteração a lista de *keywords* mantivesse intacta, bem como grande parte do código que sofreu pequenas alterações, durante o processo de tratamento de dados nomeadamente, a renomeação de parâmetros analisados e criação de outputs *Microsoft Excel* de menor dimensão, devido a limitação de *Uniform Resource Locator* (URL) por folha de cálculo.

Após serem extraídos e limpos os dados necessários, procedeu-se à análise dos dados, através do desenvolvimento de um modelo *snowflake* onde foram adicionadas novas dimensões, nomeadamente pessoas (“DimPeople”) e departamento (“DimDepartment”). O conteúdo da dimensão pessoas foi desenvolvido manualmente e posteriormente modificado, no *Microsoft Power Query*, para que o seu conteúdo estivesse em conformidade com o das outras dimensões. De seguida, para analisar a informação de forma simples e objetiva foram desenvolvidos novos *dashboards* e atualizados os parâmetros anteriormente estudados.

Por último, para um estudo mais detalhado verificou-se a necessidade de desenvolver a comparação entre os resultados dos *dashboard* desatualizados e os novos.

3.1. Enquadramento do Projeto

O presente documento corresponde a terceira fase de desenvolvimento de um projeto iniciado em 2020 realizado sob a orientação do ISEG, com base nos seus valores, missão e visão.

Todos os desenvolvimentos do projeto pretendem apoiar a implementação e, conseqüentemente, a integração de técnicas de sustentabilidade nas atividades desenvolvidas no ISEG.

Segundo o SDSN *Australia/Pacific* (2017), as IES podem contribuir para os ODS através de quatro vertentes: a investigação, o ensino, o campus e a sociedade. Embora todas sejam relevantes, para o ISEG, no âmbito deste projeto, apenas serão consideradas duas vertentes, nomeadamente a investigação e o ensino. Nas atividades de investigação foram consideradas as teses de mestrado, teses de doutoramento e artigos científicos. Relativamente as atividades de ensino considerou-se as unidades curriculares dos anos 2019 e 2021.

O primeiro desenvolvimento realizado por Marcelino (2020) teve como finalidade classificar e comunicar aos *end-user* o impacto das atividades de ensino e de investigação realizadas pelo ISEG entre 2011 e 2020, de acordo com a matriz dos ODS definida na Agenda 2030, através da edificação de *dashboards* (Marcelino, 2020).

O segundo desenvolvimento surgiu da necessidade de melhoria de um algoritmo de classificação dos ODS já existente. Portalete (2021) analisou um conjunto de abordagens a implementar, que permitiu reduzir o número de falsos positivos identificados durante o processo de classificação dos ODS. A adoção do algoritmo de Luhn e a condição do texto analisado conter pelo menos duas *keywords* proporcionou aos utilizadores finais (presidência e grupo de trabalho dos ODS do ISEG) dados mais concisos, de acordo com a realidade da instituição (Portalete, 2021).

Relativamente ao presente projeto surgiu da necessidade de integrar novas dimensões, de forma a contemplar todas as atividades da instituição e a permitir desenvolver outras análises relevantes para o ISEG. Assim, o modelo dimensional de dados sofreu algumas

alterações e foram atualizados *dashboards*, que quantificam as atividades de ensino e de investigação realizadas pela instituição durante o período de 2011 a 2021.

3.2. Identificação das Fontes de Dados

As fontes de dados identificadas por Marcelino (2020) e continuamente adotadas por Portalete (2021), sobre formato ficheiro Excel, cuja origem são os repositórios interno (Aquila) e externo (SCOPUS) do ISEG, mantiveram-se. No entanto, foi adicionado uma nova fonte ao repositório interno, o Fénix.

Como já foi referido, anteriormente, durante o ano letivo 2021/2022, O ISEG, atravessou um período de mudança da plataforma de gestão académica. Levando a que os dados localizados na plataforma Aquila fossem transitados, para a atual plataforma implementada pela instituição, o Fénix. Assim o repositório interno passou a ser o Fénix e o Aquila e o repositório externo, apenas foi considerada a plataforma Scopus.

Após serem definidas as diferentes fontes de dados, foram extraídos e por sua vez atualizados os ficheiros necessários.

Primeiramente, temos um ficheiro “Scopus”, proveniente da plataforma Scopus, que corresponde à listagem de informação, relativamente aos artigos científicos que foram realizados por docentes e/ou estudantes; um ficheiro para “Master” e outro para “PhD”, que irão listar os trabalhos finais de mestrado e doutoramento, respetivamente, desenvolvidos no ISEG; e um ficheiro “Course Units”, que irá identificar as unidades curriculares lecionadas nos anos 2019 e 2021 no ISEG. Os ficheiros “Master”, “PhD” e “Course Units”, provém das plataformas Aquila e/ou Fénix, com exceção do ficheiro “Scopus”.

3.3. Conjunto de *Keywords*

Para alcançar o principal objetivo do projeto, foi necessário adotar um conjunto de *keywords* capazes de identificar os diferentes 17 ODS. Assim, decidiu-se manter a lista já identificada por Marcelino (2020) que culminou da análise de três listas de *keywords* provenientes de diferentes fontes, nomeadamente: Monash realizada pela Universidade de Monash e pela SDSN *Australia/Pacific* ; Confraria que é uma compilação, que engloba os 16 primeiros ODS (Marcelino, 2020); e Siris desenvolvida pela consultora SIRIS *Academic* e engloba os 16 primeiros ODS (Duran-Silva et al., 2019).

Esta lista é composta por um conjunto de *keywords* em inglês e português, que contêm sinónimos e palavras derivadas, que caracterizam cada ODS.

3.4. Ambiente de Desenvolvimento

Considerando que o presente projeto corresponde à continuação do trabalho desenvolvido por Marcelino (2020) e Portalete (2021), não foram feitas alterações ao nível do ambiente de desenvolvimento.

Assim a linguagem de programação adotada, durante o tratamento de dados foi o Python, que se trata de uma linguagem de programação *high-level*, orientada a objetos e lançada por Guido Van Rossum em 1991 (Portalete, 2021).

O Python possui uma sintaxe clara e concisa, que favorece a legibilidade do código fonte. Além disso fornece diversas estruturas de alto nível, nomeadamente, listas, dicionários, bibliotecas, entre outras, que suportam uma diversidade de tarefas de programação (Borges, 2010).

Esta linguagem, foi desenvolvida com o intuito de ser de fácil leitura e compreensão, por exemplo evitar o uso excessivo de pontuação sendo um dos traços característicos, que a torna mais simples. Assim, os programadores não irão despender tempo com a complexidade da linguagem, mas sim a pensar no problema em questão (Yegulalp, 2020).

Esta escolha, teve em consideração o contexto em que se encontra o presente projeto, bem como outros fatores, nomeadamente: o *Python* é uma das linguagens de programação lecionadas durante o mestrado, pelo que existe familiaridade na sua utilização; por ser fácil de perceber e estruturar algoritmos para analisar grandes conjunto de dados (Kruglyk & Lvov, 2012); dispõe de um vasto leque de estruturas de alto nível úteis, para o tratamento e análise de dados; e pode ser programado num software cuja instalação e licenciamento é gratuito, por exemplo, o Anaconda (Borges, 2010).

Assim, o *software* escolhido foi a Anaconda, que se trata de uma plataforma *open-source*, com diversos *packages* e bibliotecas, focada para *data science* e *machine learning* (Anaconda, 2020). A plataforma Anaconda além de disponibilizar fácil instalação e licenciamento gratuito, apresenta uma interface *user-friendly* e possibilita a utilização da linguagem *Python*. Neste contexto é utilizado o *notebooks Jupyter*, que permitem criar e editar documentos cujo input e output por *default* será em linguagem *Python* ou R. Em cada *notebook*, são importadas as bibliotecas, que consistem num conjunto de módulos,

que contém diversas funções, pré-escritas. Estas funções podem ser utilizadas durante a construção do código sem que seja necessária uma instrução de importação o que facilita a programação, uma vez que reduz a necessidade de gerar novos comandos de código (Python, 2020).

No contexto em que está inserido este projeto as bibliotecas mais importantes utilizadas são *NumPy*, *Pandas*, *FlashText* e *Natural Language ToolKit* (NLTK).

A biblioteca *NumPy* (*Numerical Python*) é reconhecida por fornecer estruturas de alto nível mais especificamente, *array*'s multidimensionais, em que os valores/elementos podem ser identificados por um índice. Esta biblioteca suporta, assim, uma abordagem orientada a objetos e funções matemáticas nomeadamente a lógica, a manipulação de forma, a álgebra, e outros tipos de operações numa larga escala de dados (NumPy, 2020).

A utilização da biblioteca *Pandas* permite trabalhar com uma enorme variedade de estruturas de dados e funções desenhadas com o intuito de tornar mais fácil e rápida a manipulação de dados. Além disso, esta biblioteca ainda cria dois tipos de estruturas de dados: séries que são objeto com linhas, de uma só dimensão e os *dataframes* que correspondem a objetos em formato de tabelas (McKinney, 2013).

A *FlashText* é uma biblioteca criada com o propósito de procurar e substituir vocábulos em textos ou lista de palavras, que se baseia no algoritmo de *Aho-Corasick* e de *Trie Dictionary*. Estes vocábulos, que são habitualmente chamadas de *keywords*, serão procuradas e/ou substituídas na frase fornecida. Esta biblioteca, apenas identifica *keywords* completas compreendidas entre os caracteres de limite, *<start>keyword<eot>*, em que o *start* e o *eot* representam um espaço e sinais de pontuação (Singh, 2017).

Por último, a biblioteca NLTK tem como finalidade trabalhar com dados de linguagem natural, isto é, linguagem humana. A mesma dispõe de diversas opções para processamento de texto, nomeadamente análise, classificação e raciocínio semântico (NSS, 2017).

À semelhança do que já tinha sido desenvolvido anteriormente por Marcelino (2020) e Portalete (2021), o *software Microsoft Power BI* manteve-se, para elaborar os elementos visuais do projeto. Esta escolha deveu-se ao facto do *Power BI* permitir alterar e customizar os *dashboards* de forma simples e por último, ser uma ferramenta Microsoft e estar integrada com outras aplicações utilizadas pelo ISEG (Marcelino, 2020).

Esta ferramenta lançada em 2015 caracteriza-se por: ser uma ferramenta *user-friendly*, no que respeita a análises avançadas; permitir integrar dados de diversas fontes que podem ser locais ou em *cloud*; possibilitar a preparação, análise e a visualização de dados; desenvolver relatórios de forma interativa, que podem ser partilhados com outros utilizadores da organização (Microsoft, 2020).

Adicionalmente, existem componentes que auxiliam na criação e partilha de relatórios, nomeadamente, o *Power Query*, uma ferramenta de extração e conexão de dados, provenientes de diversas fontes, responsável pela sua transformação (Rad, 2019). Esta ferramenta foi utilizada essencialmente para o desenvolvimento de novas dimensões, como por exemplo a tabela “DimPeople” (4.3).

4. Desenvolvimento do Projeto

No decorrer deste capítulo vão ser apresentados em detalhe os diferentes componentes desenvolvidos, ao longo do projeto. Cada componente é igualmente importante, para classificar e apresentar os contributos das atividades de ensino e de investigação realizados pelo ISEG entre 2011 e 2021. No entanto, este desenvolvimento foca na **modificação do modelo dimensional** e no **estudo de novas variáveis pertinentes**.

4.1. Tratamento de Dados

Anaconda e *Microsoft Power BI* foram as ferramentas utilizadas no processo de tratamento de dados. Durante este processo são aplicados um conjunto de alterações necessárias aos ficheiros de onde provém a informação do projeto e que necessitam de estar coerentes. Estes ficheiros continham uma compilação de dados compreendidos maioritariamente entre 2011 e 2021, período estabelecido pelo grupo de trabalho dos ODS. Relativamente, ao ano 2022, uma vez que ainda não estar finalizado, não houve interesse em utilizar os seus dados para este estudo.

Primeiramente no notebook jupyter, para cada um dos ficheiros extraídos foi executado um conjunto de alterações de acordo com as especificidades de cada um, nomeadamente alteração dos parâmetros analisados e respetiva renomeação (4.1.1.), bem como dividir outputs em formato Excel (4.1.2.).

Após o “primeiro” tratamento de dados realizado no notebook jupyter, todos os dados das tabelas de facto, foram transferidos para o *software Microsoft Power BI*. Aqui com o

auxílio da interface *Power Query* foi possível apagar parâmetros de informação, renomear parâmetros através da opção “Utilizar Primeira Linha como Cabeçalhos”, unir conjuntos de dados numa única tabela permitindo e assim desenvolver novas dimensões (4.1.3), substituir valores específicos selecionando (4.1.4) entre outras atividades. Para algumas tabelas desenvolvidas foi necessário anular a opção “Ativar Carregamento”, uma vez que não eram pertinentes para o modelo dimensional (4.3) e/ou para as análises que são desenvolvidas nos *dashboards* (4.4).

Estas alterações permitiram automatizar ao máximo o processo de transformação, obter coerência entre os atributos das colunas das diferentes dimensões e a possibilidade de as interligar numa fase posterior (4.3).

4.1.1. Alteração dos Parâmetros Analisados em *Python*

Durante a análise dos dados extraídos das diferentes fontes, foi necessário primeiramente redirecionar o input atualizado. Após este passo tendo por base os parâmetros utilizados no segundo desenvolvimento, as necessidades atuais do ISEG e dados disponibilizados foram renomeados novos parâmetros. A figura 3 corresponde aos parâmetros utilizados no segundo desenvolvimento e a figura 4 diz respeito ao presente desenvolvimento.

```
dataFrame = pd.read_excel(r"C:\Users\maria.c.de.sa\Desktop\ISEG\Tese\Melhorias\INPUT PYTHON\AQUILA - listaTesesRealizadas MASTER
.rename(columns={
    'Número': 'Author',
    'Título do TFM': 'Title',
    'Abstract PT': 'Abstract_PT',
    'Abstract EN': 'Abstract_EN',
    'Palavras Chave PT': 'Keywords_PT',
    'Palavras Chave EN': 'Keywords_EN',
    'Tipo': 'Type',
    'Data': 'Date',
    'Mestrado': 'Course',
    'Orientador': 'Advisor'})
```

Figura 3 - Renomeação dos parâmetros dos dados sobre as teses desenvolvidas nos mestrados entre 2011 e 2020

```
dataFrame = pd.read_excel(r"C:\Users\ASUS\Desktop\ISEG\tese\input python ou power BI\python\teses de mestrado.xlsx") \
.rename(columns={
    'Autor': 'Author',
    'Título': 'Title',
    'Abstract PT': 'Abstract_PT',
    'Abstract EN': 'Abstract_EN',
    'Palavras Chave PT': 'Keywords_PT',
    'Palavras Chave EN': 'Keywords_EN',
    'Tipo de TFM': 'Type',
    'Data de realização da prova': 'Date',
    'Curso': 'Course',
    'Orientador': 'Advisor'})
```

Figura 4 - Renomeação dos parâmetros dos dados sobre as teses desenvolvidas nos mestrados entre 2011 e 2021

4.1.2. Divisão de *Outputs* em Formato Excel em *Python*

Para transitar os dados da ferramenta Anaconda para o *Power BI* o formato de dados escolhido foi o documento Excel. Embora o *software Microsoft Excel* suporte uma vasta quantidade de dados apresenta algumas limitações predefinidas, nomeadamente a largura

das colunas não pode exceder os 255 caracteres, o número máximo de URL por folha de cálculo é de 65530 linhas, entre outras. Assim para solucionar a limitação do número de URL foi efetuada a divisão do output, através do desenvolvimento de duas variáveis que recebem os dados de uma *dataframe* “df_final” e retornam dois intervalos de dados (figura 5).

```
In [39]: df_1 = df_final.iloc[:46716,:]
df_2 = df_final.iloc[46717:,:]
```

Figura 5 - Algoritmo de divisão dos outputs

4.1.3. Unir outputs no Power Query

O *Power Query* permite de forma intuitiva unir informação, através da funcionalidade “Acrescentar Consulta”, localizada na opção “Base” do menu principal. A opção “Acrescentar Consulta” foi utilizada neste desenvolvimento pelo menos uma vez, nomeadamente na dimensão “FactScopus”. Esta dimensão surgiu da união de dois outputs “FonteScopus1” e “FonteScopus2” provenientes da ferramenta Anaconda. Como eram necessários estes dois outputs numa única dimensão, estes foram unidos em ambiente *Microsoft Power Query*, que suportava a quantidade de dados. A figura 6 representa de forma simplificada o processo de criação da tabela.

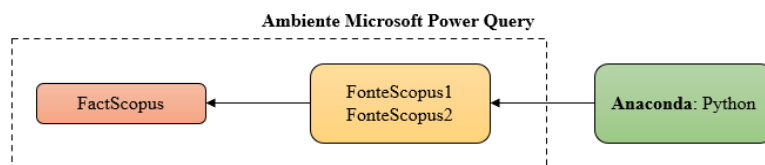
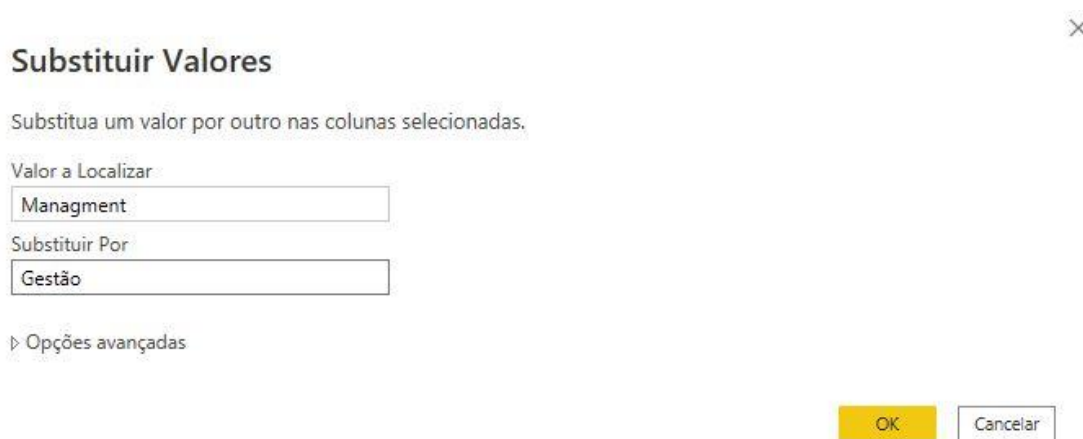


Figura 6 - Processo de criação das tabelas em ambiente Microsoft Power Query

(Fonte: Próprio Autor)

4.1.4. Substituição de Valores

Durante o desenvolvimento foi necessário substituir valores, para que as dimensões se relacionassem. Assim, através da funcionalidade “Substituir Valores...” foi disponibilizado um formulário (Figura 7), que permite identificar primeiramente o dado que se pretende modificar e de seguida solicita o novo.



Substituir Valores ✕

Substitua um valor por outro nas colunas selecionadas.

Valor a Localizar

Substituir Por

▸ Opções avançadas

Figura 7 - Formulário "Substituir Valores"

4.2. Algoritmo de Classificação dos ODS

Neste desenvolvimento manteve-se o algoritmo de classificação aplicado no segundo desenvolvimento realizado por Portalete (2020). O algoritmo vai identificar um conjunto de *keywords*, onde, para cada variável do ficheiro, caso nesse texto exista uma *keyword*, é feita a associação da *keyword* ao/s respectivo(s) ODS a que essa *keyword* corresponde. Assim, as atividades analisadas podem ser classificadas a nenhum, a um ou a vários ODS, sendo que cada ODS só é contabilizado uma vez, independentemente de existir uma ou várias *keywords* identificadas (Portalete, 2020).

4.3. Modelo Dimensional

Relativamente ao modelo dimensional, manteve-se a estrutura *snowflake*, que foi o modelo adotado por Marcelino (2020) e Portalete (2021). Este tipo de estrutura consiste num conjunto de tabelas de facto, associadas a uma ou várias tabelas de dimensão, que por sua vez estão associadas a uma ou várias tabelas de dimensão. Embora a estrutura *snowflake* seja mais complexa comparativamente a outros modelos, existe um interesse em ter tabelas de dimensões a relacionarem-se entre si, que normalizam as tabelas de dimensão. Identicamente ao primeiro desenvolvimento “foi dada prevalência à não repetição das informações ao logo das tabelas de dimensões principais ao contrário da rapidez de processamento, uma vez que os dados vão sofrer atualizações anuais” (Marcelino, 2020, p.25).

Uma das melhorias futuras apontadas por Marcelino (2020) para este projeto, foi integrar novas fontes de dados, de forma a contemplar todas as atividades do ISEG e a possibilidade de desenvolver análises e adquirir conhecimentos relevantes para a

instituição, bem como adicionar uma nova dimensão denominada “autores” juntamente com a respetiva filiação.

Assim o *Microsoft Power BI*, no separador Modelo, foi a ferramenta implementada para o desenvolvimento do novo modelo dimensional (figura 8), que irá apresentar novas dimensões, nomeadamente a “DimPeople” que engloba docentes do ISEG.

Para chegar a atual versão do modelo dimensional foram testados e propostos outros modelos (anexo I) que tentavam solucionar a problemática da dimensão “autores”.

Identicamente ao primeiro desenvolvimento as fontes de dados identificadas no subcapítulo 3.2. correspondem às tabelas de factos. Assim a tabela “FactMaster” e “FactPhD” vão ser agrupadas numa tabela facto “FactThesis”, devido ao facto de terem a mesma granularidade e ambas consistirem em trabalhos finais de curso. O modelo final apresenta três tabelas de facto, sendo que cada uma das tabelas refere-se a atividades diferentes.

No que toca às tabelas de facto, foram desenvolvidas três, nomeadamente:

- FactThesis – Tem como fonte de dados o Aquila, lista os trabalhos finais de mestrado e de doutoramento apresentados entre 2011 e 2019 e os trabalhos finais de mestrado de 2020 a 2021;
- FactCourseUnits – Apresenta as unidades curriculares lecionadas em 2019 e 2021. Esta tabela ao contrário das outras tabelas de facto apresenta duas fontes de dados, Fénix e Aquila;
- FactScopus – Identifica, entre 2011 e 2021, os artigos científicos desenvolvidos e tem como fonte dados o Scopus.

Relativamente às tabelas de dimensão foram mantidas cinco (DimSDG, DimGroup, DimCourse, DimType e DimDate) e desenvolvidas quatro novas tabelas, que englobam atributos únicos que descrevem os dados das tabelas de facto, estas são:

- DimSDG – Enumera os 17 ODS com as suas respetivas descrições;
- DimGroup – Apresenta as cinco áreas, a que os ODS pertencem, com as suas respetivas descrições;
- DimCourse – Identifica todos os cursos lecionados no ISEG e é proveniente do Scopus;

- DimDepartment – Lista os departamentos do ISEG, que podem ser “Ciências Sociais”, “Economia”, “Gestão” e “Matemática”;
- DimType – Refere a lista de todas as atividades consideradas no mapeamento, que podem ser classificadas como “artigos”, “tese”, “projetos”, “estágios” e unidades curriculares”;
- DimDate – Apresenta uma lista de todos os dias compreendidos deste 1/1/2011 até à data atual;
- DimPeople – Lista os docentes do ISEG e respetivos departamentos a que pertencem. Esta tabela exigiu a necessidade de atribuir de forma manual os departamentos a cada docente.

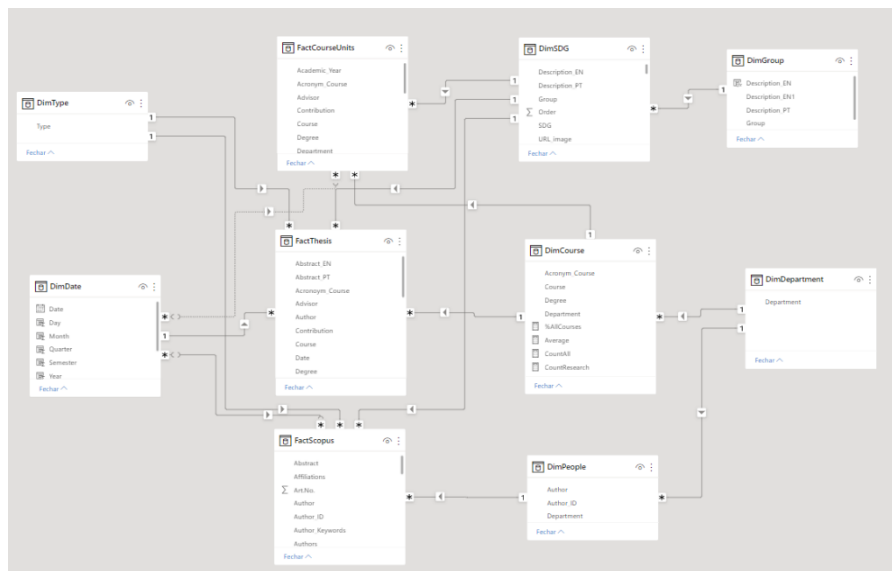


Figura 8 - Modelo dimensional de dados

4.4. Visualização dos Dados

Para visualizar os dados é necessário existir uma etapa, que retrate de que forma esta poderá ser realizada. Assim, tendo como ponto de partida o trabalho desenvolvido por Portalete (2021), foram utilizados os *dashboards* concebidos em *Power BI*, que procuram evidenciar a informação tratada em *Python*, proveniente dos ficheiros Excel. Nesta última etapa foram projetados e implementados novos *dashboard*, bem como atualizados os

dashboards desenvolvidos por Portalete (2021), tendo em consideração as necessidades dos diferentes end-users e as diretrizes fornecidas pelo ISEG.

Os *dashboards* deste desenvolvimento apresentam uma visão global das contribuições e uma análise mais aprofundada, sobre cada uma das atividades realizadas pelo ISEG. Durante o processo de construção dos *dashboards*, no *Power BI*, foi necessário elaborar um conjunto de métricas em linguagem *Data Analysis Expressions* (DAX), que são armazenadas nas respetivas tabelas a que correspondem (Kimball & Ross, 2013).

Através da linguagem DAX é possível extrair valores pertinentes ao estudo, como por exemplo valores numéricos específicos e percentuais das contribuições identificadas, a média de contribuições por ano, entre outras medidas.

Além da construção destas métricas foi necessário aplicar um conjunto de vários tipos de filtros, interno e externos, para a obtenção de informações mais detalhadas possíveis. Relativamente aos filtros internos, localizados geralmente no canto superior esquerdo dos *dashboards* foram desenvolvidos os seguintes filtros: barras temporais que compreendem a identificação do intervalo temporal, uma grelha que permite selecionar um ou mais ODS e uma segmentação de dados com orientação horizontal e fundos coloridos, para os diversos itens. Estes itens podem ser atividades, tipo de pessoa, grupo de ODS, curso, grau académico, autor/orientador e departamentos.

No que toca aos filtros externos, localizado na barra “Filtro”, secção “Filtros nesta página”, quando utilizados foram aplicados em todos os elementos do *dashboard* ou em métricas específicas. O filtro externo, apenas filtram dados temporais em anos específicos.

Foram atualizados doze *dashboards* e desenvolvidos nove *dashboards*. Entre os doze *dashboards* atualizados quatro sofreram uma pequena alteração na medida em que lhes foi adicionado um novo filtro interno “*Departement*” que permitia filtrar os dados por departamento. Importa referir que ao oposto do primeiro e segundo desenvolvimentos, as unidades curriculares apresentam uma data de realização, uma vez que houve interesse por parte dos *end-user* comparar as contribuições das UCs de 2019 *versus* 2021.

4.4.1. Análise e Discussão dos Novos *Dashboards*

Ao longo do projeto, como já referido, criaram-se nove *dashboards*. Estes tiveram em consideração as novas dimensões presentes neste desenvolvimento, bem como o design implementado no passado e as necessidades inerentes ao ISEG.

O *dashboard* da figura 9 descrever um conjunto de parâmetros, sobre a investigação durante o ano 2021. Primeiramente foi desenvolvido um gráfico de barras, que estuda o número de contribuições da investigação por ODS, desta forma, é possível observar as ODS com maior número de contribuições na investigação. Para complementar a informação são ainda indicados os valores numéricos dos artigos, das teses de mestrado e contribuições totais da investigação. Em paralelo a estes valores são apresentadas duas listas, um dos artigos e outra das teses de mestrado e doutoramento. Assim, para permitir a análise mais detalhada foram adicionados um conjunto de quatro filtros. Neste sentido, três estão localizados no canto superior esquerdo, que são “Choose the SDG”, “Type of Contribution”, “Group of SDG” e “Department” e o quarto filtro (externo) filtra todos os parâmetros do *dashboard* de forma a que, apenas os dados do ano de 2021 sejam considerados.

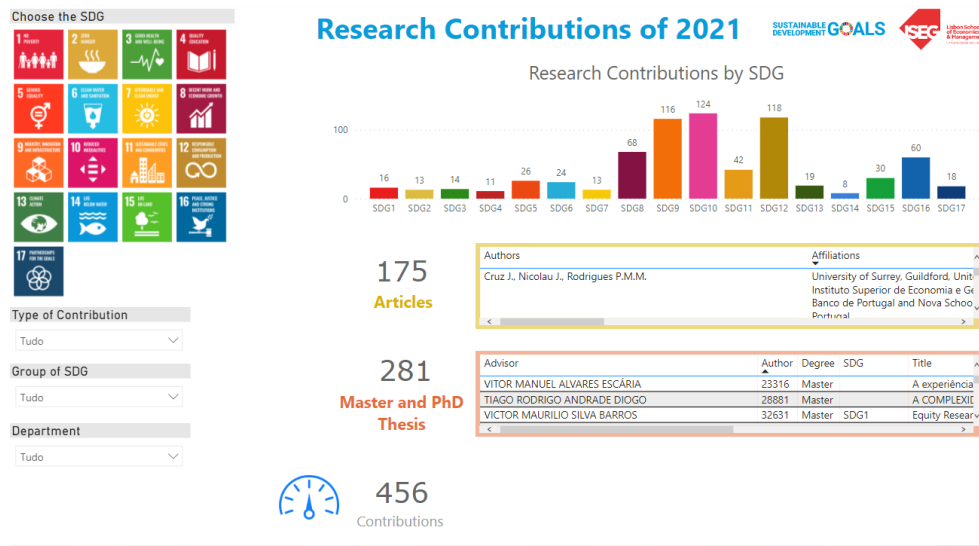


Figura 9 - Dashboard Research Contribution of 2021

Como alternativa ao *dashboard* da figura 9 foi elaborada uma segunda versão, figura 10. No que respeita aos parâmetros analisados são apresentados os valores percentuais, os valores numéricos e as listas dos artigos, bem como das teses de mestrado e doutoramento. Relativamente aos filtros, apenas foram utilizados dois filtros, um interno e outro externo. O interno permite ao utilizador filtrar a informação por SDG e o filtro externo é aplicado

a todos os parâmetros em estudo de forma a que, apenas os dados de 2021 sejam considerados.

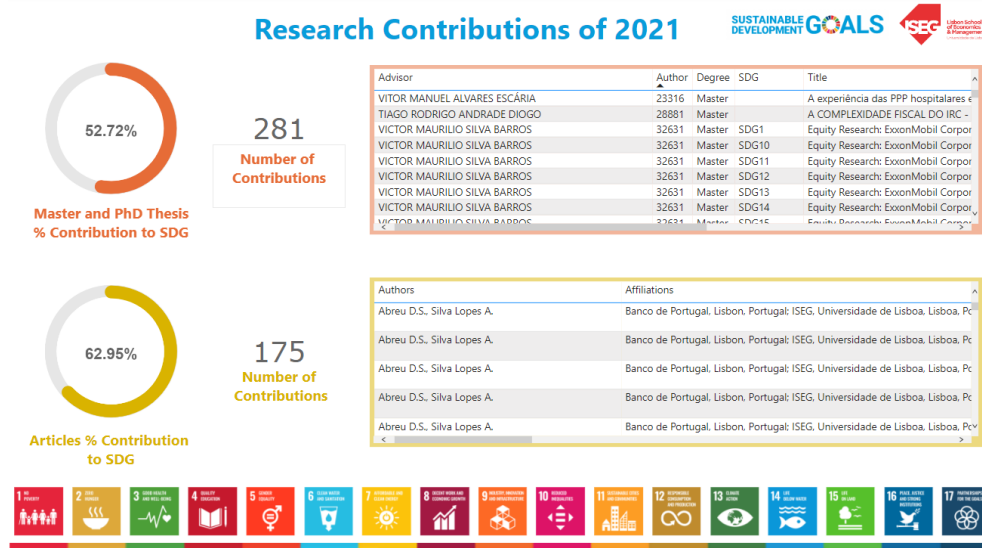


Figura 10 - Alternativa do Dashboard Research Contributions of 2021

Assim, através da análise destes dois *dashboards* é possível verificar, que na investigação as teses de mestrado e doutoramento contribuíram em maior quantidade, para os ODS em 2021. O ODS 9 (Indústria, Inovação e Infraestruturas), o ODS 10 (Reduzir as Desigualdades) e o ODS 12 (Produção e Consumo Sustentáveis) de acordo com o *dashboard* da figura 9 identicamente aos anos anteriores continuam a ser os objetivos, que apresentam maior número de contribuições. Por outro lado, os ODS com menor número de contribuições são os objetivos 14 (Proteger a Vida Marinha), 4 (Educação de Qualidade), 2 (Erradicar a Fome) e 7 (Energias Renováveis e Acessíveis).

De seguida, tendo por base o *dashboar* “Master and PhD Thesis” desenvolvido por Portalete (2020) foi criado um *dashboard*, figura 11, que faz uma análise aos mesmos parâmetros, no entanto considera apenas os dados incompletos do ano 2021. Posicionado no centro, no novo *dashboard*, são apresentadas duas métricas, a percentagem de trabalhos finais de mestrado e doutoramento que contribuem para os ODS, bem como o número específico de contribuições realizadas em 2021. Paralelamente a estes valores foram desenvolvidos dois gráficos de barras, o “Course with SDG Thesis”, que mostra os cursos das respetivas contribuições realizadas e o “Top SDG Contributions”, que apresenta os ODS ordenados por número de contribuições. Também foi desenvolvida uma tabela que lista as informações detalhadas, sobre os trabalhos finais realizados durante 2021. Identicamente aos outros *dahboards*, este dispõe de um conjunto de quatro

filtros, três internos e um externo. Assim, “Choose the SDG”, “Type of Contribution”, “Advisor” e “Department” são os filtros internos que permitem realizar uma análise mais completa e assim contribuir para a boa tomada de decisões. Relativamente ao filtro externo, este também irá filtrar todos os parâmetros temporalmente, para o ano 2021. Numa breve análise é possível identificar, que em 2021 os trabalhos finais de mestrado contribuíram mais, para os ODS e que no total foram identificadas 281 contribuições. Os objetivos que apresentam maior número de contribuições são o ODS 8 (Trabalho Digno e Crescimento Económico), ODS 9 (Indústria, Inovação e Infraestruturas), ODS 10 (Reduzir as Desigualdades) e ODS 12 (Produção e Consumo Sustentáveis).

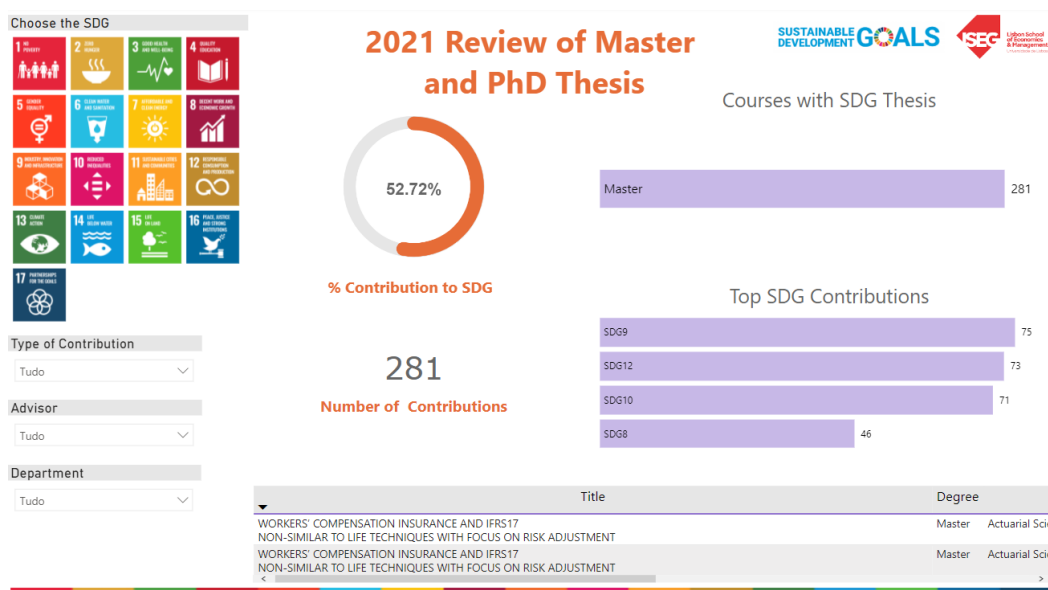


Figura 11 - Dashboard 2021 Review of Master and Phd Thesis

Na figura 12, é desenvolvido um estudo, no período temporal entre os anos de 2019 a 2021, através da utilização de filtros externos que selecionam dados dos anos 2019, 2020 e 2021 separadamente. O *dashboard* apresenta um conjunto de métricas, que dizem respeito ao número específico e percentual das teses de mestrado e doutoramento por ano. Paralelamente a estas métricas foram desenvolvidos três gráficos de barras, que estudam o número de contribuições por SDG em anos específicos: “Contributions by SDG from 2019”, “Contributions by SDG from 2020” e “Contributions by SDG from 2021”

Ao analisar o *dashboard* “Master and PhD Thesis Contributions From 2019 a 2021” o ano que desenvolveu maior número de contributos foi 2019. Ao olhar para os diferentes ODS o que apresenta maior número de contribuições, nos três anos consecutivos foram o décimo e nono objetivos. De modo geral, o número de contribuições aumento em todos

os ODS, principalmente o ODS 8 (Trabalho Digno e Crescimento Económico) e 16 (Paz, Justiça e Instituições Eficazes). No entanto, ao olhar para o objetivo 12 (Produção e Consumo Sustentáveis) verifica-se um decréscimo no ano 2020, que foi colmatado no ano 2021.

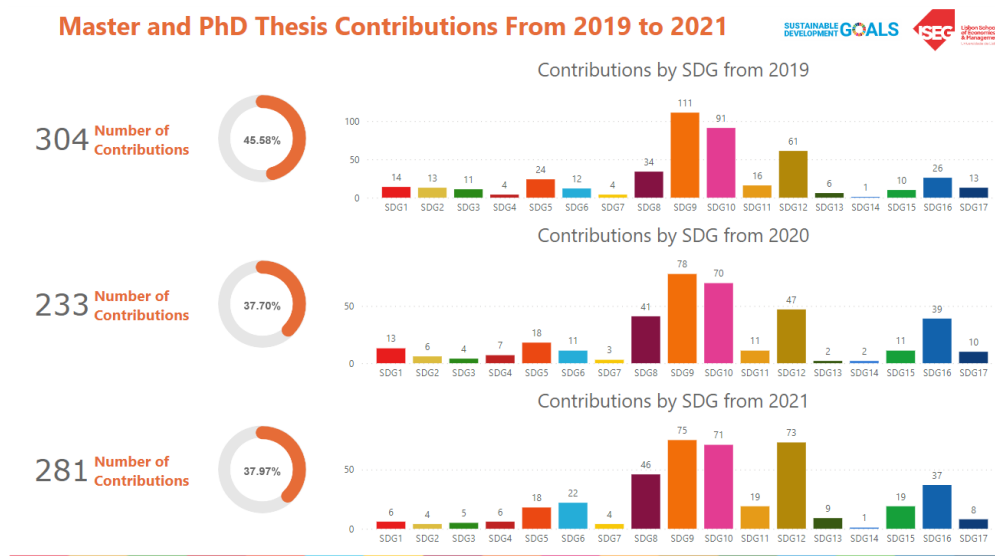


Figura 12 - Dashboard Master and Thesis Contributions from 2019 to 2021

O *dashboard* da figura 13 apresenta métricas idênticas do *dashboard* 11, no entanto faz um estudo sobre os artigos publicados no ano 2021. Identicamente ao que foi desenvolvido anteriormente, o *dashboard* da figura 15 analisa as contribuições nos anos 2019, 2020 e 2021. Numa breve análise é possível verificar que foram identificadas 175 contribuições e que destas, 53 são direcionadas para o décimo ODS e 45 para o décimo segundo objetivo, que diz respeito à Produção e Consumo Sustentáveis.

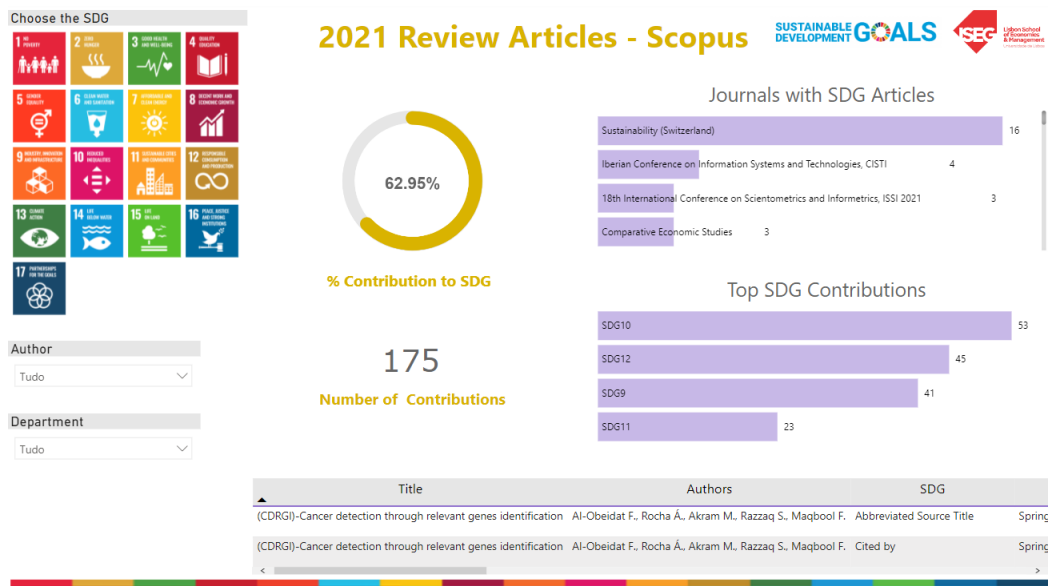


Figura 13 - Dashboard 2021 Review Articles - Scopus

O dashboard da figura 14, como já foi anteriormente referido, faz uma análise dos anos 2019, 2020 e 2021. Assim, verificasse que o décimo ODS continua a apresentar maior número de contribuições. Também existe um aumento no número total de contribuições entre os diferentes anos e pequenas oscilações no número de contribuições por ODS, sendo que o objetivo 6, que diz respeito a água potável e saneamento apresenta o menor número de contribuições.

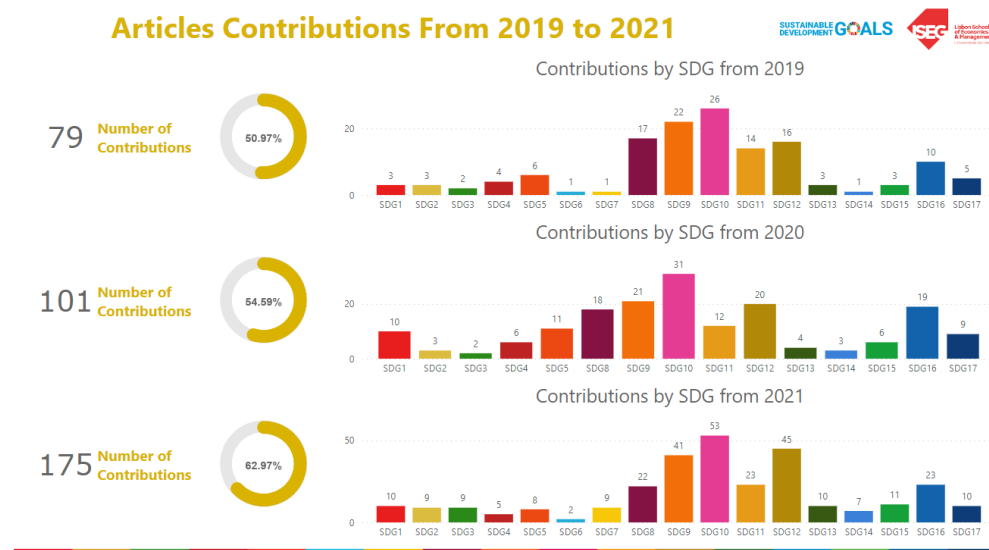


Figura 14 - Dashboard Articles Contributions From 2019 to 2021

O dashboard da figura 15 apresenta métricas idênticas ao dashboard da figura 11, no entanto faz um estudo sobre as contribuições das unidades curriculares. Assim, são identificadas 427 contribuições, destes total 263 contribuições dizem respeito ao ODS 10

(Reduzir as Desigualdades) e 161 ao ODS 9 Indústria, Inovação e Infraestruturas). Ainda é possível verificar “Management” é o curso que contribui mais para a agenda de 2030.

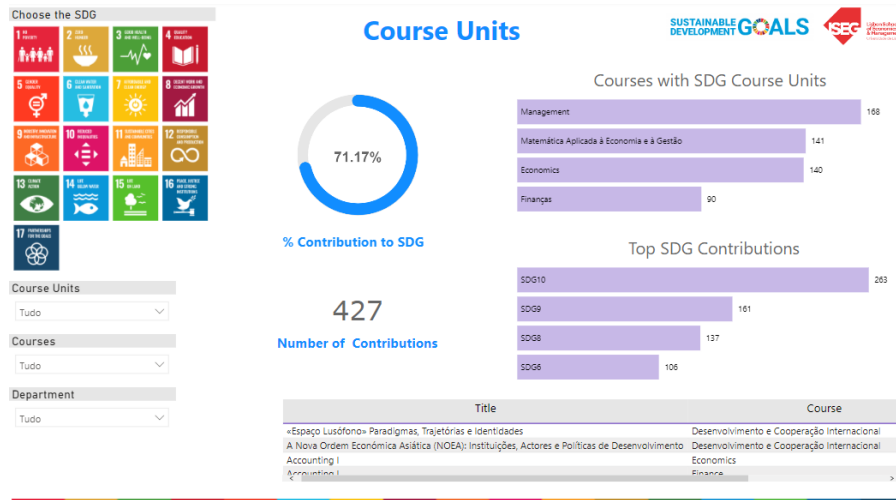


Figura 15 - Dashboard Course Units

Na figura 16 é elaborado um *dashboard* referente às unidades curriculares identificadas nos anos 2019 e 2021. Apresenta quatro métricas, o número específico e percentual das contribuições para os ODS. Em paralelo foram desenvolvidos dois gráficos de barras, que estudam o número de contribuições das UCs por SDG em 2019 e 2021: “Contributions by SDG from 2019” e “Contributions by SDG from 2021”. Apenas foram estudados estes dois anos, uma vez que os dados extraídos são referentes aos mesmos. Assim, verifica-se que as UCs em 2019 apresentam maior número de contributos do que em 2021. No que toca aos ODS em si, o objetivo com maior número de contribuições é o ODS 10 (Reduzir as Desigualdades) em ambos os anos.

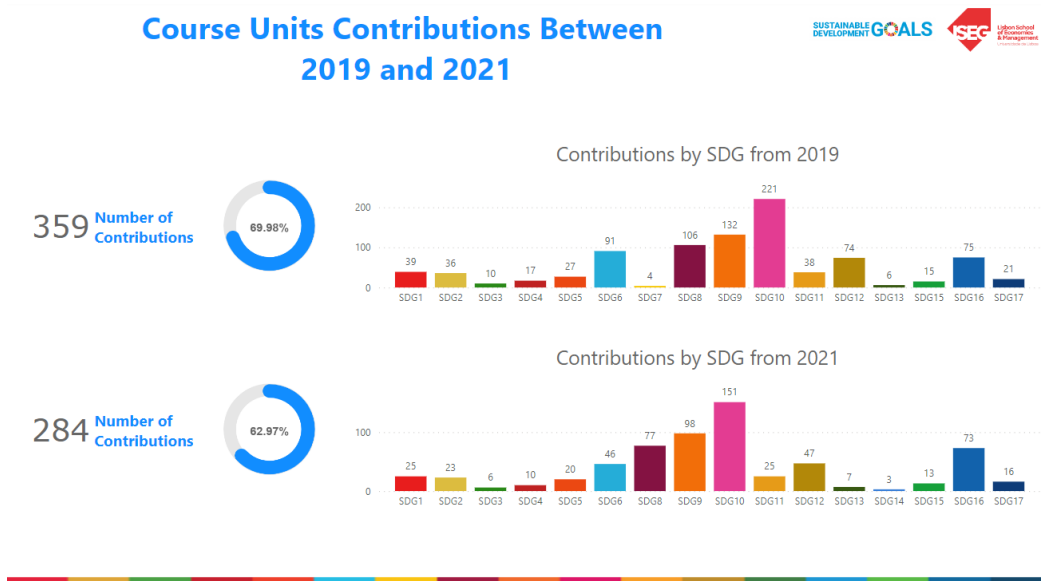


Figura 16 - Dashboard Course Units Contributions Between 2019 and 2021

Por último, foi desenvolvido o *dashboard* da figura 17 que estuda as contribuições desenvolvidas, para os ODS por departamento. Assim, é possível identificar que o departamento de “Economics” e “Management” são os que mais contribuem. Além deste facto, como já era espectável o décimo ODS continua a ser o objetivo com maior número de contribuições.

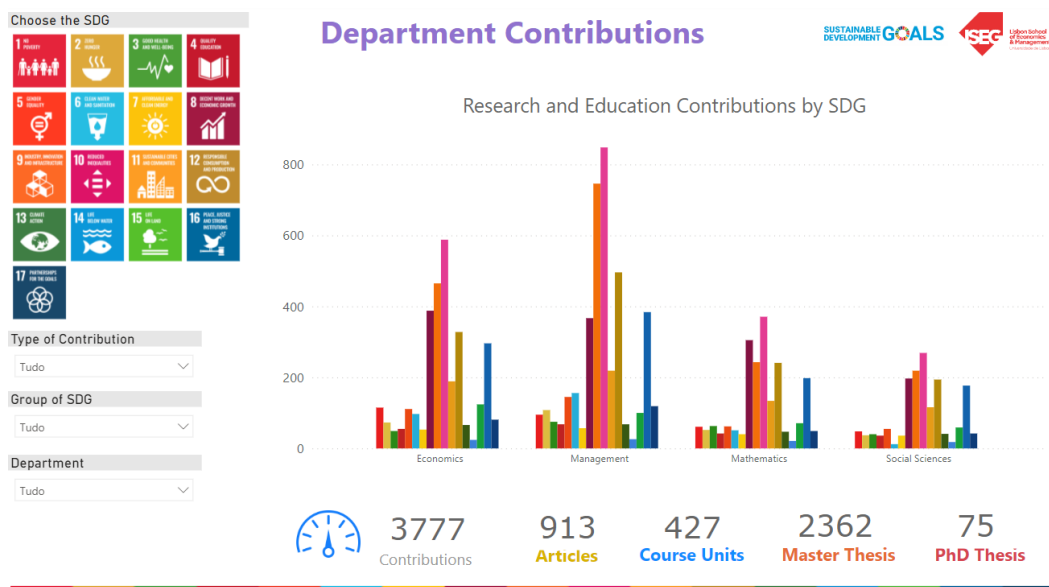


Figura 17 - Dashboard Department

4.4.2. Análise e Discussão dos *Dashboards* Atualizados

Durante este desenvolvimento, foram adicionados novos dados do ano 2020 e 2021 possibilitando analisar um período temporal de dez anos (2011 a 2021). Assim, espera-se que o número de contribuições nestes últimos dois anos aumente. Além desta suposição e dado que os números de colocações no ensino superior aumentaram gradualmente, ao longo dos anos (tabela IV), é expectável um aumento significativo nos valores das contribuições dos trabalhos finais de mestrado e doutoramento, uma vez que estão a ser produzidos em maior quantidade.

Durante os últimos dois anos (2020 e 2021) vários documentos foram desenvolvidos, chegando assim a um total de 3350 contribuições. Constata-se que este valor é superior às 2593 contribuições identificadas no segundo desenvolvimento. Assim, após uma breve análise, verifica-se que o número de contribuições nos artigos nas unidades curriculares e nos trabalhos finais de mestrado e doutoramento, aumentaram globalmente. Neste âmbito, o ODS que apresenta maior número de contribuições foi o décimo objetivo (Reduzir as Desigualdades). Embora o aumento gradual das colocações no ensino superior, tivesse ocorrido, verifica-se um comportamento interessante nos dados. No primeiro ano pandémico foram registadas 383 contribuições em 2020 ocorreu uma quebra, uma vez que foram apenas identificadas 334 contribuições. Esta situação pode dever-se ao facto de os alunos não continuarem e/ou não concluírem o seu percurso académico; ou os trabalhos finais de mestrado não irem ao encontro dos ODS. Para colmatar esta situação possivelmente deveriam ser implementadas medidas de incentivo na continuação e/ou finalização dos estudos académicos, bem como disponibilizar um conjunto de temas inseridos nos diferentes mestrados e doutoramentos, que vão ao encontro dos ODS. Apesar da quebra identificada 2021 surpreendeu pela positiva, uma vez que foi o ano em que se identificou o maior número de contribuições, desde 2011.

Desta forma, dá-se como finalizada a análise e discussão dos *dashboards* desenvolvidos na ferramenta *Microsoft Power BI*, resultante da extração de novos dados e modificação do modelo dimensional.

Tabela IV - Valores das colocações no ensino superior

Ano	1ª Fase	2ª Fase	3ª Fase	Total
2014	37778	10492	1794	50064
2015	42068	9410	1629	53107
2016	42958	9577	1555	54090
2017	44914	9831	1602	56347
2018	43992	9452	1385	54829
2019	44500	9274	1402	55176

(Fonte: DGES, 2021)

5. Conclusão e Propostas de Investigação Futuras

O presente projeto surgiu da necessidade de dar resposta ao ISEG que enquanto instituição de ensino superior é responsável por incorporar os ODS nas suas atividades, na educação, e na própria estratégia de gestão e interação com a sociedade. Assim, Marcelino (2020) propôs a realização do mapeamento e a edificação de um sistema de BI para a identificação dos ODS, nas atividades de ensino e de investigação realizadas pelo ISEG. Mais tarde, em 2021, Portalete (2021) desenvolveu melhorias no algoritmo de classificação o que ajudou na tomada de decisões estratégicas baseada em dados mais adequados à realidade. No entanto, são necessárias melhorias contínuas do sistema, nomeadamente a criação de novas dimensões, bem como atualização dos dados que sucessivamente são produzidos. Assim, surge a terceira fase de desenvolvimento descrito neste documento. Embora os desenvolvimentos sejam diferentes todos tem um objetivo principal em comum, que é classificar e comunicar o impacto das atividades de ensino e de investigação realizadas pelo ISEG para a concretização dos ODS, de forma a permitir a tomada de decisões estratégicas adequadas. Com o desenvolvimento deste projeto foi possível consolidar conhecimento adquirido durante mestrado e familiarizar com uma nova plataforma, *Microsoft Power BI* e a linguagem, *Data Analysis Expressions*.

Para alcançar o principal objetivo foram definidos objetivos secundários. Estes são igualmente importantes. No entanto, o terceiro desenvolvimento foca no objetivo d) (Adição de novas dimensões ao modelo dimensional de dados) e f) (Desenvolvimento de novos *dashboards*, com base na aplicação de boas práticas de desenho), que vão dar resposta a algumas das necessidades de melhoria apontadas por Marcelino (2020), durante o desenvolvimento do seu projeto.

No que toca ao objetivo d) foram desenvolvidas diversas versões dos possíveis modelos dimensionais até chegar ao modelo atualmente aplicado. Os diferentes modelos surgiram conforme as necessidades do ISEG, bem como os dados disponibilizados pela instituição. Relativamente ao ponto f), os novos *dashboards* tiveram em consideração os *dashboards* desenvolvidos por Portalete (2021). Dado que diversas métricas já estavam estudadas, o seguinte desenvolvimento foca-se nas novas dimensões elaboradas no passo d) e na evolução temporal em anos anteriores, durante e após a pandemia Covid-19 (2019, 2020 e 2021).

Relativamente aos outros objetivos também cruciais para o desenvolvimento do projeto, foram atingidos com sucesso apesar das dificuldades sentidas, foi possível retirar conclusões pertinentes para o ISEG. Nomeadamente verificar o aumento gradual do número de contribuições ao longo dos anos; o ODS 10 (**Reduzir as Desigualdades**) é o objetivo que apresenta maior número total de contribuições; e o ano com maior número de contribuições identificadas até ao momento é 2021.

Após três desenvolvimentos não podemos considerar este projeto, na sua globalidade, como terminado já que a sua conclusão poderá ser um ponto de partida, para um novo estudo. No futuro existem melhorias a serem consideradas tanto no processo de desenvolvimento como no projeto em si.

Primeiramente em futuros desenvolvimentos recomendo que o processo de extração de dados seja agilizado, uma vez que é imperativo adquirir todos os dados necessários, antes do tratamento dos mesmos.

Também é indispensável a revisão do projeto num todo e averiguar a possível existência de dados em falta ou incompletos, de novas tecnologias e fontes de dados, que possam melhorar o processo de análise. Além de manter os dados atualizados é importante as tecnologias implementadas acompanhem as necessidades do ISEG.

Durante este desenvolvimento, devido a falta de agilidade na disponibilização de dados, a dimensão “DimPeople” poderia ser melhorada englobando os docentes do ISEG, bem como os autores dos artigos identificados na plataforma Scopus externos ao ISEG. Seria interessante também adicionar uma nova dimensão, que permitisse classificar estas pessoas como internas ou externas à instituição. Ainda neste ambiente os dados em falta da “DimScopus” deveriam ser adicionados.

Deve-se ainda ter em consideração, outras sugestões referidas por Marcelino (2020) e Portalete (2021) nomeadamente, repensar no sistema atual que trata da listagem das *keywords* dos ODS, uma vez que esta é realizada através de uma lista “manual” criada nos *notebooks* do *Jupyter*; e desenvolver uma nova categoria de ODS no mapeamento, que esteja relacionado com as atividades, que contribuam para todos os ODS num todo.

Referências Bibliográficas

Anaconda (2020). Data science technology for a better world [Em linha]. Disponível em: <https://www.anaconda.com/> [Acesso em: 2020/08/22].

Angeloni, M. T. & Reis, E. S. (2006). Business Intelligence como Tecnologia de Suporte a Definição de estratégias para melhoria da qualidade do ensino [Em linha]. Disponível em: http://www.anpad.org.br/diversos/down_zips/10/enanpad2006-adid-0815.pdf [Acesso em 2022/05/02].

Azeroual, O. & Theel, H. (2018) The Effects of Using Business Intelligence Systems on an Excellence Management and Decision-Making Process by Start-Up Companies: A Case Study, *International Journal of Management Science and Business Administration*, 4(3), pp. 30–40.

Barbieri, C. (2001). BI – business intelligence: modelagem e tecnologia, 1ªEd. Rio de Janeiro: Axcel Books.

Borges, L. E. (2010). Python para Desenvolvedores [Em linha]. Disponível em: https://ark4n.files.wordpress.com/2010/01/python_para_desenvolvedores_2ed.pdf [Acesso em: 2022/05/18].

Cabrito, B., Cerdeira, L., Mucharreira, P. & Rodrigues, A., (2019). A importância da relação entre as instituições de ensino superior e a sociedade – colaboração, inovação e transferência de conhecimento [Em linha]. Disponível em: https://publicacoes.riqual.org/wp-content/uploads/2021/06/Forges_19_427_432.pdf [Acesso em: 2022/05/19].

Caldeira, C. P. & Lopes, A. (2018). Aplicação de Técnicas de Business Intelligence a Base de Dados Prosopográficas [Em linha]. [Acesso em: 2022/05/19].

Castro, J. (2016). Tendências de Business Intelligence [Em linha]. Disponível em: <http://dspace.uevora.pt/rdpc/bitstream/10174/23459/1/5983-20467-1-PB.pdf> [Acesso em: 2020/09/22].

Centro de Informação Regional das Nações Unidas para a Europa Ocidental (2018). Guia sobre Desenvolvimento Sustentável [Em linha]. Disponível em: https://unric.org/pt/wp-content/uploads/sites/9/2019/01/SDG_brochure_PT-web.pdf [Acesso em: 2022/05/02].

Coelho, J. (2011). Introdução à Base de Dados [Em linha]. Disponível em: <https://repositorioaberto.uab.pt/bitstream/10400.2/3462/1/Introdu%C3%A7%C3%A3o%20%C3%A0%20Base%20de%20Dados.pdf> [Acesso em: 2022/05/12].

Daniel, L. (2020). Business Intelligence como Fator para Vantagem Competitiva [Em linha]. Disponível em: <https://www.repository.utl.pt/handle/10400.5/20792> [Acesso em: 2022/05/19].

DGES (2021). Regime Geral: Ensino Superior Público Concurso Nacional de Acesso [Em linha]. Disponível em: <https://www.dges.gov.pt/pt/pagina/regime-geral-ensino-superior-publico-concurso-nacional-de-acesso> [Acesso em: 2022/09/19].

Duran-Silva, N., Fuster, E., Massucci, F. A. & Quinquillà, A. (2019). A controlled vocabulary defining the semantic perimeter of Sustainable Development Goals [Base de dados], dezembro 2019. Zenodo. Disponível em: <https://zenodo.org/record/3567769#.Xw3wEihKiUk>.

Few, S. (2006). Information Dashboard Design: The Effective Visual Communication of Data. O'Reilly Media, Inc.

Gomes, B. (2011). Previsão de Churn em Companhias de Seguros [Em linha]. Disponível em: <http://repositorium.sdum.uminho.pt/handle/1822/28301> [Acesso em: 2020/09/22]

Haupt, R., Scholtz, B. & Calitz, A. (2015). Using Business Intelligence to Support Strategic Sustainability Information Management. Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists, 1–11.

ISEG, L. S. of E. & M. (2021). Estratégia de Sustentabilidade do ISEG 2030 [Em linha]. Disponível em: <https://www.iseg.ulisboa.pt/wp-content/uploads/2021/06/Estrate%CC%81gia-de-Sustentabilidade-2030-PT.pdf> [Acesso em: 2022/09/19].

ISEG, L. S. of E. & M. (2021). Plano de Atividades para 2020 do Instituto Superior de economia [Em Linha]. Disponível em: [PLANO-DE-ACTIVIDADES-ISEG-2020-vfinal-dez-2019.pdf \(ulisboa.pt\)](#) [Acesso em: 2022/08/19].

Junior, R. M., Fien, J. & Horne, R. (2019). Implementing the UN SDGs in Universities: Challenges, Opportunities, and Lessons Learned. *Sustainability: The Journal of Records* 12 (2), pp 129–133.

Kimball, R. & Ross, M. (2013) *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, 3ªEd. Indiana Edition: Wiley. Kimball Group

Kruglyk, V. & Lvov, M. (2012). Choosing the first educational programming language. *CEUR Workshop Proceedings*, pp.188–198.

Malik, S. (2005) *Enterprise Dashboards - Design and Best Practices for IT*, John Wiley & Sons, Inc. John Wiley & Sons, Inc

Marcelino, A. C. L. (2020). Mapeamento das Atividades que contribuem para os ODS com o apoio do Business Intelligence - Aplicação Prática numa Instituição de Ensino Superior. *Lisbon School of Economics & Management - Universidade de Lisboa* [Em linha]. Disponível em: <https://www.iseg.ulisboa.pt/aquila/getFile.do?fileId=1503971&method=getFile> [Acesso em: 2022/03/13].

Marques, P. (2015). *Business Intelligence* [Em linha]. Disponível em: <https://run.unl.pt/bitstream/10362/37697/1/TGI0150.pdf> [Acesso em: 2022/09/22].

McKinney, W. (2013) *Python for Data Analysis*. O'Reilly Media, Incorporate

Mohammed, A. J. (2020). Decade of Action, Un Sdg [Em linha]. Disponível em: <https://www.un.org/sustainabledevelopment/decade-of-action/> [Acesso em: 2022/06/06].

Moody, D. & Kortink, M. A. (2000). From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design. *Proceedings of the International Workshop on Design and Management of Data Warehouses*, pp. 5–12.

Moura, A. (2017). Criação de um Data Warehouse Empresaria [Em linha]. [Acesso em: 2020/09/22].

Microsoft (2020). What is Power BI Desktop? [Em linha]. Disponível em: <https://docs.microsoft.com/pt-pt/power-bi/fundamentals/desktop-what-is-desktop> [Acesso em: 2022/08/15].

Niu, L., Lu, J. & Zhang, G. (2009). Business intelligence. In: Cognition-Driven Decision Support for Business Intelligence, Berlin: Springer, pp. 19–29.

NSS (2017). FlashText – A library faster than Regular Expressions for NLP tasks. Analytics Vidhya [Em linha]. Disponível em: <https://www.analyticsvidhya.com/> [Acesso em: 2022/09/20].

NumPy (2020). About Us. [Em linha]. Disponível em: <https://numpy.org/about/> [Acesso em: 2022/08/22].

Ong, I., Siew, P. & Wong, S. (2011). A Five-Layered Business Intelligence Architecture. Communications of the IBIMA 2011, pp. 1–11.

Peters, M. & Brath, R. (2004). Dashboard Design: Why Design is Important. DM Review, 16(3) pp. 5–8.

Portalete, M.C.R.B. (2021) Análise e Melhoria De Algoritmo de Detecção de Falsos Positivos na Manutenção de Keywords para os ODS. Lisbon School of Economics & Management - Universidade de Lisboa [Em linha]. Disponível em: <https://www.repository.utl.pt/bitstream/10400.5/23357/1/DM-MCRBSP-2021.pdf> [Acesso em: 2022/06/10].

Python (2020) *The Python Standard Library* [Em linha]. Disponível em: <https://docs.python.org/3/library/intro.html> [Acesso em: 2022/06/15].

NumPy (2020). About Us [Em linha]. Disponível em: <https://numpy.org/about/> [Acesso em: 2022/06/22].

Reddy, P. V. & Schroeder, C. G. (1999). Data Warehouse Architecture for Army Installations [Em linha]. Disponível em: https://archive.org/details/DTIC_ADA371882/page/n51/mode/2up [Acesso em: 2022/06/10].

Rad, R. (2019). From Rookie to Rock Star (Vol. I). RADACAD Systems Limited.

Regia-Douro Park (2022). SDGs Labs – Making the SDGs our Business [Em linha]. Disponível em: <https://www.regiadouro.com/erasmus/> [Acesso em: 2022/05/28].

República Portuguesa (2017). Relatório nacional sobre a implementação da Agenda 2030 para o Desenvolvimento Sustentável [Em linha]. Disponível em: <https://www.cig.gov.pt/wp-content/uploads/2017/07/Portugal2017.pdf> [Acesso em: 2022/03/12].

Rieckmann, M., Mindt, L. & Gardiner, S. (2017). Education for Sustainable Development Goals Learning Objectives. Paris: United Nations Educational, Scientific and Cultural Organization [Em linha]. Disponível em: https://books.google.pt/books?hl=ptPT&lr=&id=Fku8DgAAQBAJ&oi=fnd&pg=PP4&dq=sustainable+development+goals+and+education&ots=ZNGzjIb9j7&sig=aY4hdTe4VzBs0oLL2x39FBkSGKI&redir_esc=y#v=onepage&q=sustainable+development+goals+and+education&f=false [Acesso em: 2022/06/10].

SDGS- Labs, (2022) Making the SDGs our Business [Em linha]. Disponível em: <https://sdgs-labs.eu/> [Acesso em: 2022/04/28].

SDSN Australia/Pacific (2017). Getting started with the SDGS in Universities: A guide for universities, higher education institutions, and the academic sector. [Em linha]. Disponível em: <https://resources.unsdn.org/getting-started-with-the-sdgs-in-universities> [Acesso em: 2022/05/19].

Silva, A. (2021). Desenvolvimento de um dashboard de indicadores de gestão científica numa Instituição de Ensino Superior [Em linha]. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/134953/2/484403.pdf> [Acesso em: 2022/09/22].

Silva, L. (2016). ETL in the Big Data Era [Em linha]. Disponível em: <https://scholar.tecnico.ulisboa.pt/records/3TeLE25hp8omGbdi-KWaKkCrUq9ld6WqBB0A> [Acesso em: 2022/09/22].

Singh, V. (2017). Replace or Retrieve Keywords In Documents at Scale [Em linha]. Disponível em: <https://arxiv.org/abs/1711.00046> [Acesso em: 2022/05/22].

UNESCO (2012). Education for sustainable development: Sourcebook [Em linha]. [Acesso em: 2022/09/05].

Universidade de Évora (2020). UÉ, uma das primeiras IES a mapear ODS [Em linha]. Disponível em: <https://www.uevora.pt/ue-media/noticias?item=29485> [Acesso em: 2022/08/20]

Watson, H. J. (2009) Tutorial: Business Intelligence – Past, Present, and Future. *Communications of the Association for Information Systems*, 25 (39), 487–510.

Yegulalp, S. (2020) Python 2 EOL: How to survive the end of Python 2, InfoWorld. Disponível em: <https://www.infoworld.com/article/3365221/python-2-end-of-life-howto-survive-the-end-of-python-2.html> [Acesso em: 2022/06/10].

Yigitbasioglu, O. M. & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*.

Anexos

Anexo I:

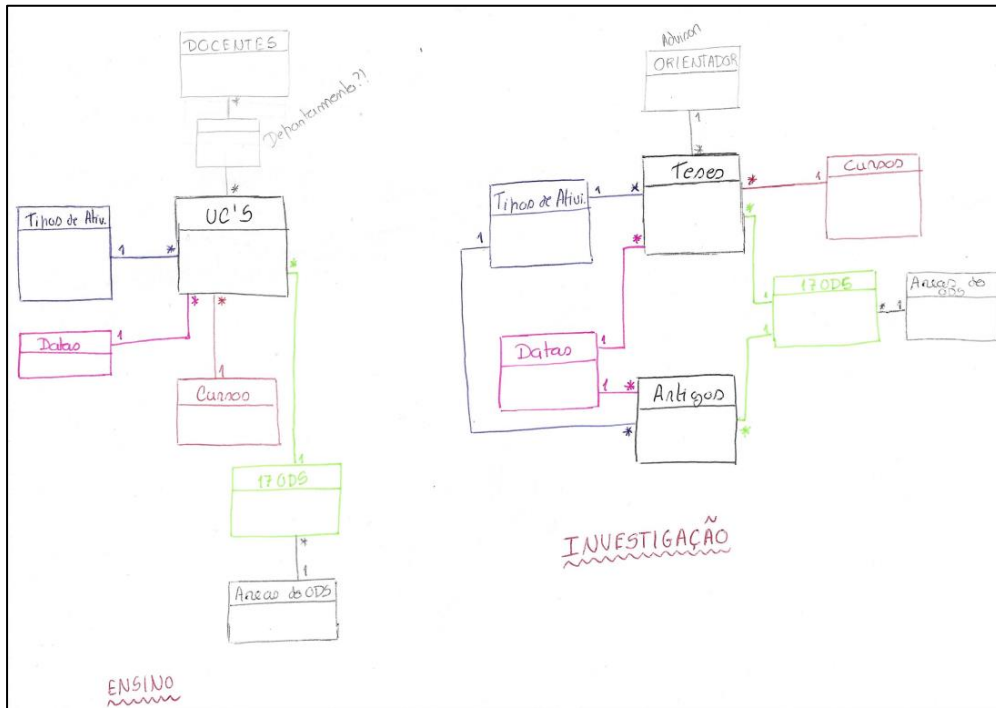


Figura 19 - Versão 1 do Modelo dimensional

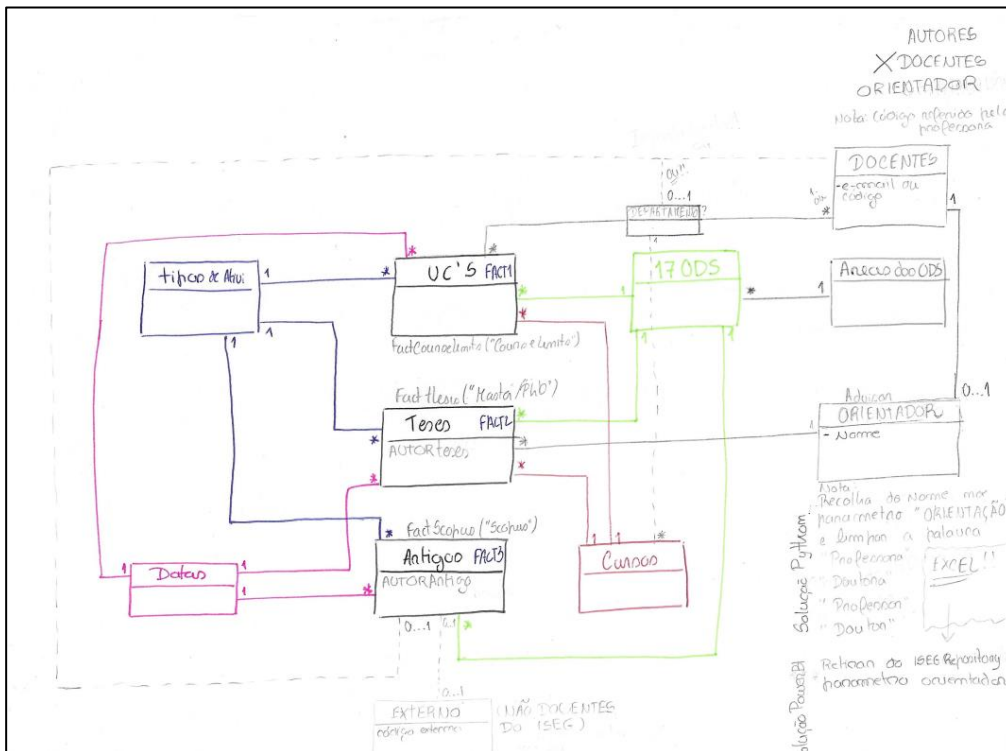


Figura 20 - Versão 2 do Modelo dimensional

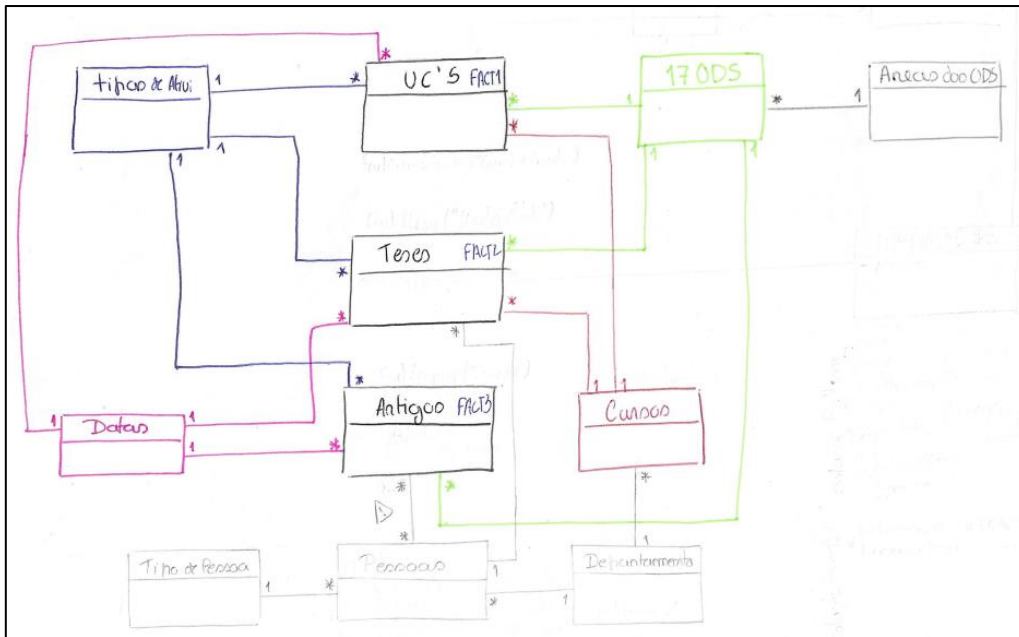


Figura 21 - Versão 3 do Modelo dimensional

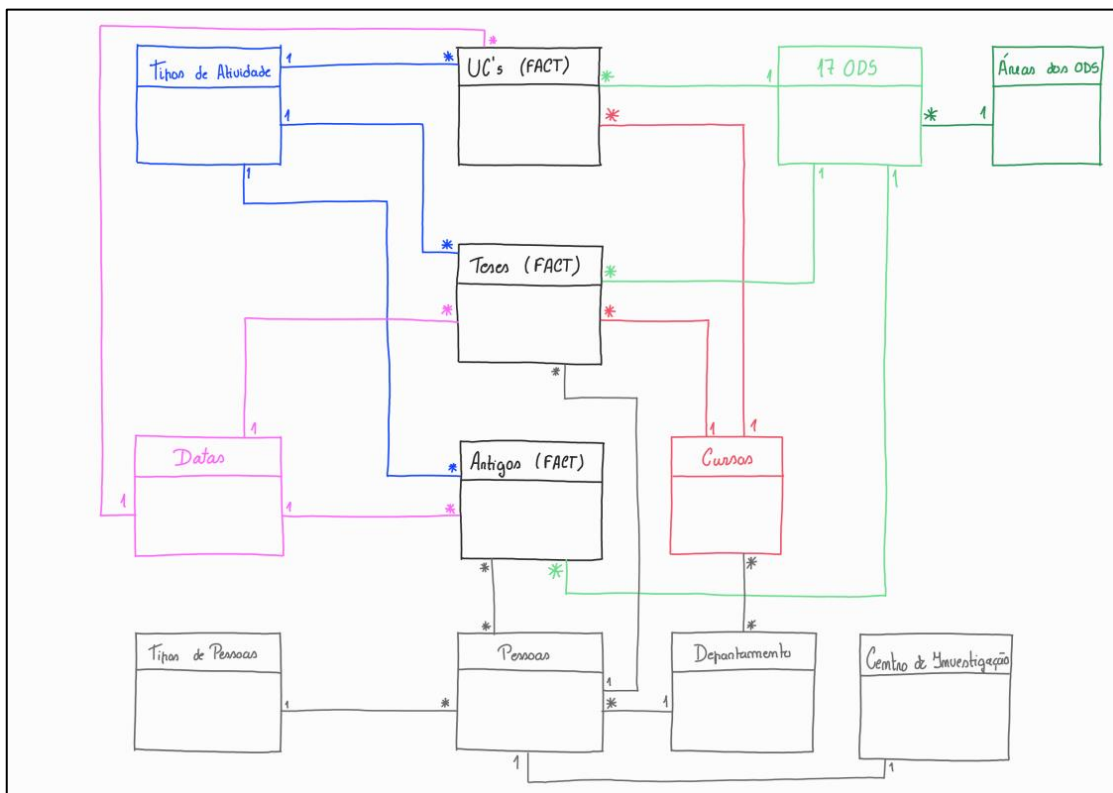


Figura 22 - Versão 4 do Modelo dimensional