

Mestrado em Engenharia Informática

Dissertação

Relatório Final

Projeto DW-UC

Desenvolvimento de uma *Data Warehouse*
para a Universidade de Coimbra

Adolfo Joaquim Santos Fonseca Pinto

ajpinto@student.dei.uc.pt

Orientador:

Bruno Miguel Brás Cabral

bcabral@dei.uc.pt

Coimbra, 5 de Setembro de 2019



FCTUC DEPARTAMENTO
DE ENGENHARIA INFORMÁTICA
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Resumo

Dada a heterogeneidade e quantidade de dados existentes na Universidade de Coimbra, tornou-se necessário desenvolver um sistema que desse apoio na tomada de decisões.

Este sistema, inserido no projeto UC-Num, pretende disponibilizar uma análise OLAP (*Online Analytical Processing*) com base numa *Data Warehouse*. Através de uma interface web, e com auxílio a dashboards interativas com análises de KPIs (*Key Performance Indicators*), este sistema pretende auxiliar na tomada de decisão em áreas específicas da universidade.

O objetivo deste estágio passa por desenvolver uma componente da *Data Warehouse*, em conjunto com um sistema para realizar análises para a Universidade de Coimbra.

Estas análises dão informação essencial à administração e gestão da universidade. Aqui, os *stakeholders* assumem um papel de relevo tanto na concepção do indicador como na validação final do mesmo. Tais indicadores, inseridos no plano estratégico e de ação da universidade, foram alvo de uma definição inicial de maneira a perceber como se poderia fazer a recolha de dados que dessem suporte ao mesmo. Através destes indicadores e de métricas específicas é possível saber o comportamento da instituição e ainda se esta atinge os objetivos propostos.

Ao estarem disponíveis na plataforma, estes indicadores resultam de um trabalho de recolha, modelação de dados, automatização do processo de recolha de dados e disponibilização numa página *web*.

Sendo este estágio um projeto de BI (*Business Intelligence*), este pretende seguir o desenvolvimento típico de um projeto deste tipo de acordo com a metodologia recomendada por *Ralph Kimball*, um dos pioneiros no conceito de *Data Warehousing*.

Palavras-Chave

Business Intelligence, Data Warehouse, Online Analytical Processing, ETL, KPIs

Índice

1. Introdução	8
1.1. Enquadramento	8
1.2. Objetivo	11
1.3. Estrutura do relatório	12
2. Plataforma de análise de dados	13
2.1. Objetivos	13
2.2. Principais funcionalidades	17
2.3. Requisitos funcionais	18
2.4. Requisitos não funcionais	20
2.5. Áreas e indicadores já disponíveis	21
2.6. Arquitetura de Software	22
2.6.1. Tecnologias	24
3. Análise de requisitos	25
4. Implementação	27
5. Plano de trabalhos	31
5.1. Primeiro semestre	31
5.2. Segundo semestre	32
5.3. Resultados	33
6. Conclusão	34
Referências	35
Anexos	36

Lista de Figuras

Figura 1 - Enquadramento do projeto UC-num	9
Figura 2 - Dados económico-financeiros	10
Figura 3 - Disponibilização dos indicadores	21
Figura 4 - Arquitetura do sistema	22
Figura 5 - Modelo de dados do indicador de empreendedorismo e inovação	27
Figura 6 - Extração dos dados dos alunos do NONIO	28
Figura 7 - Carregamento da tabela de factos	28
Figura 8 - <i>Schema Workbench</i> , criação de cubos	29
Figura 9 - Modelo de dados do indicador de comunicação	31
Figura 10 – <i>Dashboard</i> do indicador de Comunicação	32
Figura 11 - Planeamento de um projeto de BI de acordo com Ralph Kimball	31
Figura 12 - Plano para o primeiro semestre	32
Figura 13 - Plano para o segundo semestre	33
Figura 14 - Indicador do módulo Académicos	43
Figura 15 - Indicador do módulo Comunicação	44
Figura 16 - Indicador do módulo Comunicação	47

Lista de Tabelas

Tabela 1 - Requisitos funcionais com prioridade alta	17
Tabela 2 - Requisitos funcionais com prioridade média	18
Tabela 3 - Requisitos funcionais com prioridade baixa	18
Tabela 4 - Requisitos não funcionais	19
Tabela 5 – Indicadores do ambiente de testes para validação	38
Tabela 6 – Indicador de gestão orçamental	25
Tabela 7 – Indicador de académicos	25
Tabela 8 – Indicador de Investigação	26
Tabela 9 – Indicador de Económico-Financeiros	26
Tabela 10 – Indicador de Académicos	26
Tabela 11 – Indicador de Comunicação	26
Tabela 12 – Indicador de Comunicação	26

Glossário

BI	<i>Business Intelligence</i>
ETL	<i>Extract Transform Load</i>
ELT	<i>Extract Load Transform</i>
KPI	<i>Key Performance Indicator</i>
OLAP	<i>Online Analytical Processing</i>
UC	Universidade de Coimbra
SAP	<i>Systems, Applications & Products</i>
NÓNIO	Sistema de Gestão Académica
DW	<i>Data Warehouse</i>
FCT	Faculdade de Ciências e Tecnologias
CSV	<i>Comma-separated values</i>
RDBMS	<i>Relational Database Management System</i>
MDX	<i>Multidimensional Expressions</i>
API	<i>Application Programming Interface</i>
CDE	<i>Community Dashboard Editor</i>
CDA	<i>Community Data Access</i>

Capítulo 1

Introdução

Este capítulo introduz os objetivos principais deste estágio. Começa-se por fazer um enquadramento do trabalho no projeto UC-Num, seguida da motivação para o mesmo. Após isso são apresentados os objetivos do estágio e na última secção é apresentada a restante estrutura do relatório.

1.1 Enquadramento

A Universidade de Coimbra é composta por diversas faculdades que se focam em diferentes áreas do saber. Apesar de cada faculdade ter a sua própria autonomia e projetos (departamentos, institutos, centros, associações, etc.), ao longo do tempo torna-se difícil e moroso o processo de gestão e análise de dados.

A UC, de forma a tomar decisões com base nestas instituições autónomas, precisa de uma ferramenta que auxilie na tomada de decisões. Esta ferramenta insere-se num projeto que resultou numa candidatura da UC ao Sistema de Apoios à Modernização Administrativa (SAMA) [1]. Este projeto, denominado “UC-Cloud – Infraestruturas e serviços de TIC para a *cloud*” tem como um dos principais objetivos agilizar o processo de resposta a todos os utilizadores da UC. Este engloba também outros subprojectos de grande interesse á comunidade académica tais como:

- **UC-com** - Infraestruturas de suportes a serviços
- **UC-gest** - Integração e extensão de sistemas de gestão
- **UC-pay** - Sistema de pagamentos eletrónicos
- **UC-contract** - Integração com plataformas de contratação pública
- **UC-numTabBoard** - Indicadores para a gestão e monitorização da estratégia

Toda esta informação pode ser condensada na figura abaixo:

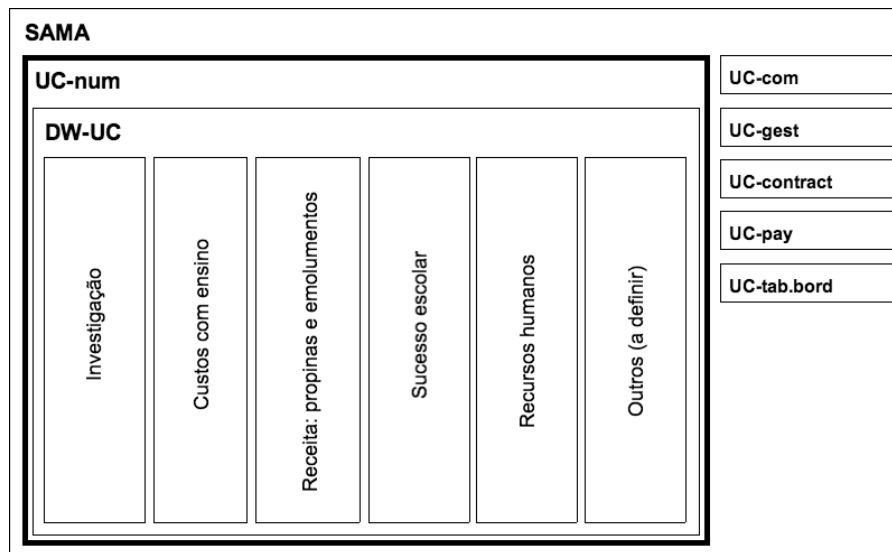


Figura 1 - Enquadramento do projeto UC-num

Idealmente, todo este processo de recolha de dados seria automático, mas isso não se verifica para todos os indicadores. De facto, todos os anos são utilizados diversos recursos humanos para recolher informação manualmente de uma dada área da universidade, como por exemplo, o Financiamento de Projetos "per capita" ou a Percentagem de Docentes em Unidades de I&D reconhecidas pela FCT.

Estes indicadores de desempenho (compostos por um valor possível de calcular), presentes no plano estratégico e de ação da UC 2011-2015 e mais recentemente, o planeamento estratégico 2015-2019, ajudam na tomada de decisão. Analisando tais indicadores, é possível perceber se uma organização está a atingir os objetivos definidos no seu plano. Para se ter uma ideia de que tipo de indicadores se está a considerar, tome-se como exemplo na área de **Ensino** os indicadores:

- Taxa de sucesso escolar
- Grau de empregabilidade
- N° de estudantes de doutoramento

ou a área de **Investigação**:

- N° de teses de doutoramento concluídas
- N° de projetos europeus e internacionais

Percebe-se, portanto, que estes indicadores, tanto de alto nível (comportamento global da organização) e baixo nível (processos específicos a cada instituição) são fundamentais para uma boa gestão da universidade. O projeto UC-Num, em que este estágio se insere, foi criado com base nas ideias expressas anteriormente. Composta por uma *Data Warehouse*, que por sua vez tem várias *data marts* (subconjunto de dados de uma DW), com um sistema OLAP, o acesso a toda esta informação pretende ser agilizado o que torna mais fácil o processo de tomada de decisões.

No arranque do estágio esta *Data Warehouse* já se encontra em funcionamento, trabalho que foi desenvolvido por alunos anteriores/bolseiros e orientados pelo prof. Dr. Bruno Cabral.

Esta plataforma já recolhe indicadores das diversas áreas do plano estratégico e de ação, através de um sistema ETL e que disponibiliza a informação via web em *dashboards* interativas. A figura 2 mostra como um destes indicadores da área Económico-Financeira se encontra na plataforma atualmente.

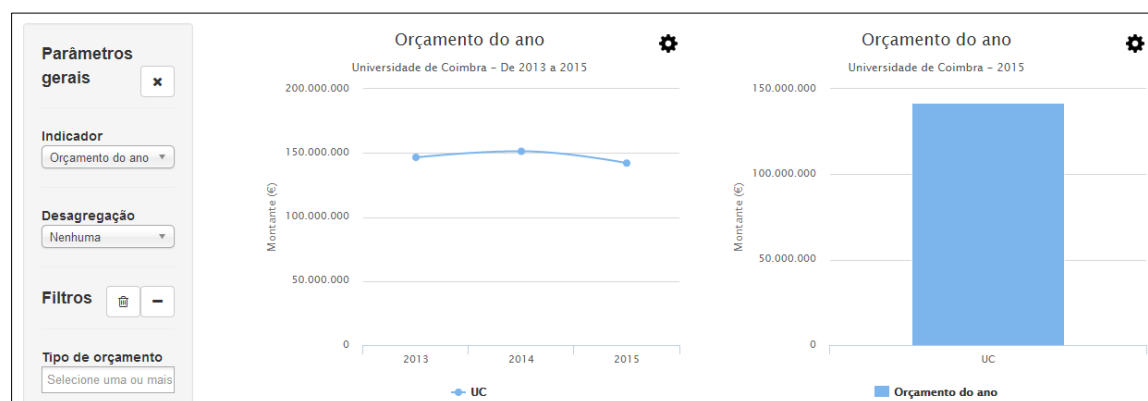


Figura 2 - Dados económico-financeiros

1.2 Objetivo

O objetivo deste estágio é desenvolver uma nova componente da solução de *data warehousing* e análise OLAP para a UC. O trabalho focar-se-á em áreas identificadas como sendo importantes dado o estado atual da plataforma:

- Projetos de investigação
- Investigação
- Dados académicos
- Comunicação

Mais concretamente:

- Definir em detalhe os novos indicadores a implementar;
- Definir o plano e a arquitetura do ETL para aquisição de dados para os novos indicadores;
- Implementação do ETL e do *front end* OLAP;
- Validação do indicadores e colocação dos mesmos em produção;
- Manutenção e correções à plataforma UC-Num;

1.3 Estrutura do relatório

No capítulo 2 é apresentada a plataforma de análise de dados, detalhando quais os objetivos, principais funcionalidades, áreas e KPIs já disponíveis e a arquitetura de software da mesma.

No capítulo 3 são apresentados o levantamento de requisitos.

No capítulo 4 é apresentado a implementação de indicadores (modelo de dados, ETL, cubos OLAP e *dashboards*).

No capítulo 5 é apresentado o plano de trabalhos para o primeiro e segundo semestre e seus resultados.

No capítulo 6 é apresentada a conclusão e após isso as referências utilizadas.

A última secção é dedicada aos anexos.

Capítulo 2

Plataforma de análise de dados

Neste capítulo pretende-se apresentar os objetivos desta plataforma, que se encontra em produção atualmente, nomeadamente as suas principais funcionalidades e as áreas e indicadores do PES que se encontram disponíveis. Por fim é apresentada a arquitetura de software do sistema, um dos pontos mais importantes de qualquer *software* de elevada dimensão.

2.1 – Objetivos

Os principais objetivos desta plataforma, são de disponibilizar indicadores para a avaliação da qualidade pedagógica e desempenho dos serviços da UC.

Esta aplicação também disponibiliza uma análise OLAP com base numa *Data Warehouse*, cujo processo de carregamento é automático. Esta análise é apresentada através de uma página *web*, que apresenta um conjunto de ecrãs interativos com análises de *KPIs* e outros indicadores de relevância a membros específicos da comunidade da UC (equipa reitoral, diretores de faculdades e departamentos e conselho de gestão). Torna-se, assim, bastante útil no apoio à tomada de decisão em diversas áreas. [2]

Desta maneira é possível diminuir o tempo de espera e garantir o processo de monitorização estratégica da UC e do contributo individual de cada instituição da UC, através dos indicadores chave (*KPIs*) [2].

Estes indicadores, fazem parte de um conjunto de métricas identificadas pela UC, que se inserem nas seguintes áreas:

- **Investigação** – Nesta área inserem-se indicadores de grande importância á UC como o “Nº. de teses de doutoramento concluídas”, “Nº. de projetos europeus e internacionais”, “Percentagem de docentes em unidades de I&D reconhecidas pela FCT” entre outros. Esta informação vem maioritariamente da *Web of Science* [3], que

gere uma grande quantidade da literatura científica e acadêmica publicada internacionalmente.

- **Ensino** – Nesta área encontram-se indicadores que descrevem toda a parte acadêmica da UC: unidades curriculares, pautas, empregabilidade, candidaturas à universidade, estágios, etc. Atualmente a UC já tem um sistema que trata desta informação que suporta todos os processos de gestão acadêmica, o NÓNIO, sendo que os dados destes indicadores vêm desta fonte de dados (automática).
- **Comunidade** – Esta área engloba o turismo (nº de visitas a espaços turísticos), biblioteca (nº de títulos disponibilizados, nº de empréstimos) e ainda a rede estudantil (nº de antigos estudantes ativos na rede UC). Outros indicadores mais gerais como o “Número de candidatos à Universidade de Verão” ou “Grau de bilinguismo do site UC” também são considerados importantes.
- **Pessoas** – Nesta área é possível ter uma visão geral de toda a UC: rácio de trabalhadores, demografia, índices etários, serviço docente, etc. Como este pilar engloba grandes quantidades de dados da UC e um processo de recolha manual dependeria de muitos recursos humanos, esta informação é disponibilizada através de *web services* SAP. Ou seja, periodicamente, esta informação é recolhida utilizando serviços já desenvolvidos anteriormente pela UC de maneira a agilizar todo este processo.
- **Económico-Financeiros** – Aqui pode-se ver dados como a evolução da despesa, orçamento do ano, a receita arrecadada e dívidas de clientes, entre outros. Esta informação também é disponibilizada através de SAP.
- **Infraestruturas** – Os indicadores presentes neste pilar não têm um elevado grau de importância, ao contrário dos indicadores referidos anteriormente. “Custo por m² com conservação, manutenção e requalificação dos edifícios”, “Grau de cobertura por rede wireless” são indicadores possíveis de encontrar aqui. Note-se aqui que, dada a baixa importância desta área, esta informação é recolhida de forma manual.

- **Internacionalização** – Indicadores relativos á parte internacional da universidade como “Percentagem de estudantes de nacionalidade estrangeira”, “Nº. de publicações em colaboração internacional”, “Percentagem de estudantes de mobilidade” constituem este pilar. Esta área contém dados que podem vir de diversas fontes: NÓNIO, SAP ou *Web Of Science*.
- **Cidadania e inclusão** - A inclusão social e o exercício da cidadania são temas de foco a UC. Aqui podemos encontrar informação sobre bolseiros alojados em residências ou nº. de estudantes dirigentes associativos jovens que participam em ações de voluntariado entre outros indicadores a desenvolver no futuro.
- **Marca UC** – Nesta área interessa saber informação sobre produtos da universidade como “Receita da venda de produtos com marca UC”, “Nº. de prémios obtidos pela UC ou a membros da comunidade académica” entre outros.
- **Comunicação** - De maneira a ter uma presença na web e dar a conhecer a pessoas estrangeiras a Universidade de Coimbra, indicadores relativos a redes sociais, (*Facebook, Twitter, Instagram*) são contabilizados aqui. Através de informação vinda do *Google Analytics* é possível saber também o grau de relacionamento dos utilizadores com as páginas web da universidade (nacionalidade, demografia, tempo de permanência numa página, etc.)
- **Ambiente** – “Consumo de eletricidade, gás ou água por m² utilizado”, “Posição no *GreenMetric Ranking of World Universities*”, constituem indicadores presentes no plano estratégico da UC e que são alvo de monitorização pela UC.

Torna-se claro que, antes de estes indicadores serem disponibilizados numa pagina *web*, é necessário haver um levantamento e análise dos mesmos, processo comum a muitos projetos de *software*, onde numa primeira fase se reúnem todos os requisitos que se pretendem implementar e que são alvo de uma análise posterior. Após isso é feita a especificação formal destes indicadores que geram um documento que contém toda a informação relativa ao identificador a implementar.

Dada a grande dimensão da UC, estes indicadores foram reunidos por diferentes áreas de ação (ensino, pessoas, económico-financeiros, etc) com a presença dos elementos responsáveis pela área em análise. Depois deste processo, que, regra geral, ocupa uma grande parte do tempo do desenvolvimento previsto, o *output* deste trabalho passa pela definição e especificação dos indicadores. Após este trabalho, é feita a recolha da informação nos sistemas fonte. Sendo que alguns indicadores se focam em áreas específicas da UC, foram extraídos, tratados, integrados e armazenados os dados na *data warehouse* dos sistemas SAP e NÓNIO, sistemas estes que são usados nos serviços da administração da UC.

Depois destes dados estarem presentes na *data warehouse*, foram criados ecrãs (*dashboards*) com análises sobre os indicadores. Muitos destes indicadores são possíveis de parametrizar e ainda filtrar os resultados de acordo com uma métrica específica, como por exemplo, o ano letivo.

2.2 – Principais funcionalidades

Desde o início deste projeto, que a plataforma tem vindo a evoluir e a englobar dados das mais diversas áreas. As funcionalidades (comuns a todos os módulos), também chamadas de requisitos funcionais, já se encontravam especificadas e implementadas exceto as funcionalidades relativas aos indicadores que serão implementados este ano.

Dado este projeto ter sido começado em anos anteriores, muitas destas funcionalidades foram desenvolvidas por alunos/bolseiros que se encontravam a fazer estágio neste projeto. Estas funcionalidades servem de base a quase todos os indicadores e encontram-se atualmente em produção.

2.3 – Requisitos Funcionais

Segue-se de seguida uma análise não exaustiva das principais funcionalidades do sistema, ou seja, os requisitos funcionais:

Funcionalidade 1: Autenticação
Descrição: Sendo esta plataforma acessível a um grupo restrito de utilizadores e composta por dados de toda a UC, esta deve apresentar ao utilizador um sistema de <i>login</i> (utilizador/palavra-passe). As permissões de acesso terão de ser dadas pelas entidades responsáveis ou pelo administrador de sistema.
Funcionalidade 2: Terminar a sessão
Descrição: Na pré-condição do utilizador estar autenticado, este consegue sair da aplicação de forma segura sendo reencaminhado para a página de autenticação.
Funcionalidade 3: Terminar a sessão por inatividade
Descrição: Semelhante à funcionalidade anterior, se não houver interação com a plataforma durante um determinado período de tempo, o utilizador terá de inserir novamente os seus dados de autenticação.
Funcionalidade 4: Autorização
Descrição: A plataforma disponibiliza vários níveis de acesso aos dados, isto é, um utilizador pode, por exemplo, ter acesso apenas aos dados académicos ou económico-financeiros (entre outros), dependendo das permissões dados pelo administrador.
Funcionalidade 5: Navegação interna
Descrição: Neste sistema, composto por uma DW, existem os conceitos de “ <i>drill down</i> ” e “ <i>drill up</i> ”, que significa navegar através das dimensões hierárquicas de dados. Assim, é possível fazer “ <i>drill down</i> ” dentro de OLAP, ou seja, ver os dados que se encontram em hierarquias mais baixas e “ <i>drill up</i> ”, ou seja, ver os dados em hierarquias mais altas (<i>roll up</i>).
Funcionalidade 6: Parâmetros gerais
Descrição: A plataforma disponibiliza diversos parâmetros possíveis de aplicar aos dados apresentados. Aqui é possível aplicar filtros dependentes dos valores apresentados (e.g. género, nacionalidade, faixa etária, etc) e ainda filtros temporais.

Funcionalidade 7: Parâmetros de tempo
Descrição: Grande parte dos dados ocorrem durante um período de tempo, pelo que, é possível especificar que ano/anos se pretende visualizar dentro dos disponíveis.
Funcionalidade 8: Secção de ajuda
Descrição: Para utilizadores não familiares com a plataforma esta disponibiliza uma secção de ajuda com os casos de uso mais frequentes.
Funcionalidade 9: Informação auxiliar
Descrição: Em caso de dúvida relativamente aos dados apresentados, são fornecidas ferramentas que ajudam o utilizador a perceber os dados que está a visualizar, como por exemplo, <i>tooltips</i> com informação complementar.
Funcionalidade 10: Visualização de dados
Descrição: Mostrar informação na forma de gráfico-tabela é uma das maneiras mais eficazes de compreender os dados apresentados, funcionalidade que, esta plataforma também disponibiliza.

Tabela 1 – Requisitos funcionais com prioridade alta

A tabela 1 apresenta os requisitos que têm prioridade alta no sistema. Existem também requisitos com mais baixo nível de importância como se mostra nas tabelas 2 e 3:

Funcionalidade 1: Navegação entre módulos
Descrição: A plataforma é composta por vários módulos, sendo que o utilizador deve poder aceder a estes dados a partir de qualquer instante.

Tabela 2 - Requisitos funcionais com prioridade média

Funcionalidade 1: Ocultar parâmetros
Descrição: De maneira a otimizar o espaço no ecrã, o utilizador pode esconder o <i>container</i> onde se encontram os parâmetros gerais e de tempo.
Funcionalidade 2: Exportar informação da tabela/gráfico
Descrição: No caso de se querer utilizar a informação apresentada para outros fins, é possível exportar para formato <i>excel</i> ou CSV a informação presente nas tabelas apresentadas ao utilizador.
Funcionalidade 3: Zoom de gráficos
Descrição: Efectuar <i>zoom in</i> e <i>out</i> de um dos gráficos

Tabela 3 – Requisitos funcionais com prioridade baixa

2.4 – Requisitos não funcionais

Em termos de qualidade que a plataforma deve possuir, existem funcionalidades bem definidas e que devem ser comuns a todos módulos implementados anteriormente e no futuro. Tais funcionalidades, também chamadas de requisitos não funcionais são apresentados na tabela 4:

Funcionalidade 1: Atualização de dados
Descrição: O <i>backend</i> (sistema que implementa as regras de negócio) da DW é suportado por um processo de ETL e vários cubos OLAP, pelo que este processo deve ser autónomo.
Funcionalidade 2: Compatibilidade de navegadores
Descrição: Dado existir uma miríade de browsers disponíveis no mercado, esta plataforma deve dar suporte aos browsers com maior taxa de utilização (Google <i>Chrome</i> , Firefox, Opera e Safari).
Funcionalidade 3: Compatibilidade de sistemas operativos
Descrição: Atualmente o sistema encontra-se desenvolvido em Linux, pelo que no futuro deve ser dado suporte a este SO.
Funcionalidade 4: Licenças de <i>software</i>
Descrição: Sempre que possível, todo o <i>software</i> utilizado deve ser <i>freeware</i> .
Funcionalidade 5: Segurança
Descrição: Para aumentar a segurança dos dados deve ser utilizado HTTPS com um certificado SSL para encriptar e proteger a informação entre o servidor e o utilizador final.

Tabela 4 – Requisitos não funcionais

2.5 – Áreas e indicadores chave já disponíveis

Atualmente, das áreas do plano estratégico e ação identificadas anteriormente, as seguintes áreas já tem dados disponíveis para visualização na plataforma:

- Ensino
- Recursos-Humanos
- Económico-Financeiros/Gestão orçamental
- Ação social

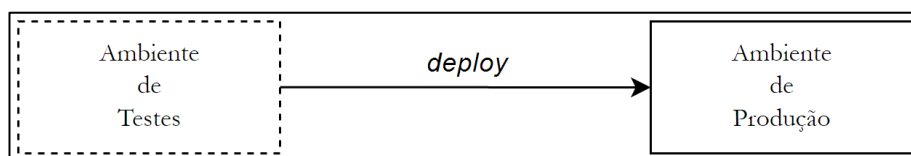


Figura 3 - Disponibilização dos indicadores

Dada a grande quantidade de KPI's a implementar e o processo de validação dos mesmos ser um processo demorado, o número de indicadores disponível em ambiente de testes é superior ao número de indicadores no ambiente de produção. De facto, estes só são colocados em produção depois de haver uma validação pelas entidades competentes, que regra geral, são pessoas da parte administrativa da UC.

Quando a Data Warehouse foi desenhada criou-se um ambiente de desenvolvimento local, de testes e de produção. Neste ambiente de desenvolvimento local encontrava-se todo o trabalho desenvolvido por alunos, onde cada um trabalhava num módulo específico da DW.

Estas alterações eram enviadas para o ambiente de testes e após uma validação, estes seriam colocados em produção. Ainda assim, indicadores que sejam colocados em testes podem deixar de funcionar correctamente ao fim de certo tempo. Isto deve-se a erros no processo ETL ou algum indicador que dependa de um *web service* que deixou de funcionar, por exemplo. Nesta fase, teoricamente, estes indicadores deveriam ser enviados novamente para o ambiente de desenvolvimento (*roll back*) para corrigir eventuais erros críticos.

2.6 – Arquitetura de Software

A definição da arquitetura de software é um processo fundamental a qualquer sistema robusto. A construção deste sistema de suporte à decisão, pretende seguir um processo típico de um sistema de inteligência no negócio.

A figura 4 apresenta a arquitetura atual do sistema:

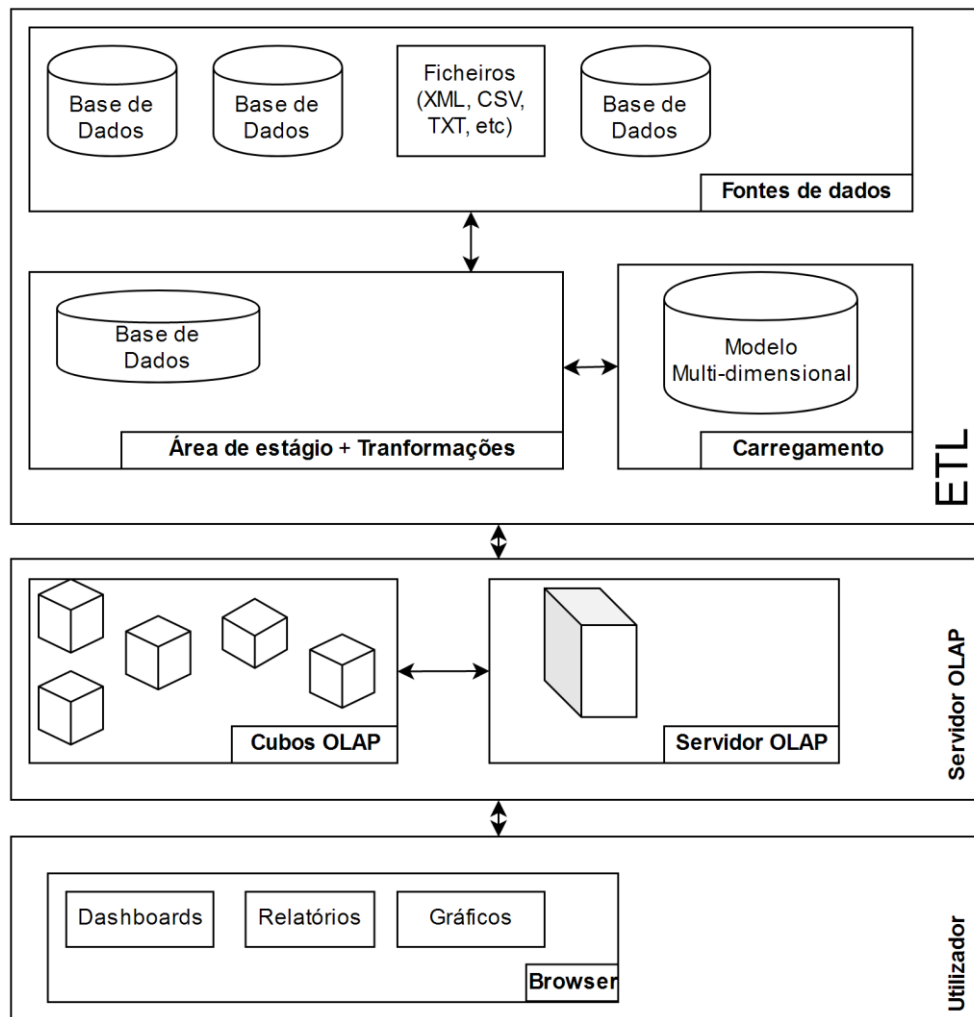


Figura 4 - Arquitetura do sistema

Numa primeira fase, é necessário extrair os dados que se encontram geralmente em bases de dados relacionais, mas podendo também estar noutro formato como ficheiros de texto, por exemplo. Depois de extraídos estes dados, estes são enviados para a área de estágio. Esta área de estágio é composta por uma base de dados que irá armazenar a informação proveniente das fontes de dados originais nas respectivas tabelas criadas para o efeito. Nesta etapa a primeira fase do processo **ETL**, a extração, é concluída. A segunda fase,

transformação (ETL), passa por “limpar” os dados. Dependendo dos dados, é comum lidar com valores nulos, eliminação de duplicados e conversão de dados. Ou seja, estes dados vão sofrer um conjunto de transformações, processo que é fundamental para evitar problemas futuros.

Se não houver problemas com o processo anterior, estes são carregados para o modelo multi-dimensional. Já na área do servidor OLAP, esta é composta por vários cubos que permitem representar os dados em vários níveis. Isto ajuda no momento das pesquisas pela forma que os dados são representados. Finalmente, os utilizadores podem aceder aos dados através de *dashboards* interativas (gráficos, tabelas) utilizando um *browser*.

Todas estas áreas estão ligadas entre si, representadas pela seta bidirecional da figura 4, o que significa que se algum dos processos falhar o indicador não pode ser mostrado ao utilizador final apresentando para isso uma mensagem de erro (*feedback*). Sendo que existe um sistema de *logs* para cada operação realizada na *Data Warehouse*, a pessoa que desenvolveu determinada funcionalidade pode consultar os ficheiros de logs que registam eventuais erros que ocorram durante a análise OLAP.

2.6.1 – Tecnologias

Nesta secção pretende-se apresentar informação complementar à figura 4, arquitetura do sistema.

Idealmente, a escolha de base de dados da *data warehouse* seria a mesma da fonte de dados para evitar lidar com diferentes implementações de dados. Rapidamente se percebe que, as fontes de dados são muito heterógeneas, sendo suportadas por OracleSQL, Access, MySQL, Excel, entre outras. Resta, portanto, utilizar um motor RDBMS robusto e flexível que se adequa às necessidades da DW. Foi escolhido utilizar **PostgreSQL**, tanto por ser uma alternativa *open-source* com desenvolvimento ativo, como o elevado leque de funcionalidades como por exemplo, vistas materializadas ou o uso de índices. Estas duas últimas capacidades são de extrema importância numa *data warehouse* em termos de performance. Apesar disto, outros motores de bases de dados como Oracle seriam também possíveis de utilizar. Como foi referido anteriormente, toda esta escolha de tecnologias já se encontrava bem definida e que foi alvo de uma análise (vantagens / desvantagens).

Na área do **ETL** o software que tem vindo a ser utilizado pelos alunos anteriores ao projeto, é o *Kettle (Pentaho Data Integration)*, ferramenta muito utilizada para projetos de *business intelligence* e que se adequa ao nosso caso de uso. Apresenta uma interface gráfica de criação de processos de ETL, uma grande quantidade de plugins disponíveis e a grande capacidade de orquestração em coordenar e combinar transformações.

Para a criação de **dashboards** foi utilizado o *pentaho BI Server*, com auxílio de vários *plugins* pela facilidade de criação de *dashboards* em formato web. Neste passo, o utilizador apenas terá de aceder através do *browser* ao *website* onde se encontra o sistema, para consultar os vários gráficos e tabelas disponíveis.

Em suma, todo o trabalho a ser desenvolvido deve recair sobre o uso deste *software*.

Capítulo 3 – Análise de requisitos

Uma das etapas essenciais num projeto de software é a recolha de requisitos. Nesta fase, foi feita uma reunião onde se fez uma seleção de indicadores a implementar.

Numa primeira fase, de entre diversos indicadores, atribui-se a cada indicador um grau de importância: alta prioridade, que como o nome indica, seriam os indicadores que deveriam ser implementados o mais rápido possível e baixa prioridade, ou seja, os indicadores que nesta fase, não eram de grande importância, mas que poderiam vir a ser no futuro. Esta lista de indicadores encontra-se disponível no Anexo B.

Outro aspeto importante foi a seleção de indicadores em que o processo de extração fosse automático, ou seja, não seria necessário a disponibilização de recursos humanos por parte da universidade para auxiliar no desenvolvimento dos mesmos.

Desta lista de indicadores foi feita uma subseleção, visto ser impossível de implementar todos os indicadores importantes, que resultou nas tabelas abaixo, em que é especificado os indicadores de prioridade elevada de recolha automática:

Indicador	Receitas com visitas aos espaços turísticos
Módulo	GO (Gestão Orçamental)
Plano Estratégico	Sim
Fonte de informação	SAP
Estado	Automático
Observações	Verificar se o <i>web service</i> da receita arrecadada responde a este indicador, por espaço visitado

Tabela 6 – Indicador de Gestão Orçamental

Indicador	N.º de estudantes a frequentar unidades curriculares na área de empreendedorismo e inovação (1.º, 2.º e 3.º ciclo)
Módulo	INV (Investigação)
Plano Estratégico	Sim
Fonte de informação	NÓNIO
Estado	Automático

Tabela 7 – Indicador de Académicos

Indicador	Volume de financiamento obtido no programa Portugal 2020 e no Horizonte 2020
Módulo	PR (Projetos (Investigação))
Plano Estratégico	Sim
Fonte de informação	SAP
Estado	Automático

Tabela 8 – Indicador de Investigação

Indicador	Peso do financiamento de origem internacional, por tipologia
Módulo	GO (Gestão Orçamental)
Plano Estratégico	Sim
Fonte de informação	SAP
Estado	Automático
Observações	De que forma é que esta informação pode ser recolhida de SAP?

Tabela 9 – Indicador de Económico-Financeiros

Indicador	Nº de estudantes dirigentes associativos jovens, estudantes membros de órgãos da UC, estudantes com participação em atividades de reconhecido mérito universitário e estudantes em ações de voluntariado credenciadas
Módulo	AC (Académicos)
Plano Estratégico	Sim
Fonte de informação	NÓNIO
Estado	Automático
Observações	É possível obter através da situação especial do estudante. A aguardar envio da lista de situações especiais.

Tabela 10 – Indicador de Académicos

Indicador	Número de visitas (sessões) a páginas (do domínio uc.pt)
Módulo	COM (Comunicação e marca UC)
Plano Estratégico	Sim
Fonte de informação	<i>Google Analytics</i>
Estado	Filtros: por origem geográfica, por tipo de dispositivo de acesso, por idioma do browser do visitante, por origem do tráfego.

Tabela 11 – Indicador de Comunicação

Indicador	Número de seguidores nas redes sociais
Módulo	COM (Comunicação e marca UC)
Plano Estratégico	Sim
Fonte de informação	<i>Facebook/Twitter/Instagram/Youtube</i>
Estado	Automático

Tabela 12 – Indicador de Comunicação

Capítulo 4 - Implementação

No caso do indicador da tabela 7, n.º de estudantes a frequentar unidades curriculares na área de empreendedorismo e inovação, foi elaborado o modelo de dados presente na figura 5. Este modelo tem como tabela de factos *ac_f_inovacao* que é constituída por chaves estrangeiras para as dimensões que suportam os dados deste indicador. Neste caso as dimensões são: unidades curriculares, tempo, curso, demografia e unidade orgânica.

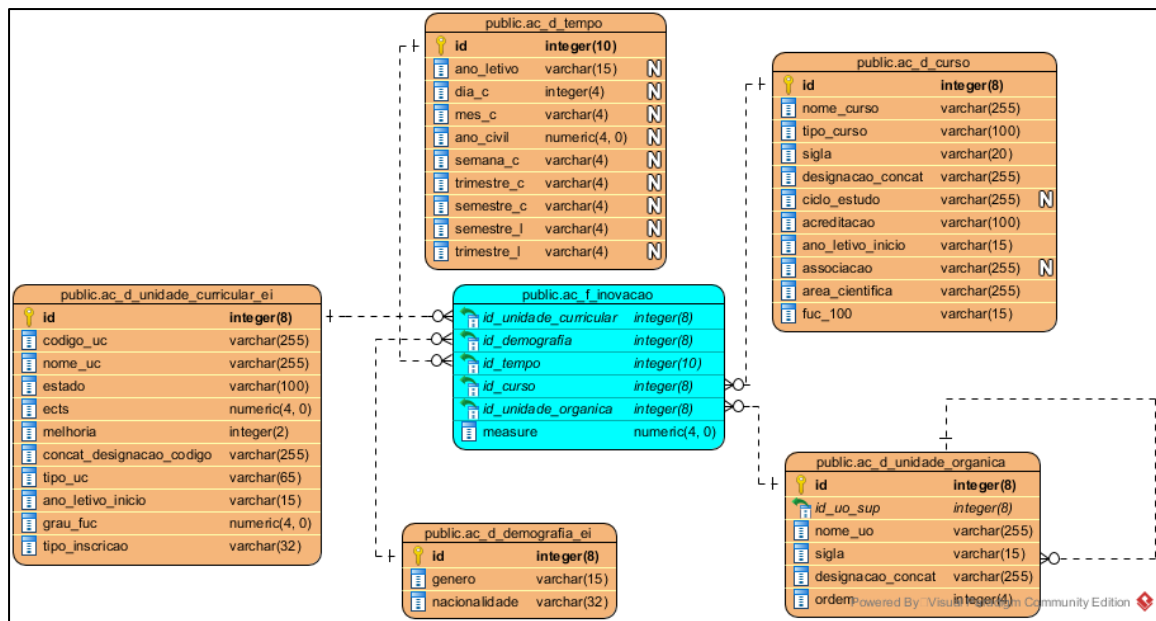


Figura 5 - Modelo de dados do indicador de empreendedorismo e inovação

Sendo que este indicador lida com dados de estudantes da universidade de Coimbra, estes dados encontram-se armazenados nas bases de dados do NONIO, pelo que foi necessário extrair esta informação desta fonte. Estes dados foram disponibilizados através de uma vista materializada, que por definição, contém o resultado de uma *query*. Esta vista era composta pelas inscrições em disciplinas dos alunos da universidade. Depois de saber como ir buscar os dados, o próximo passo foi a implementação do ETL. Como referido anteriormente a ferramenta usada foi o *Kettle (Pentaho Data Integration)*.

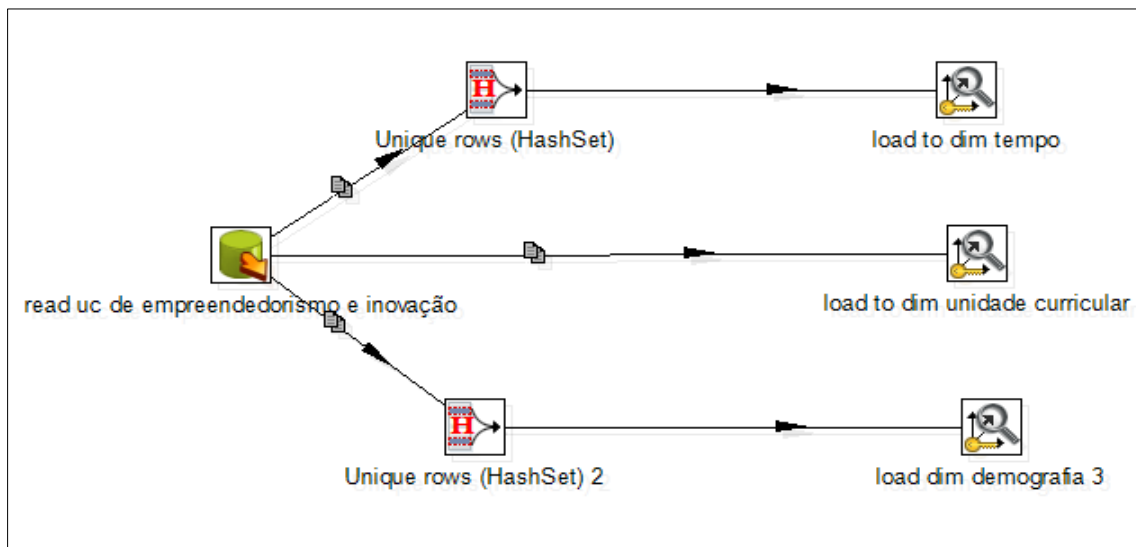


Figura 6- Extração dos dados dos alunos do NONIO

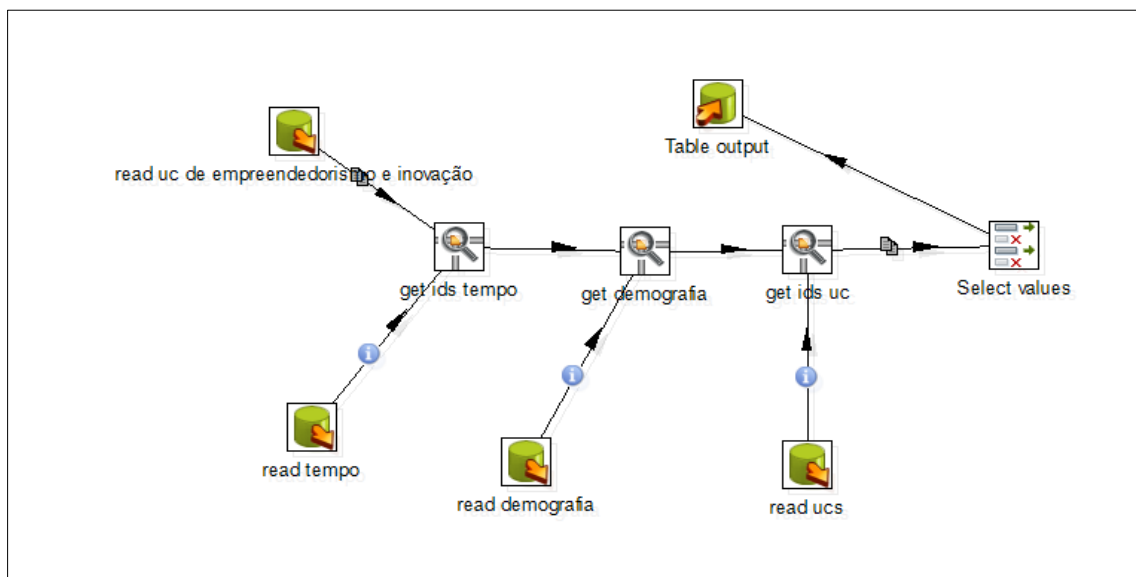


Figura 7 - Carregamento da tabela de factos

Depois destes dados estarem nas respetivas tabelas, representado por um *schema*, o próximo passo foi a criação de cubos tendo em conta as dimensões e a tabela de factos. Neste passo foi utilizado o *Mondrian Schema Workbench* que permite a criação de cubos OLAP visualmente.

O *Mondrian* é um sistema OLAP que permite ler fontes de dados na forma JDBC. Este é facilmente integrável com o servidor *pentaho* onde a plataforma é disponibilizada. Muitas das dimensões utilizadas na *data warehouse* são partilhadas entre cubos, como por exemplo, a dimensão do tempo, pelo que se evita rescrever esta dimensão aquando da criação de um novo cubo.

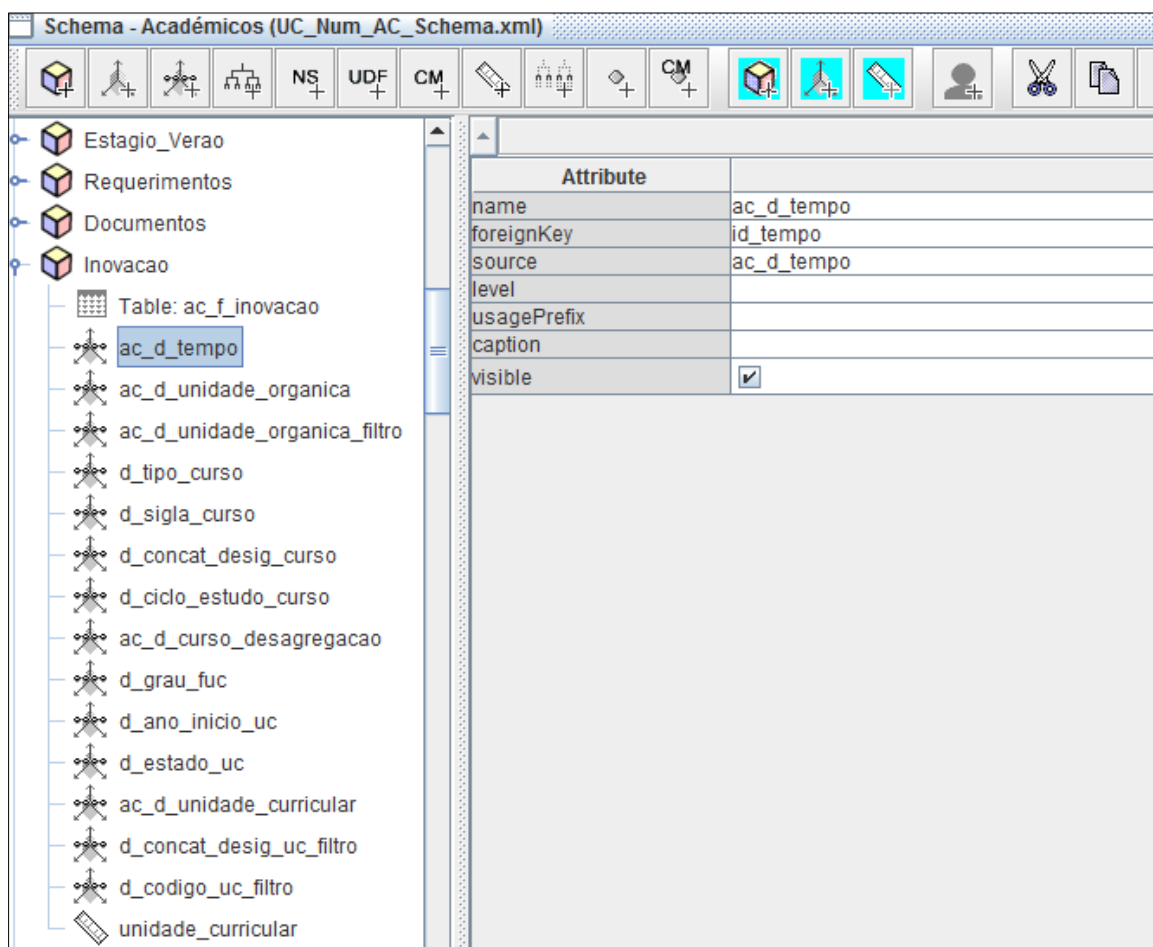


Figura 8 - Schema Workbench, criação de cubos

Nesta ferramenta é também possível escrever e testar *queries* MDX, antes de colocar em testes ou produção. Por exemplo a seguinte *query* MDX, permite saber o número de alunos em cada ano letivo que frequentam unidades curriculares de empreendedorismo ou inovação.

```
select NON EMPTY [Measures].[unidade_curricular]
ON COLUMNS, NON EMPTY {
Hierarchize({[ac_d_tempo.hie_ano_letivo]. [ano_letivo].Members})
}
ON ROWS
FROM [Inovacao]
```

No exemplo acima “Inovacao” é o nome do cubo OLAP de onde se pretende extrair os dados. “[Measures]. [unidade_curricular]” é um agregador que internamente realiza um SELECT DISTINCT nos dados selecionados. Por fim é também necessário especificar o eixo.

Esta linguagem MDX, semelhante a SQL, serve, portanto, para selecionar dados de cubos OLAP e ainda fazer cálculos sobre os dados. Tomando como exemplo a *query* acima representada esta irá correr na *dashboard* final que mostrará o indicador.

Em relação ao indicador que lida com o número de visitas a páginas do domínio *uc.pt* (tabela 11) era importante saber que tipo de dispositivo um dado utilizador usou para aceder a uma dada página, bem como outras métricas como o idioma, a geografia, a língua utilizada no browser entre outros. Para isso o modelo de dados que dá suporte a este indicador é apresentado na figura abaixo:

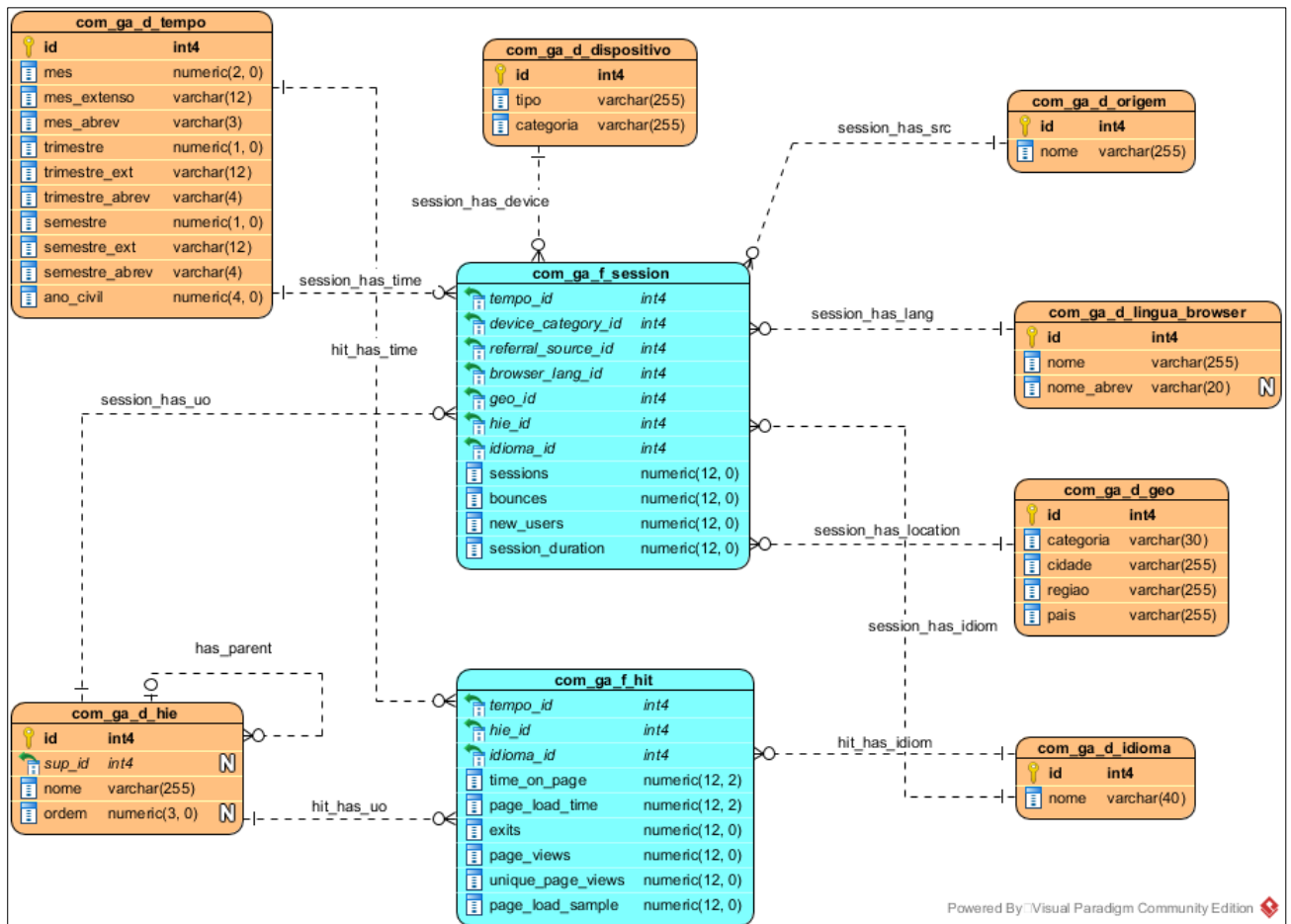


Figura 9 - Modelo de dados do indicador de comunicação

Associado a este indicador, a implementação do número de seguidores nas redes sociais (tabela 12) passou por utilizar as APIs públicas para extração dos dados. Aqui apenas se considerou duas métricas: o tempo e o número de seguidores. *Facebook*, *Twitter*, *Youtube* e *Instagram* oferecem aos utilizadores uma API onde cada utilizador que pretenda utilizar este serviço necessita de registar uma chave de acesso. Desta maneira, estas plataformas conseguem evitar e limitar o número de pedidos por utilizador para aceder a certos dados.

Finalmente, para dar suporte ao modelo de dados e cubos OLAP, as *dashboards* apresentadas ao utilizador final mostram os dados recolhidos. Cada *dashboard* apresentada ao utilizador final é suportada por duas ferramentas:

- *Pentaho CDE*, que simplifica a criação, *design* e visualização dos dados. Aqui é possível definir que *scripts* é que a *dashboard* final apresentará ao utilizador.
- *Pentaho CDA*, que simplifica o acesso aos dados fonte. Aqui é definido qual o ficheiro *Mondrian*, ficheiro XML que contém informação sobre os cubos OLAP, é que o servidor deve aceder para fazer *queries* sobre os dados.

Todos estas ferramentas devem estar em perfeita harmonia para que não sejam levantados problemas aquando da disponibilização final dos dados de um indicador. Por fim estes dados são mostrados ao utilizador final, o que resulta numa *dashboard*, como se pode ver na figura abaixo:

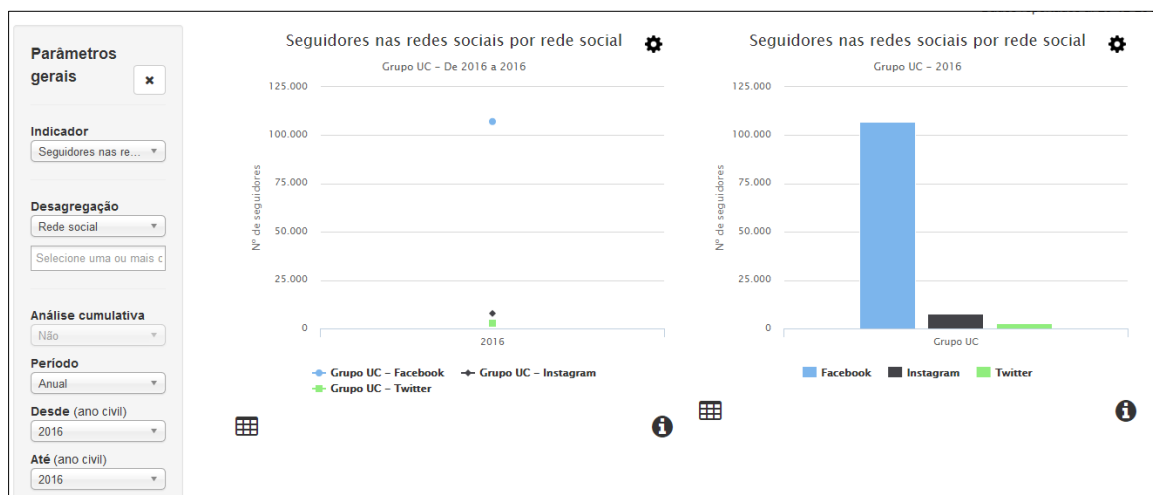


Figura 10 – Dashboard do indicador de Comunicação

Capítulo 5

Plano de trabalhos

Neste capítulo é apresentado o planeamento do estágio para o primeiro e segundo semestre. Este planeamento pretende ter como inspiração a metodologia proposta por *Ralph Kimball* [4], que demonstrou ao longos dos anos ser uma das referências na construção de *Data Warehouses*.

5.1 – Primeiro Semestre

No início deste estágio foi definido o plano relativo ao semestre em que se pretendia fazer inicialmente uma contextualização do projeto. Dado este projeto ter várias pessoas externas que auxiliam na recolha dos dados (equipa do nónio, pessoal dos recursos humanos, entre outros) foi feita uma reunião inicial com o Engenheiro Jorge China, diretor do serviço SGSIIC, Nelson Costa, chefe de divisão do SGSIIC e o professor Dr. Bruno Cabral. Aqui foi feito um contexto inicial á plataforma e um primeiro contacto com a mesma.

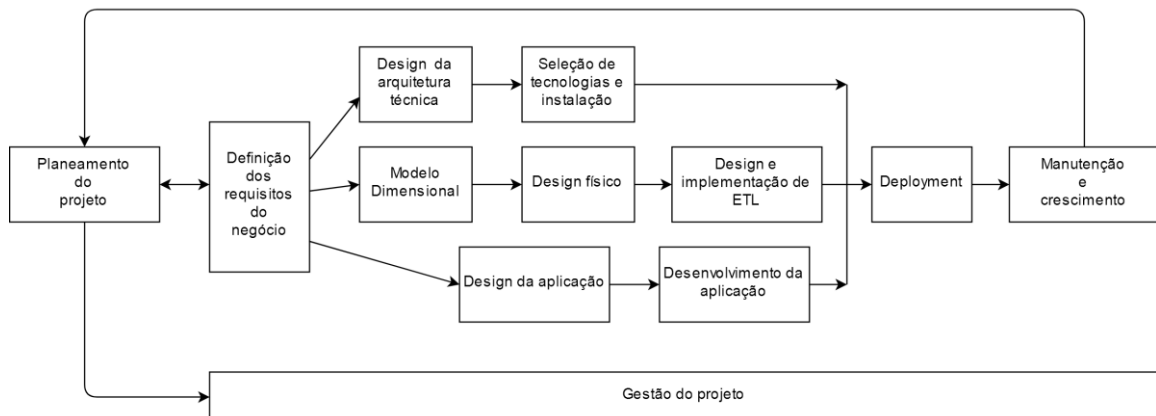


Figura 11 - Planeamento de um projeto de BI de acordo com *Ralph Kimball*

O próximo passo foi de tomar conhecimento das tecnologias que estavam a ser utilizadas no projeto. Este passo revelou-se importante pois toda a integração com bases de dados, processos ETL e disponibilização dos dados ao utilizador final é foco deste projeto e

estágio. De acordo com o plano inicial, grande parte do tempo seria dedicado a isto, ou seja, aprendizagem das metodologias de desenvolvimento, das tecnologias e ferramentas a utilizar.

O estado da arte, em que se inseria o trabalho realizado por alunos anteriores, foi também um passo importante para a compreensão das escolhas existentes no sistema. Dada a boa organização do projeto, este apresenta uma documentação interna onde é registada todos os artefactos inerentes ao projeto (código fonte, requisitos, fichas de indicadores, dados de acesso á máquinas, etc). Aliado a isto, os relatórios forneceram também uma ajuda importante na medida em que se justificava decisões de *design* tomadas.

Nas várias reuniões tidas ao longo do semestre, foi identificado um plano de reparações ao sistema atual, porque, como referido anteriormente, alguns dados deixaram de funcionar e pretendia-se resolver o mais rápido possível estes erros antes de avançar para qualquer implementação nova de indicadores.

Este plano de reparações focou-se em dois aspectos importantes:

- Dos indicadores disponíveis em produção e testes, identificar que indicadores deixaram de funcionar;
- Corrigir os indicadores identificados no passo anterior

Esta auditoria revelou-se extremamente importante, pois deu a conhecer como o sistema está desenvolvido, os vários ambientes de desenvolvimento, produção e testes. Estes ambientes, permitem seguir uma abordagem segura no desenvolvimento das *data marts*, pois só depois do código ter sido desenvolvido e testado é que é colocado em produção.

Na parte final do primeiro semestre ficaria reservada para a escrita do relatório e resolução de eventuais problemas encontrados ao longo do semestre.

Tarefa	Fevereiro			Março			Abril				Maio				Junho				
	2W	3W	4W	1W	2W	3W	4W	1W	2W	3W	4W	1W	2W	3W	4W	1W	2W	3W	4W
Contextualização																			
Estado da arte																			
Estudo das Tecnologias																			
Auditoria ao Sistema Atual																			
Plano de Reparções																			
Sprints de desenvolvimento																			
Escrita do relatório																			

Figura 12 - Plano para o primeiro semestre

Este plano inicial sofreu alterações, tanto por causa de atrasos na realização das tarefas propostas por parte do estagiário como resolução de alguns problemas de configuração de máquinas de desenvolvimento. Isto levou a que, o tempo reservado a algumas tarefas diminuiu drasticamente o que levou ao incomprimento das mesmas. Ainda assim, visto este semestre ser um primeiro contacto com as tecnologias e aprendizagem das mesmas, está previsto que o trabalho previsto para o segundo semestre, seja mais estruturado e com o cumprimento das tarefas previstas.

5.2 – Plano 2º semestre

O plano para o segundo semestre passou por fazer a definição dos indicadores chave (ficha formal do indicador e sua validação). De seguida, a criação da arquitetura, onde é feita a definição dos elementos de software internos e externos (fontes de dados) a utilizar. Nesta fase são também feitos o modelo de dados e o posterior desenho do processo ETL.

Tarefa	Setembro			Outubro				Novembro				Dezembro				Janeiro			
	2W	3W	4W	1W	2W	3W	4W	1W	2W	3W	4W	1W	2W	3W	4W	1W	2W	3W	4W
Definição de KPIs																			
Arquitetura																			
Implementação																			
Verificação/Validação																			
Deployment																			
Manutenção																			
Documentação Interna																			
Escrita do relatório																			

Figura 13 - plano para o segundo semestre

A fase de implementação, segue uma abordagem iterativa, onde é desenvolvido um indicador a cada iteração e que passa pela implementação do ETL, modelo de dados e do *front end*. Após isto, são feitos testes ao que foi desenvolvido, é criada a ficha de validação (e validação da mesma) do indicador implementado e como passo final o *deployment* do trabalho realizado. Qualquer encontrado nesta fase deve ser corrigido.

Como a documentação também assume um passo importante num projeto de *software*, irá ser criada uma documentação interna que auxilie e explique o trabalho realizado.

Ao longo do semestre, irá ser escrito o relatório final que passa por descrever todos estes passos.

5.3 – Resultados

Neste estágio foram identificadas as áreas de desenvolvimento a realizar, que encobriu as áreas académicas, investigação, projetos de investigação do PES e comunicação.

Ao longo da primeira parte do estágio, visto ainda não ter uma área em específico em que trabalhar, foi proposta a resolução de problemas que existiam na plataforma. Esta primeira fase do estágio focou-se sobretudo na resolução de problemas existentes na plataforma resultantes de factores externos.

Ainda assim, ficou clara e cumprido o plano de estudo das tecnologias, estado da arte e compreensão de como o sistema opera, de maneira a que, no segundo semestre, este trabalho fosse mais rápido, pois pressupõe-se que o aluno já conhece o sistema em que está a trabalhar. Esta resolução de problemas existentes na plataforma não foi cumprido na totalidade, mas que deu para perceber quais as áreas em que se deve focar mais no desenvolvimento de novos indicadores.

A segunda parte do estágio, em contraste com a primeira parte, correu bem melhor, visto já ter todas as condições necessárias ao desenvolvimento de indicadores. Foram feitas diversas reuniões entre *stakeholders* e pessoas externas ao projeto que ajudaram no desenvolvimento destes indicadores. Aqui, o *feedback* entre as várias partes interessadas foi fundamental para a concepção e validação destes indicadores.

Ao longo do estágio foi feito o relatório e a escrita de documentação interna para o repositório do projeto, pois uma boa documentação é fundamental para o sucesso de qualquer projeto de *software*.

Capítulo 6 - Conclusão

Neste capítulo são apresentadas as conclusões sobre o trabalho realizado ao longo do estágio.

Apesar do primeiro semestre ter sido marcado por vários momentos difíceis como a integração na equipa do projeto e problemas na infraestrutura deixados por alunos anteriores, foi feito um esforço por parte do estagiário para se manter atualizado.

A criação de condições para que o decorrer do projeto fosse sem percalços foi também tomada em conta. Aqui a marcação de reuniões com os responsáveis pela infraestrutura para perceber eventuais problemas na plataforma revelaram-se essenciais.

Considero que o trabalho feito poderia ter sido bem melhor. A recolha de indicadores e sua implementação são um processo chave deste projeto, mas que por vezes tomam mais tempo do que o planeado, tanto por fatores internos como externos.

Pessoalmente penso que este projeto tem futuro, mas que depende em grande parte de pessoas externas ao projeto que por vezes pode complicar a realização do mesmo.

Acrescento ainda que, apesar de este projeto apresentar várias etapas até a disponibilização final de um indicador, todo este trabalho seria melhor coordenado com várias pessoas a trabalhar no mesmo, pois a delegação de tarefas seria mais rápida e eficaz.

A nível técnico seria interessante discutir, e talvez introduzir, a ideia de ETL vs ELT no processo de extração de indicadores.

Para concluir, *data warehouses* são um conceito extremamente relevante para os dias de hoje, dada a quantidade de dados que é necessário analisar e é para onde a indústria de *big data* se encaminha.

Referências

[1] UC-cloud - Infraestruturas e serviços de TIC para a cloud -
https://www.uc.pt/ciuc/UC-cloud_Relatorio_final.pdf

[2] Projeto UC-Num

http://www.uc.pt/damc/ficheiros/anexo_II.pdf

[3] Pesquisa em bases de dados bibliográficas -
https://www.dec.uc.pt/~ifranca/Bases_de_Dados.pdf

[4] KIMBALL, R, ROSS, M.2002 The Data Warehouse Toolkit – The Complete Guide to Dimensional Modeling (2nd edition)

Anexos

- A.** Especificação de indicadores para testes/validação
- B.** Grau de importância dos indicadores do PES
- C.** Protótipo rápido do indicador do módulo Académicos: “N.º de estudantes a frequentar unidades curriculares na área de empreendedorismo e inovação”
- D.** Protótipo rápido do indicador do módulo de comunicação: “N.º de seguidores por rede social”
- E.** Protótipo rápido do indicador do módulo de comunicação: “N.º de visitas a páginas do domínio UC”
- F.** Ficha de indicador: “N.º de estudantes a frequentar unidades curriculares na área de empreendedorismo e inovação”
- G.** Ficha de indicador: “N.º de seguidores por rede social”

Anexo A – Especificação de indicadores para testes/validação

Nota: IND_{módulo}_{número}

e.g IND_AC_024 (Indicador do módulo Académicos número 24)

<i>Módulo</i>	<i>Designação</i>
AC	Académicos
SE	Sucesso Escolar
RH	Recursos Humanos
GO	Gestão Orçamental
SAS	Serviços Ação Social

<i>Código</i>	<i>Indicador</i>
IND_AC_024	N.º de pautas emitidas
IND_AC_025	Tempo de disponibilização de classificações definitivas aos estudantes
IND_SE_001	Taxa de conclusão no tempo previsto
IND_AC_005	N.º de matriculados inscritos através do concurso nacional de acesso
IND_AC_009	Taxa de procura (concurso via escola)
IND_AC_012	Candidatos a unidades curriculares isoladas
IND_AC_038	N.º total de cursos
IND_AC_039	N.º de unidades curriculares
IND_SE_001	Conclusões em n anos
IND_SE_008	Duração da interrupção
IND_SE_001 + IND_SE_002	Taxa de crescimento do n.º de graduados por ciclo
IND_SE_001 + IND_SE_002	Duração do tempo de conclusão das teses de doutoramento
IND_RH_015	Média etária dos docentes de carreira
IND_RH_037	Percentagem de docentes de carreira com idade média inferior a 40 anos (média etária dos docentes de carreira) (nº)
IND_RH_037	Pessoal técnico com nível de habilitação igual ou superior a licenciatura
IND_RH_019	Rácio de trabalhadores
IND_RH_007	Antiguidade média
IND_RH_008	Índice de absentismo
IND_RH_010	Índice de admissão de trabalhadores
IND_RH_011	Índice de saída de trabalhadores
IND_RH_012	Índice de reposição
IND_RH_013	Índice de rotação
IND_RH_014	Índice de envelhecimento
IND_RH_015	Nível etário
IND_RH_016	Leque etário
IND_RH_017	Índice de emprego jovem
IND_RH_018	Índice de rejuvenescimento

IND_RH_020	Admissões
IND_RH_021	Saídas
IND_RH_030	Custo hora
IND_RH_032	Leque salarial ilíquido
IND_RH_033	Carga Salarial
IND_RH_034	Despesa com pessoal “per capita”
IND_RH_036	Remuneração base média
IND_RH_037	N.º de trabalhadores
IND_RH_038	ETIs
IND_RH_039	Despesa com remunerações
IND_RH_040	Despesa total com pessoal
IND_RH_041	Despesa média mensal com remunerações
IND_RH_050	Despesa média mensal total com pessoal
IND_RH_050	N.º de acidentes
IND_RH_002	Suspensão
IND_RH_035	Taxa de complementos e encargos sociais
IND_GO_025	Taxa de crescimento do financiamento competitivo
IND_GO_021	Equilíbrio entre a despesa executada e a receita arrecadada
IND_GO_019	Taxa de crescimento da receita resultante da prestação de serviços especializados
IND_GO_019	Financiamento competitivo
IND_GO_033	Receita de produtos com a marca UC
IND_GO_034	Taxa de crescimento da receita de produtos com a marca UC
IND_GO_020	Dívidas de clientes ou terceiros
IND_AC_021	N.º de estudantes em cursos de língua portuguesa para estrangeiros (taxa de crescimento de x%)
IND_AC_038	N.º de ciclos de estudo lecionados em parcerias com instituições estrangeiras de ensino superior
IND_AC_021 + IND_RH_037	N.º países diferentes na comunidade académica (superior a x)
IND_RH_037	N.º de trabalhadores de nacionalidade estrangeira (docentes, investigadores, pessoal técnico)
IND_AC_038	N.º de cursos lecionados em parceria internacional
IND_AC_039	N.º de unidades curriculares lecionadas em língua estrangeira
IND_AC_038	N.º de instituições internacionais parceiras (área de ensino)
IND_PES_SAS_ALIM_009	N.º de refeições servidas
IND_PES_SAS_ALIM_010	N.º médio de refeições por dia
IND_PES_SAS_ALIM_040	N.º de dias de funcionamento das unidades alimentares
IND_GO_033	Receita de produtos com marca UC (volume e taxa de crescimento)

Tabela 5 – Indicadores do ambiente de testes para validação

Anexo B – Grau de importância dos indicadores do PES

Ambiente	
Indicador	Prioridade
Área de painéis solares fotovoltaicos	Baixa
Produção de energia renovável	Baixa
Consumo de eletricidade por m2 utilizado	Baixa
Consumo de gás por m2 utilizado	Baixa
Consumo de água por m2 utilizado	Baixa
Área de painéis solares térmicos	Baixa
Posição no GreenMetric Ranking of World Universities	Baixa

Comunicação	
Indicador	Prioridade
Nº. de visitas à página da UC	Alta
Nº. de seguidores nas redes sociais	Alta
Média bienal do índice de notoriedade da UC, medido através do AVV	Baixa
Posição no Ranking Web of Universities e no 4 International Colleges & Universities	Baixa

Marca UC	
Indicador	Prioridade
Receita da venda de produtos com marca UC	Alta
Posição no QS World University Ranking, no Academic Ranking of World Universities, no Times Higher Education World University Ranking e no Scimago Institutions Rankings	Baixa
Nº. de prémios obtidos pela UC ou a membros da comunidade académica	Baixa

Cidadania e inclusão	
Indicador	Prioridade
Nº. de estudantes dirigentes associativos jovens, estudantes membros de órgãos da UC, estudantes com participação em atividades de reconhecimento mérito universitário e estudantes em ações de voluntariado credenciadas	Alta
Nº. de bolsiros alojados em residências	Alta
Porcentagem de estudantes que conhecem a política de promoção da igualdade de oportunidades, inclusão dos grupos sub-representados e valorização da diversidade	Baixa
Porcentagem de estudantes beneficiários de apoios sociais	Baixa

diretos	
Montante de apoios sociais diretos concedidos pela UC	Baixa

Internacionalização

Indicador	Prioridade
Percentagem de estudantes de nacionalidade estrangeira	Alta
Nº de estudantes ao abrigo do Estatuto do Estudante Internacional	Alta
Percentagem de estudantes de mobilidade <i>incoming</i>	Alta
Nº. de ciclos de estudo lecionados em parceria com instituições estrangeiras de ensino superior	Alta
Nº. de países representados na comunidade académica	Alta
Nº. de publicações em colaboração internacional,	Alta
Nº. de trabalhadores de nacionalidade estrangeira	Alta
Peso do financiamento de origem internacional	Alta
Receita de propinas provenientes de estudantes internacionais	Alta
Nº. de estudantes em cursos de língua portuguesa para estrangeiros	Alta

Infraestruturas

Indicador	Prioridade
Grau de cobertura por rede wireless dos espaços úteis com necessidades identificadas	Baixa
Grau de cobertura por fibra ótica dos espaços úteis com necessidades identificada	Baixa
Grau de implementação do plano com vista a assegurar as instalações adequadas a todas as unidades e serviços	Baixa
Área útil média por docente, estudante e pessoal técnico, por tipologia de espaços	Baixa
Custo por m2 com conservação, manutenção e requalificação dos edifícios	Baixa
Grau de abrangência do princípio do utilizador-pagador dos gastos operacionais das instalações	Baixa

Económico-Financeiros

Indicador	Prioridade
Volume de financiamento competitivo	Alta
Volume de financiamento obtido nos programas Horizonte 2020, Portugal 2020 e Centro 2020	Alta
Nível de diversificação da estrutura de financiamento	Alta
Receita resultante da prestação de serviços especializados	Alta
Taxa de aprovação de projetos	Alta

Pessoas	
Indicador	Prioridade
Nº. de docentes de carreira com idade inferior a 40 anos	Alta
Média etária dos docentes de carreira	Alta
Percentagem de docentes de carreira com doutoramento noutra instituição que não a UC	Alta
Percentagem de pessoal técnico com nível de habilitação igual ou superior a licenciatura	Alta
Percentagem de pessoal técnico com plano pessoal de desenvolvimento	Baixa
Percentagem de pessoal técnico que frequentou ações de formação	Baixa

Comunidade	
Indicador	Prioridade
Receita com visitas a espaços turísticos	Alta
Nº. de estudantes atletas de alto rendimento e atletas da UC	Alta
Nº. de estudantes integrados em atividades culturais da UC	Alta
Nº. de estudantes a frequentar unidades curriculares na área de empreendedorismo e inovação	Alta
Nº. de antigos estudantes ativos na Rede UC	Alta
Nº. de visitantes aos espaços turísticos	Alta
Nº. de estágios de ligação ao mercado de trabalho	Alta
Percentagem de publicações em colaboração com a indústria	Alta
Volume de negócios das spin-offs e start-ups	Baixa
Nº. de comunicações de invenção	Baixa
Nº. de spin-offs e startups criadas	Baixa
Nº. de empregos criados por spin-offs e start-ups	Baixa
Nº. de patentes ativas	Baixa
Nº. de utilizadores de instalações desportivas	Baixa

Ensino	
Indicador	Prioridade
Nº. de estudantes de doutoramento	Alta
Percentagem de estudantes colocados em 1ª opção no concurso nacional de acesso	Alta
Taxa de conclusão no tempo previsto	Alta
Taxa de sucesso escolar	Alta
Taxa de abandono escolar efetivo	Alta
Grau de satisfação dos estudantes	Alta
Nº. de estágios de ligação ao mercado de trabalho	Alta
Grau de empregabilidade	Alta
Nº. de ECTS completados em ensino à distância	Alta
Percentagem dos melhores 25% melhores candidatos no	Baixa

concurso nacional de acesso que escolhem a UC	
-----------------------------------------------	--

Investigação	
Indicador	Prioridade
Nº. de artigos nas 25% revistas de maior impacto	Alta
Nº. de artigos nas 10% revistas de maior impacto	Alta
Nº. de artigos nas 1% revistas de maior impacto	Alta
Nº. de publicações por docente doutorado ETI na Web of Science	Alta
Nº. de citações por docente doutorado ETI na Web of Science	Alta
Nº. de teses de doutoramento concluídas	Baixa
Nº. de projetos europeus e internacionais	Baixa
Percentagem de docentes em unidades de I&D reconhecidas pela FCT	Baixa

Anexo C – Protótipo rápido do indicador do módulo Académicos: “N.º de estudantes a frequentar unidades curriculares na área de empreendedorismo e inovação”

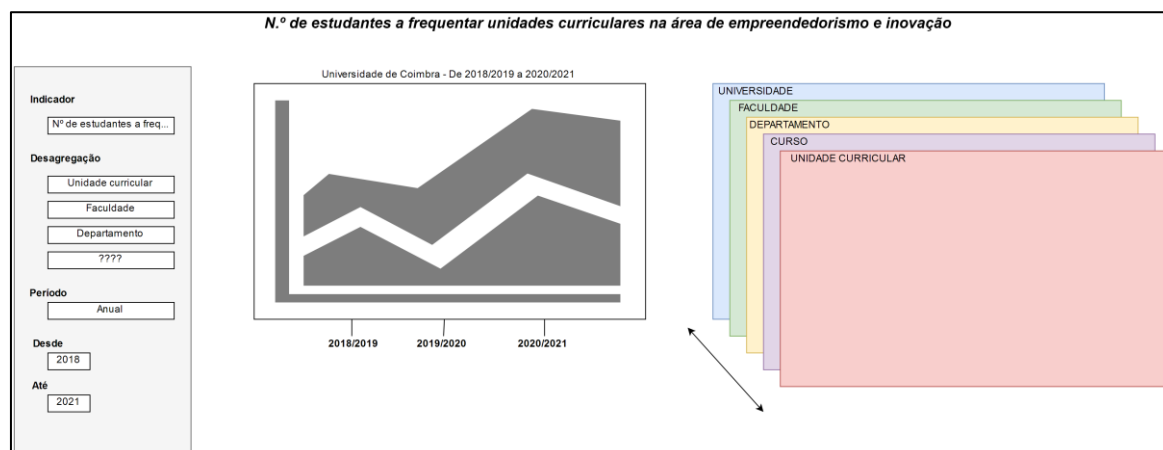


Figura 14 – Indicador do módulo Académicos

Neste indicador importava saber o número de estudantes a frequentar unidades curriculares na área de empreendedorismo e inovação. Na área lateral esquerda, é apresentado um menu de opções onde é possível filtrar por unidade curricular, faculdade, departamento. É possível também especificar o período a que correspondem os dados por exemplo (2015-2016).

Os gráficos apresentados correspondem aos valores dos filtros onde no primeiro gráfico a é mostrado os dados a que correspondem os filtros e no segundo gráfico é possível especificar que dados se quer.

Para ilustrar a ideia, se se quiser o número de alunos a frequentar unidades curriculares na área de empreendedorismo e inovação na FCTUC e no departamento de engenharia informática (*drill down* nos dados) o gráfico da esquerda serve para esse efeito.

Anexo D – Protótipo rápido do indicador do módulo de comunicação: “N.º de seguidores por rede social”

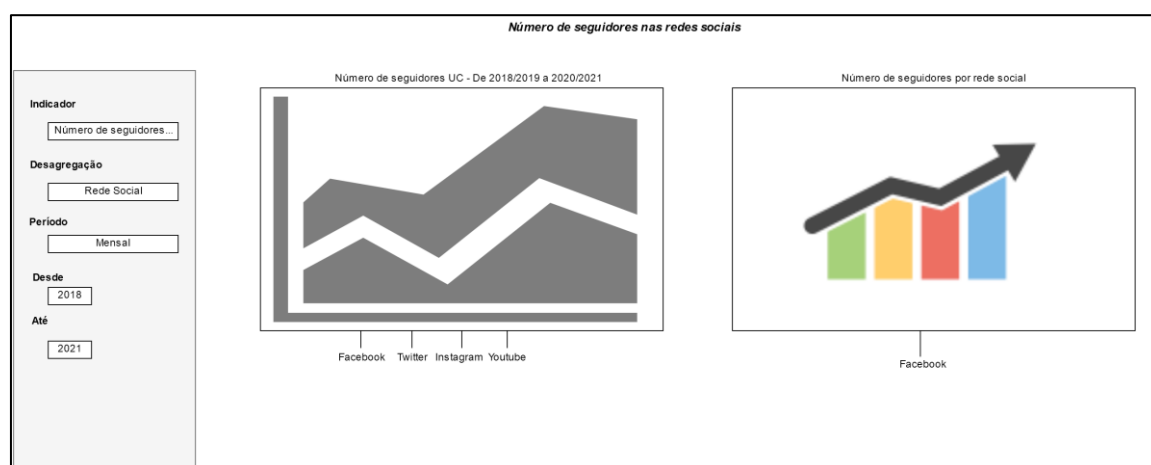


Figura 15 – Indicador do módulo de Comunicação

Neste indicador importava saber o número de seguidores por rede social. Os casos considerados foram as redes sociais *Facebook*, *Twitter*, *Instagram* e *Youtube*. Aqui, é possível discriminar os dados por um certo período de tempo por exemplo de 2014 a 2015.

O primeiro gráfico mostra os dados de todas as redes sociais e no gráfico da esquerda é possível especificar uma rede social e observar os dados independentemente.

Anexo E – Protótipo rápido do indicador do módulo de comunicação: “N.º de visitas a páginas do domínio UC”

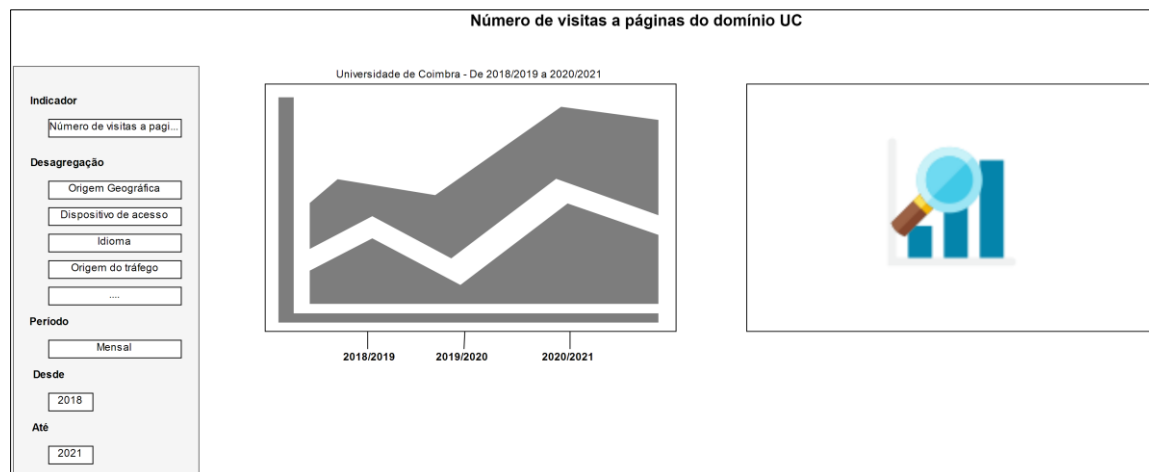


Figura 16 – Indicador do módulo de Comunicação

Anexo F – Ficha de indicador: “N.º de estudantes a frequentar unidades curriculares na área de empreendedorismo e inovação”

Histórico de versões para a ficha IND_COM_001:

Versão #	Autor	Data	Descrição
0.1	Adolfo Pinto	2019	Criação do documento e preenchimento inicial.

ID DO INDICADOR	NOME	DESCRIÇÃO
IND_INV_001	N.º de estudantes a frequentar unidades curriculares na área de empreendedorismo e inovação	N.º de estudantes a frequentar unidades curriculares na área de empreendedorismo e inovação (1.º, 2.º e 3.º ciclo)

GLOSSÁRIO

UC	Unidade Curricular (apenas se considera as que têm “empreendedorismo” e “inovação” no nome)
UO	Unidade Orgânica

FÓRMULA NA SUA GRANULARIDADE MÍNIMA

De todas as unidades curriculares que o aluno se encontra inscrito, filtrar por “empreendedorismo” e/ou “inovação” no nome.

DADOS DE ENTRADA

Descrição dos Dados	Validações
Inscrições em disciplinas (Nónio)	-

NÍVEIS DE GRANULARIDADE

Nível	Operações Associadas
Unidade curricular	Visualização da unidade curricular em que o aluno se encontra inscrito.
Faculdade	Visualização da faculdade (FCTUC, FDUC, FEUC, FFUC, FLUC, FMUC, FPCEUC, FCDEF, Colégio das Artes, Instituto de investigação interdisciplinar, Reitoria, Tribunal Universitário Judicial e Europeu, ICNAS, Entidade Subsidiária de Direto Privado, Centro de Investigação e Desenvolvimento)

TEMPO

Período de Tempo	Operações Associadas
Ano	Evolução temporal discriminada por ano letivo.

Permite análise cumulativa?

Sim Não

Permite análise período homólogo?

Sim Não

FILTROS GERAIS

Nome	Descrição
Faculdade	Faculdades da Universidade de Coimbra
Unidade curricular	Cadeira/disciplina
Tipo de disciplina	Isolada, avulsa, plano de estudos

FONTES DE INFORMAÇÃO

Sistemas Fonte	Vistas/Tabelas/WS	Campos
Nónio	MVIEW_INSCRICOES_DISC	

PERIODICIDADE DA RECOLHA DE DADOS

Semestral

RESPONSÁVEL PELO MÓDULO

Dra. Madalena Alarcão

PESSOA(S) E SERVIÇO(S) RESPONSÁVEL(EIS) PELA ESPECIFICAÇÃO

Divisão de planeamento, gestão e desenvolvimento	<serviços>
--------------------------------------------------	------------

ESPECIFICAÇÃO VALIDADA POR

DATA VALIDAÇÃO

Filipe Rocha	
--------------	--

OBSERVAÇÕES

--

Anexo G – Ficha de indicador: “Nº de seguidores por rede social”

Histórico de versões para a ficha IND_COM_001:

Versão #	Autor	Data	Descrição
0.1	Ana Ilharco	19-01-2016	Criação do documento e preenchimento inicial.
0.2	Adolfo Pinto	16-01-2019	Alteração dos dados

ID DO INDICADOR NOME DESCRIÇÃO

IND_COM_021	Número de seguidores nas redes sociais.	Número de seguidores nas redes sociais.
-------------	-----------------------------------------	-----------------------------------------

GLOSSÁRIO

API	<i>Application Programming Interface</i>
GC	Gabinete de Comunicação (Assessoria de Imprensa)
PIMC	Projeto Especial Imagem, Media e Comunicação (http://www.uc.pt/pimc)
UC	Universidade de Coimbra
UO	Unidade Orgânica

FÓRMULA NA SUA GRANULARIDADE MÍNIMA

Número de seguidores ou “gostos” nas redes sociais da UC, por tipo de rede social, por mês.

DADOS DE ENTRADA

Descrição dos Dados	Validações

NÍVEIS DE GRANULARIDADE

Nível	Operações Associadas
Grupo UC	
Unidades Grupo UC	
Unidades orgânicas	

TEMPO

Período de Tempo	Operações Associadas
Ano civil	Evolução temporal discriminada por ano civil.
Semestral	Evolução temporal discriminada por semestre civil.
Trimestral	Evolução temporal discriminada por semestre civil.
Mensal	Evolução temporal discriminada por um intervalo de meses

Permite análise cumulativa?

Sim Não

Permite análise período homólogo?

Sim Não

FILTROS GERAIS

Nome	Descrição
Tipo de rede social	Discriminação ou filtragem dos dados por tipo de rede social (e.g. Facebook, Twitter, Instagram, Youtube)

FONTES DE INFORMAÇÃO

Sistemas Fonte	Vistas/Tabelas/WS	Campo s
Twitter	https://developer.twitter.com/en/docs.html	
Facebook	https://developers.facebook.com/docs/graph-api/	
Youtube	https://developers.google.com/youtube/reporting/v1/code_samples/	
Instagram		

PERIODICIDADE DA RECOLHA DE DADOS

Mensal

RESPONSÁVEL PELO MÓDULO

VR Clara Almeida Santos

PESSOA(S) E SERVIÇO(S) RESPONSÁVEL(EIS) PELA ESPECIFICAÇÃO

Joana Tereso	PIMC
Joana Ferreira	SASUC

ESPECIFICAÇÃO VALIDADA POR**DATA VALIDAÇÃO**

<responsável pela especificação>	
----------------------------------	--

OBSERVAÇÕES

--