

João Daniel Aguiar Castro

Estudo de utilização do Repositório de dados da Universidade do
Porto

Dissertação realizada no âmbito do Mestrado em Ciência da Informação, orientada pela
Professora Doutora Maria Cristina de Carvalho Alves Ribeiro

Faculdade de Engenharia e Faculdade de Letras
Universidade do Porto

Julho de 2013

Estudo de utilização do Repositório de dados da Universidade do Porto

João Daniel Aguiar Castro

Dissertação realizada no âmbito do Mestrado em Ciência da Informação, orientada pela
Professora Doutora Maria Cristina de Carvalho Alves Ribeiro

Membros do Júri

Professor Doutor Gabriel de Sousa Torcato David - Presidente
Faculdade de Engenharia - Universidade do Porto

Professora Doutora Ana Alice Rodrigues Pereira Baptista - Arguente
Universidade do Minho

Professora Doutora Maria Cristina de Carvalho Alves Ribeiro - Orientadora
Faculdade de Engenharia - Universidade do Porto

.....

Agradecimentos

A presente dissertação de mestrado contou com importantes incentivos de diversas pessoas. Às seguintes expresse um sincero agradecimento...

À minha orientadora, Professora Cristina Ribeiro, pelo apoio, disponibilidade em acompanhar todas as fases deste trabalho, e acima de tudo pelo espírito crítico e desafios propostos ao longo deste percurso;

Aos investigadores da Universidade do Porto que acederam em colaborar neste trabalho: ao Dr. Marcelo Moura, ao Dr. José Luís Moreira, à Dr^a Luísa Bastos, à Dr^a Ângela Lomba, ao Dr. Rosaldo Rossetti; agradecimentos estendidos ao Dr. Paco Ruiz da Universidade de Valência e ao Dr. Fábio Pinto da Costa, sem esquecer a valiosa contribuição da Dr^a Maria Eugénia Fernandes, da Reitoria da Universidade do Porto, pela mediação no contato com os investigadores;

Aos meus colegas do laboratório I123, em particular ao João Rocha da Silva pelo contributo impar e disposição para me aconselhar sempre que solicitado, e ainda ao José Pedro Barbosa e à Mariana Gouveia cujo trabalho que desenvolveram é intrínseco a este trabalho;

A todos os docentes do Mestrado em Ciência da Informação pela partilha de conhecimentos;

A toda a minha família pelo apoio, sacrifício e paciência ilimitada, em especial à tia Glória, pela aposta pessoal em mim, que me permite escrever estas palavras neste momento... e ainda à D. Fernanda e Sr. Adelino por todos os incentivos.

Por fim, à minha namorada, Alexandra, pessoa responsável pelas escolhas mais sensatas que tomei ao longo dos últimos anos...obrigado por acreditares, Sempre!

A todos OBRIGADO!

Resumo

Verifica-se atualmente uma crescente preocupação da comunidade científica em relação ao acesso e partilha dos dados resultantes do processo de investigação. Impulsionados pela tecnologia, os dados de investigação têm vindo a ser produzidos em grande quantidade, o que representa um desafio à sua gestão. Os repositórios digitais, como plataformas integradas e robustas de armazenamento, preservação e disseminação de informação, podem significar uma oportunidade de gestão eficiente em custos dos dados de investigação. Nesse sentido várias iniciativas relacionadas com repositórios digitais tendem a adotar estratégias que garantam os princípios de integridade dos dados ao longo do seu ciclo de vida, sendo a curadoria de dados um conceito emergente neste contexto. O Repositório de dados da Universidade do Porto, em fase experimental, apresenta um conjunto de funcionalidades e propõe um procedimento de depósito e gestão de dados, assente em duas plataformas integradas, de acordo com necessidades previamente identificadas para os investigadores. Antes de generalizar o uso das plataforma junto dos investigadores é necessário avaliar a sua conformidade com as necessidades dos investigadores. Para isso foram realizados testes de utilização junto de investigadores da Universidade do Porto, que por um lado permitiram identificar os pontos fortes e fracos do repositório e das ferramentas de gestão de dados, e por outro lado fazer um levantamento de sugestões de funcionalidades. A abordagem metodológica, resultante da análise da literatura de estudos de utilização em repositórios digitais, identifica os atributos de qualidade de Jakob Nielsen e o modelo de referência de Judy Jeng como critérios de avaliação a aplicar nos testes de utilização. Além disso, o trabalho desenvolvido implicou a seleção de descritores, de acordo com as recomendações da DCMI (Dublin Core Metadata Initiative), para um conjunto de dados reais já depositados no repositório experimental, nomeadamente através da elaboração de perfis de aplicação para diferentes domínios.

Palavras-chave: repositório de dados; estudo de utilização; curadoria de dados; gestão de dados de investigação; perfil de aplicação.

Abstract

Currently there is a growing concern in the scientific community regarding the access and sharing of data resulting from the research process. Driven by technology, research data has been produced in large scale, which represents a challenge to their management. Digital repositories as integrated and robust platforms for storage, preservation and dissemination, are an opportunity for cost-effective management of research data. Accordingly, various initiatives related to digital repositories tend to adopt strategies that ensure the principles of data integrity throughout its lifecycle, and data curation is an emerging concept in this context. The *Repositório de dados da Universidade do Porto*, which is currently in an experimental stage, proposes a procedure for data management and storage, based on two integrated platforms, based on researchers identified needs. Before the widespread use of these platforms among researchers, it is necessary to assess their compliance with researchers requirements. So, tests were performed with researchers from the University of Porto, allowing to identify strengths and weaknesses in the repository and the data submission process, and to make a survey of suggested functionalities. The methodological approach results from the literature review of studies of use in digital repositories, and led to the adoption of quality attributes of Jakob Nielsen and the reference model made by Judy Jeng as evaluation criteria to be applied in the usability tests. Moreover, the work involved the selection of descriptors, according to DCMI (Dublin Core Metadata Initiative) recommendations, to a set of actual pending data at the experimental repository, particularly by drawing application profiles for different domains.

Keywords: data repository; user study; data curation; research data management; application profile.

Lista de ilustrações

Figura 1: Etapas do ciclo de vida da investigação e dados (JISC).....	7
Figura 2: Entidades funcionais OAIS (CCSDS, 2012).....	16
Figura 3: Tipo de conteúdos em repositórios de acesso aberto (OpenDoar, 2013).....	21
Figura 4: Data Audit Framework.....	25
Figura 5: Requisitos para uma infraestrutura de gestão de dados de investigação da Un. de Oxford (Wilson et al, 2011).....	26
Figura 6: Modelo de Comportamento Informacional (Wilson, 1981).....	31
Figura 7: Excerto de dados de um ensaio em fratura de materiais.....	43
Figura 8: Excerto de dados de um ensaio em química analítica.....	47
Figura 9: Excerto de registo de dados na campanha experimental, domínio da dinâmica de estruturas.....	50
Figura 10: Excerto de dados do domínio gravimetria.....	55
Figura 11: Modelo de avaliação de usabilidade de bibliotecas digitais (Jeng, 2005).....	61
Figura 12: Interface inicial do Repositório de dados da Universidade do Porto.....	62
Figura 13: Casos de uso do Repositório de dados da Universidade do Porto.....	63
Figura 14: Página de coleção “Aereal Gravimetry Runs” (Repositório de dados).....	66
Figura 15: Navegação por autor (Repositório de dados).....	68
Figura 16: Navegação por assunto (Repositório de dados).....	69
Figura 17: Ordenação dos ficheiros por ordem descendente (Repositório de dados).....	70
Figura 18: Visualização de ficheiros associados a um item (Repositório de dados).....	72
Figura 19: Registo de metadados completo (Repositório de dados).....	72
Figura 20: Explorar e descarregar dados (Repositório de dados).....	74
Figura 21: Filtragem de dados (Repositório de dados).....	76
Figura 22: Subscrição de coleção (Repositório de dados).....	78
Figura 23: <i>Workflow</i> do processo de submissão de dados no Repositório de dados da Universidade do Porto (Barbosa, 2013).....	83
Figura 24: Casos de uso UPBox (Barbosa, 2013).....	84

Figura 25: Casos de uso DataNotes (Gouveia, 2013).....	85
Figura 26: Processo de criação de um projeto (UPBox).....	86
Figura 27: Caixa de adição de colaborador (UPBox).....	88
Figura 28: Criação de pastas e <i>upload</i> de ficheiros (UPBox).....	89
Figura 29: Visualização do estado de carregamento de ficheiros (UPBox).....	90
Figura 30: Hiperligação para <i>download</i> de ficheiro (UPBox).....	91
Figura 31: Eliminação de ficheiros (UPBox).....	92
Figura 32: Iniciar processo de anotação (UPBox).....	93
Figura 33: Seleção de ficheiro a anotar (DataNotes).....	95
Figura 34: Página de edição de metadados (DataNotes).....	97
Figura 35: Tabela de descrições (DataNotes).....	100
Figura 36: Caixa de pesquisa (DataNotes).....	101
Figura 37: Seleção de metadados (DataNotes).....	102
Figura 38: Recuperação de um campo previamente descrito (DataNotes).....	103

Lista de tabelas

Tabela 1: Objetivos dos estudos de usabilidade em repositórios digitais.....	36
Tabela 2: Técnicas e critérios de usabilidade em repositórios digitais.....	36
Tabela 3: Perfil de aplicação “fm.”	45
Tabela 4: Perfil de aplicação “qa.”	48
Tabela 5: Perfil de aplicação “de.”	51
Tabela 6: Perfil de aplicação do domínio da ecologia georreferenciada (norma INSPIRE – ISO 19-115).....	53
Tabela 7: Perfil de aplicação “grav.”.....	56
Tabela 8: Descrição dos atributos de qualidade de Nielsen (2003).....	60
Tabela 9: Resultados da tarefa 1 (Repositório de dados).....	67
Tabela 10: Resultados da tarefa 2 (Repositório de dados).....	68
Tabela 11: Resultados da tarefa 3 (Repositório de dados).....	69
Tabela 12: Resultados da tarefa 4 (Repositório de dados).....	71
Tabela 13: Resultados da tarefa 5 (Repositório de dados).....	73
Tabela 14: Resultados da tarefa 6 (Repositório de dados).....	75
Tabela 15: Resultados da tarefa 7 (Repositório de dados).....	77
Tabela 16: Resultados da tarefa 8 (Repositório de dados).....	79
Tabela 17: Recomendações de melhoria Repositório de dados da Universidade do Porto.....	82
Tabela 18: Resultados tarefa 1 (UPBox).....	87
Tabela 19: Resultados tarefa 2 (UPBox).....	88
Tabela 20: Resultados tarefa 3 (UPBox).....	90
Tabela 21: Resultados tarefa 4 (UPBox).....	91
Tabela 22: Resultados tarefa 5 (UPBox).....	92
Tabela 23: Resultados tarefa 6 (UPBox).....	94
Tabela 24: Resultados tarefa 7 (DataNotes).....	96
Tabela 25: Resultados tarefa 8 (DataNotes).....	97

Tabela 26: Resultados tarefa 9 (DataNotes).....	100
Tabela 27: Resultados tarefa 10 (DataNotes).....	101
Tabela 28: Resultados tarefa 11 (DataNotes).....	102
Tabela 29: Recomendações de melhoria UPBox e DataNotes.....	107

Lista de abreviaturas e siglas

ANDS – *Australian National Data Service*

CCSDS – *Consultative Committee for Space Data Systems*

CSDGM – *Content Standard for Digital Geospatial Metadata*

CRUS – *Center for Research on User Studies*

DAF – *Data Audit Framework*

DANS – *Data Archiving and Network for Earth*

DC – *Dublin Core*

DCB – *Double Cantilever Beam*

DCMI – *Dublin Core Metadata Initiative*

DCC – *Digital Curation Center*

DDCTF – *Digital Data Curation Task Force*

DDI – *Data Documentation Initiative*

de. – *Dinâmica de Estruturas*

DPE – *Digital Preservation Europe*

EDINA – *Edinburgh Data and Information Access*

EML – *Ecological Metadata Language*

ESSD – *Earth System Scientific Data*

FEUP – *Faculdade de Engenharia da Universidade do Porto*

fm. – *Fratura de Materiais*

grav. – *Gravimetria*

NCBI – *National Center for Biotechnology Information*

NDIIPP – *National Digital Information and Infrastructure Preservation Program*

NSF – *National Science Foundation*

JISC – *Joint Information Systems Committee*

ICPSR – *Interuniversity Consortium for Political and Social Research*

ISO – *International Organization for Standardization*

KRDS – *Keeping Research Data Safe*

OAIS – *Open Archival Information System*

OpenDoar – *The Directory of Open Access Repositories*

OKF – *Open Knowledge Foundation*

OPN – *Open Planets Foundation*

qa. – *Química Analítica*

RDA – *Research Data Alliance*

U. Porto – *Universidade do Porto*

VRE – *Virtual Research Environment*

Sumário

Introdução	1
Capítulo 1.	
Dados de investigação e Repositórios de dados	5
1.1 Dados de investigação: produção e partilha.....	5
1.1.1 Os investigadores e a partilha dos dados de investigação.....	8
1.1.2 O acesso aberto aos dados de investigação.....	10
1.2 A preservação de dados de investigação.....	12
1.3 Curadoria de dados.....	16
1.4 Repositórios digitais.....	18
1.5 Repositórios de dados.....	20
1.6.1 Repositórios de dados e a gestão dos dados de investigação.....	23
1.6.2 Gestão de dados de investigação: o caso do laboratório Bennet (U. Michigan).....	26
Capítulo 2.	
Estudos de utilização	28
2.1 Os estudos de utilização e a Ciência da Informação.....	28
2.1.1 Perspetiva Histórica.....	29
2.1.2 O modelo de Tom Wilson (1981).....	30
2.1.3 Desenvolvimento da investigação qualitativa.....	32
2.2 Estudos de utilização aplicados a repositórios digitais.....	34
Capítulo 3.	
Desenho de um perfil de aplicação para dados de investigação	39
3.1 Descrição de dados de investigação.....	39
3.2 Proposta de desenho de perfis de aplicação.....	41
3.3 Gestão do workflow dos dados de investigação.....	42
3.4 Estudo de caso do grupo de investigação no domínio da fratura de materiais.....	43
3.4.1 Perfil de aplicação para experiências de fratura de materiais	44
3.5 Gestão de dados de investigação: Estudo de caso do domínio da química analítica.....	46
3.5.1 Perfil de aplicação para experiências de química analítica	48
3.6 Descrição de dados do domínio da Dinâmica de Estruturas.....	49
3.7 Descrição de dados do domínio da Ecologia Georreferenciada.....	52
3.8 Descrição de dados do domínio da Gravimetria.....	54
Capítulo 4.	
Testes de utilização do Repositório de dados da Universidade do Porto e das plataformas UPBox e DataNotes	58
4.1 Teste de utilização do Repositório de dados da Universidade do Porto.....	62
4.1.1 Resultados do teste de utilização do Repositório de dados da Universidade do Porto....	64
4.1.2 Impressões gerais Repositório de dados da Universidade do Porto.....	79
4.1.3 Recomendações Repositório de dados da Universidade do Porto.....	82
4.2 Teste de utilização das plataformas UPBox e DataNotes.....	83
4.2.1 Resultados do teste de utilização das plataformas UPBox e DataNotes.....	86
4.2.2 Impressões gerais UPBox e DataNotes.....	103
4.2.3 Recomendações UPBox e DataNotes.....	107
Conclusão	109
Referências Bibliográficas	113

ANEXOS	118
Anexo I – Recomendações de esquemas para descrição de dados de investigação (DCMI) .	119
Anexo II – Guião de entrevista (Data Curation Tool Profile).....	123
Anexo III – Guião teste de utilização Repositório de dados da Universidade do Porto.....	129
Anexo IV– Guião de teste de utilização das ferramentas UPBox e DataNotes.....	131
Anexo V – Exemplar de relatório de teste de utilização.....	133

Introdução.

O acesso aos dados provenientes da investigação é uma questão que tem estado na agenda da comunidade científica, independentemente das áreas de investigação onde estes se originam. Os dados resultantes de atividades de investigação têm como particularidade a sua heterogeneidade, sendo que para investigadores de um domínio concreto estes dados exigem uma descrição cuidadosa que diga respeito por exemplo às grandezas envolvidas, às condições de recolha e às condições de acesso por outros.

Um dos argumentos que move a comunidade científica para a necessidade de fomentar o acesso e partilha dos dados recolhidos durante o processo de investigação é a oportunidade de reutilização. Oportunidade que permite, por exemplo, evitar a repetição de procedimentos, contribuindo assim para a agilização dos processos de investigação. Esta dinâmica pode fomentar o interesse das organizações no potencial valor comercial dos produtos resultantes da investigação científica e conduzir ao incremento do investimento privado, ao desenvolvimento de novos serviços, à empregabilidade e à produção de riqueza, para mencionar alguns.

Intimamente relacionada com a questão do acesso e reutilização de dados de investigação surge a preservação digital, sendo um dos maiores desafios no ambiente digital garantir o acesso permanente ao objeto digital, tendo em consideração a evolução tecnológica. Outro problema são as dificuldades de resposta que as organizações têm perante a produção de dados a larga escala. Neste âmbito tende a haver uma maior consciencialização geral para a curadoria de dados, que se entende como o processo de manutenção, preservação e enriquecimento dos dados de investigação ao longo do seu ciclo de vida, muito embora em sentido prático os resultados desta prática tenham, ainda, um impacto ténue.

No caso da Universidade do Porto estão em funcionamento duas plataformas com preocupações de preservação de publicações científicas. Uma é o repositório voltado para a partilha de publicações, o *Repositório Aberto da Universidade do Porto*, e outra o *Repositório Temático da Universidade do Porto*, este orientado para áreas ou públicos específicos. Contudo, ambos os repositórios apenas tornam visíveis os resultados finais das publicações, não tratando uma das componentes vitais da realização de uma investigação, isto é, os dados processados.

Com isto em mente encontra-se a motivação para o desenvolvimento de um repositório multidisciplinar, que armazene, preserve, e permita o acesso e a reutilização dos dados de

investigação. O projeto UPData, que conta na sua equipa com investigadores da Faculdade de Engenharia da Universidade do Porto (FEUP) e da Reitoria da Universidade do Porto, tem como pano de fundo o estudo da aplicabilidade na Universidade do Porto (U. Porto) de práticas de curadoria de dados provenientes da experiência internacional. No repositório experimental estão depositados alguns conjuntos de dados reais, com descrições apropriadas, que podem ser interrogados. Foi também proposto um procedimento de depósito para interação com os investigadores, que requer avaliação junto dos mesmos.

É neste contexto que se vai realizar o trabalho da dissertação, consistindo no estudo de utilização do Repositório de dados da Universidade do Porto, que compreende algumas funcionalidades para servir casos de uso dos utilizadores, e do processo de submissão de dados. O processo de submissão de dados envolve duas ferramentas de gestão de dados integradas, a UPBox para a criação de projetos e o DataNotes para anotação de ficheiros.

Trata-se de um trabalho de cariz individual na exploração de esquemas de metadados e na interação a realizar com os investigadores da U. Porto durante os testes de utilização. Parte do trabalho será desenvolvida em colaboração com outros investigadores a trabalhar nesta área, com o objetivo de estabelecer um procedimento adequado para os investigadores da U. Porto, testando e melhorando o que foi proposto no projeto UPData.

O Repositório de dados da Universidade do Porto encontra-se em fase de protótipo. Assim, o trabalho a realizar poderá ter uma natureza de intervenção no mesmo e nas plataformas propostas para gestão dos dados de investigação, permitindo adaptar as suas funcionalidades de acordo com os resultados do estudo.

Objetivos e resultados esperados

O trabalho a realizar implica a familiarização e acompanhamento da evolução tanto do Repositório de dados da Universidade do Porto, como das ferramentas UPBox e DataNotes. As atividades a desenvolver para alcançar os objetivos determinados são as seguintes:

- Seleção de um conjunto de descritores para dados de investigação, de acordo com as recomendações da DCMI (*Dublin Core and Metadata Initiative*).

Para tal será tida em conta a descrição de dados, com recurso às recomendações propostas pela “*Dublin Core Science and Metadata Community*” - esta comunidade funciona como um fórum, direcionado aos responsáveis pela curadoria de dados, de troca de informação sobre metadados para

descrever dados de investigação. Um dos grandes desafios dos repositórios digitais reside na consolidação das práticas de descrição e na normalização dos descritores a usar, sendo de prever a dificuldade em uniformizar a descrição entre domínios distintos. Entende-se que a descrição normalizada dificilmente cobrirá toda a especificidade de um conjunto de dados, no entanto a descrição de dados a nível conjunto será facilitada através da normalização.

- Desenvolvimento e validação de perfis de aplicação.

Para fazer cumprir este objetivo pretende-se criar um perfil de aplicação específico para cada um dos domínios de investigação dos quais foi possível obter dados. Importa realçar que os descritores apurados estão dependentes da validação por parte de um investigador ligado à produção dos mesmos.

-Contactos com investigadores da U. Porto e observação direta da sua interação com o Repositório de dados da Universidade Porto e com as ferramentas de gestão de dados.

No que diz respeito à interação com os investigadores, pretende-se fazer um trabalho de mediação entre a comunidade científica e a curadoria de dados. Existe um painel de investigadores que colaboraram no arranque da iniciativa UPData cuja contribuição será pedida neste trabalho. A experiência dos investigadores tem de ser tida em consideração para o sucesso deste trabalho. As dúvidas, dificuldades, comentários e sugestões de funcionalidade por parte dos investigadores podem ser capitalizadas em oportunidades para encontrar soluções adequadas para as plataformas a testar. Para suportar esta atividade objetivo elaborou-se um guião para servir de apoio a cada um dos testes de utilização.

O objetivo final a alcançar com a realização deste trabalho será a proposta de melhorias nos procedimentos de gestão e submissão de dados e do Repositório de dados da Universidade do Porto.

Estrutura da Dissertação

Este documento encontra-se estruturado em quatro capítulos distintos:

Capítulo 1 - *Dados de investigação e Repositórios de dados* - corresponde à revisão de literatura sobre a temática dos dados de investigação, realçando em primeiro lugar a crescente produção, o ciclo de vida dos dados e a importância da partilha dos mesmos. Em seguida é feita uma análise à sensibilidade dos investigadores para a partilha e discutida a disseminação aberta dos dados, sendo também abordados conceitos incontornáveis neste contexto, como a preservação e a curadoria de dados. A segunda parte deste capítulo é dedicada ao estado atual dos repositórios digitais, partindo de uma exposição geral para a exposição de casos particular, ou seja, enumeram-se projetos intimamente

ligados aos dados de investigação. Para terminar é apresentado um estudo de caso, de autoria de Akmon et al., realizado em 2011, relativo à gestão de dados num laboratório da Universidade do Michigan.

Capítulo 2 - Estudos de Utilização – apresenta a revisão da literatura referente aos estudos de utilização. Começa por enquadrar esta temática na área da Ciência da Informação, salientando a relevância deste tipo de estudos para conhecer as necessidades dos utilizadores. Neste sentido é apresentada uma perspetiva histórica dos estudos de utilização e faz-se o destaque para as principais teorias de ordem qualitativa que surgiram neste campo. Depois estreitasse a análise aos estudos de utilização aplicados a repositórios digitais, sendo feito um levantamento de trabalhos de vários autores de modo a recolher as principais técnicas que por norma são utilizadas neste tipo de estudo, para assim justificar a adoção das técnicas ao processo metodológico desta investigação.

Capítulo 3 - Desenho de um perfil de aplicação para dados de investigação - este capítulo apresenta dois estudos de casos da experiência realizada junto de investigadores das áreas da Engenharia Mecânica e Química, da FEUP, dos quais resultaram perfis de aplicação para domínios específicos destas áreas de investigação. Para além destes são sugeridos perfis de aplicação para a descrição de dados dos domínios da Dinâmica de Estruturas, da Ecologia Georreferenciada e da Gravimetria.

Capítulo 4 - Teste de utilização do Repositório de dados da Universidade do Porto e das ferramentas UPBox e DataNotes - este capítulo é dedicado a descrição da experiência realizada junto dos utilizadores. Assim são apresentadas genericamente as ferramentas a testar, o Repositório de dados da Universidade do Porto, a UPBox e o DataNotes. Em primeiro lugar são discriminados os casos de uso que influenciaram a proposta de tarefas a realizar, os participantes, o processo de recolha de dados e as variáveis do estudo. Para facilitar a leitura do documento e evitar redundância na informação transmitida as funcionalidades das plataformas são descritas juntamente com a proposta das tarefas, sendo acompanhadas por um figura relacionada, uma tabela com os resultados obtidos, uma descrição breve da interação e as sugestões recolhidas. Após cada um dos testes de utilização é feito um levantamento das impressões gerais dos participantes e, por fim, é apresentada uma tabela que cobre todas as recomendações que foram possíveis obter com o intuito de melhorar as plataformas.

Este trabalho termina com uma reflexão global à experiência realizada, num capítulo dedicado às conclusões. Em anexo pode ser consultada a documentação produzida para apoiar este trabalho.

Capítulo 1.

Dados de investigação e Repositórios de dados

1.1 Dados de investigação: produção e partilha

Os dados de investigação são um aspeto crítico na ciência, representam a sua infraestrutura na medida em que estabelecem as bases para a tomada de decisão científica. A *National Science Foundation* (NSF)¹, entidade promotora do desenvolvimento científico, adota a definição de dados de investigação como - "*the recorded factual material commonly accepted in the scientific community as necessary to validate research findings*". Os dados de investigação podem ser classificados como dados brutos (*raw data*), geralmente por processar, ou dados processados, que dizem respeito a dados em bruto que já passaram por um processo de manipulação. Estes últimos são mais refinados e perceptíveis para outras pessoas, como é o caso de dados organizados numa tabela (Lauriault et al., 2007). O valor dos dados de investigação pode estar dependente da sua estrutura e organização, e nesta perspetiva os conjunto de dados (*datasets*), correspondentes à agregação de dados relacionados entre si e registados no mesmo formato, são essenciais. De uma forma geral os dados de investigação são tudo aquilo que um investigador necessitará para reproduzir resultados publicados.²

Os dados de investigação podem ainda ser agrupados de acordo com a disciplina que os produz e em última instância os utiliza. Este tipo de agrupação pode ser útil para o entendimento das semelhanças e diferenças, assim como para o potencial uso dos dados de investigação ao longo do tempo (Willis, Greenberg and White, 2012). A National Science Board (2005) divide os dados de investigação em três grupos mediante a sua origem: **observacionais** – que não podem ser recolhidos novamente e são arquivados indefinidamente (ecologia, ciências sociais, biologia); **computacionais** – resultantes de modelos computacionais ou simulações que podem ser recapturados com informação adequada sobre hardware, software e inputs; **experimentais** – por norma reproduzíveis, embora por vezes as condições da experiência ou variáveis possam ser desconhecidas (física, química).

A produção de dados de investigação é dispendiosa e, por norma, constitui interesse para a comunidade científica a longo prazo. Quando produzidos em determinado espaço ou tempo, como no caso das observações ecológicas, tornam-se insubstituíveis (Borgman et al., 2007). Os dados de

¹ National Science Foundation. Disponível em: <http://www.lib.umich.edu/research-data-management-and-publishing-support/nsf-data-management-plans> Consultado a 27 de outubro de 2012

² Disponível em: <http://www.lib.umich.edu/research-data-management-and-publishing-support/nsf-data-management-plans>; Consultado a 9 de abril de 2013

investigação não representam somente um *output* da investigação, mas também providenciam *inputs* para novas hipóteses e replicação de estudos, isto é, abrem caminho para novo conhecimento científico e conduzem à inovação.

Atualmente a produção de dados de investigação tem atingido um volume sem precedentes, assistindo-se a uma mudança de paradigma, à qual Hey et al. (2009) designam como “*the four paradigm: data intensive scientific discovery*”, em que o conceito intensidade é chave. Para além da intensidade, a *NSF*³ verifica que a colaboração entre investigadores é também um aspeto em crescimento. A literatura tende a adotar o termo “dilúvio”, no inglês “*deluge*”, para se referir à produção de dados em larga escala. Embora o crescimento da produção não possa ser contestado, Borgman et al. (2007) fazem notar que o dilúvio de dados está a ocorrer, não em absoluto, mas de forma relativa. Estes autores consideram que enquanto para as “*big sciences*”, como a física e a astronomia, se tem vindo a formar um conjunto de ferramentas e repositórios para lidar com a intensa produção, as ciências ditas pequenas, dependentes do trabalho de campo, carecem de infraestruturas capazes de gerir a quantidade crescente de dados. A falta de um *framework* integrado para gerir este tipo de dados de investigação representa “*uma barreira significativa não apenas para aqueles cientistas que conduzem a investigação, mas também para aqueles, que subsequentemente a reutilizarão*” (Borgman et al., 2007).

A partilha de dados é uma das componentes de validação do método científico, permitindo a verificação dos resultados e um alargamento da investigação com base em resultados prévios. Num ambiente de aumento exponencial de dados a colaboração é premente. A partilha engloba o depósito e preservação dos dados, mas o principal desígnio é potenciar o acesso e reutilização dos mesmos. O incentivo para a partilha e reutilização de dados começa nas boas práticas de gestão em todas as fases do ciclo de vida dos dados (Tenopir et al., 2011). Em consideração a esta circunstância o *Joint Information Systems Committee* (JISC)⁴, propôs um *framework* do ciclo de vida dos dados (fig.1) integrado no processo de investigação, proporcionando um serviço de apoio aos investigadores em cada uma das etapas. Através do *Virtual Research Environment (VRE)*, o JISC facultava ferramentas valiosas para a colaboração entre equipas de instituições distantes ou mesmo de países diferentes.

O ciclo de vida dos dados de acordo com o *framework* do JISC⁵, tem origem no desencadeamento do processo de investigação, e não pode ser considerado à parte do ciclo de vida da investigação, já que os

³ National Science Foundation. Press Release 10-077 (2010). Disponível em: http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928; Consultado a 27 de outubro de 2012

⁴ JISC. Joint Information System Committee. Disponível em: <http://www.jisc.ac.uk/> Consultado a 05 de novembro de 2012

⁵ JISC. How JISC is helping researchers. Disponível em: <http://www.jisc.ac.uk/whatwedo/campaigns/res3/jischelp.aspx> Consultado a 05 de novembro de 2012

dados são indispensáveis à investigação científica. As etapas de produção de dados propostas pelo JISC contemplam: a **experimentação, simulação** ou **observação** - que difere substancialmente entre disciplinas, mesmo entre disciplinas do mesmo domínio; a **gestão de dados** - exigindo-se dos investigadores um plano de gestão dos dados; a **análise de dados** - podendo esta ser uma das componentes de um trabalho de investigação, nomeadamente nas ciências sociais em que se procede a investigação por questionário; a **partilha de dados** - considerando a preservação e curadoria dos dados para permitir o acesso e reutilização por outros investigadores.

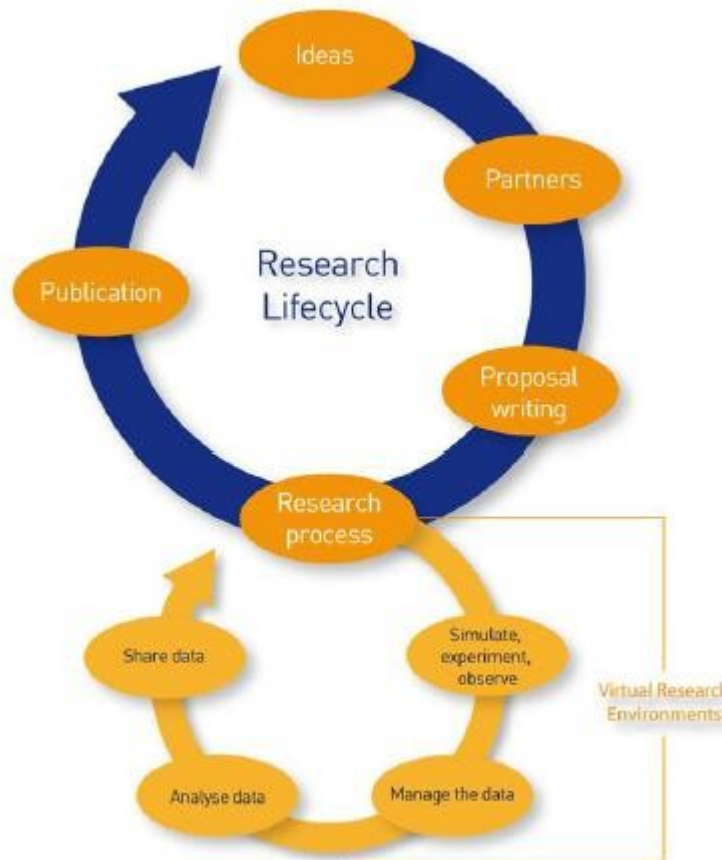


Figura 1: Etapas do ciclo de vida da investigação e dados (JISC)

Os dados são elementos preponderantes na investigação e no desenvolvimento de novos métodos, técnicas de análise e implementação de *software*. A larga comunidade científica beneficia, atendendo que a partilha de dados encoraja novas perspetivas, ao possibilitar diferentes interpretações e abordagens que conduzem ao progresso científico, nomeadamente quando essa partilha se desenrola num ambiente interdisciplinar (Tenopir et al., 2011). Outra das vantagens da disseminação dos dados é a oportunidade da sua reutilização, o que vai permitir uma maior verificação dos resultados, sendo este um aspeto crítico da ciência. A reutilização ajuda a reconhecer erros, desencoraja a fraude e promove o uso eficiente do financiamento ao evitar a repetição de procedimentos (Piwowar et al.,

2007). Para além disso, a replicação de resultados funciona como uma boa base prática para formar novos investigadores. Em suma, a partilha de dados facilita a ciência.

A disseminação aberta e o uso irrestrito são atributos essenciais de incremento aos sistemas atuais de produção científica e inovação tecnológica. Reconhecer o valor dos dados digitais é fundamental na cadeia de valor da ciência e desbloqueará mecanismos de financiamento público e privado. Em virtude de obter o máximo retorno desse investimento, o acesso, uso, gestão, e preservação devem ser uma parte integral do processo de investigação. Perante isto, os dados não podem ser entendidos como um subproduto. Em muitos casos, os dados têm valor, para além de servir o propósito para o qual foram recolhidos.

1.1.1 Os investigadores e a partilha dos dados de investigação

Assume-se que a partilha de dados é uma das componentes de validação do método científico, permitindo a verificação de resultados e um alargamento da investigação com base em resultados prévios. A literatura sobre esta temática tem discutido sobretudo as questões de motivação de partilha *versus* sonegação dos dados, isto é, o posicionamento dos investigadores face à divulgação dos dados. Nesse sentido Tenopir et al., (2011), com base num inquérito realizado a nível internacional, concluíram que estão enraizadas na cultura da comunidade científica barreiras para uma efetiva partilha dos dados. Contudo bons exemplos provêm das áreas da astronomia, geofísica e biodiversidade. A ponderação entre o que se pode ganhar ou perder com a partilha é uma inquietação que tem vindo a perturbar os cientistas ao longo dos séculos, e continua a prevalecer mesmo com o surgimento de projetos que formalizam a correta citação de dados (Pfeiffenberger e Carlson, 2011).

Entre as principais razões que motivam os investigadores a reter dados estão a falta de tempo e suporte institucional para a gestão de dados a curto ou longo prazo (Arzberger et al., 2004), os interesses comerciais (Akmon, 2011), a confidencialidade em relação aos sujeitos, isto é, a privacidade (bem patente no caso da medicina), futuras oportunidades de publicação e a retenção da exclusividade sobre os dados que levaram bastante tempo a produzir (Savage e Vickers, 2009). De acordo com um estudo de Wicherts et al. (2011) há uma motivação mais perversa para explicar a relutância de alguns autores para partilhar os dados que produzem. Segundo este estudo, que teve como referência a análise de dados de natureza estatística, os investigadores receiam que dispor os seus dados ao escrutínio de pares possa expor erros ou produzir conclusões que contradigam as próprias. Wicherts et al. (2011) associam que a deteção de erros e a partilha de dados após a publicação de resultados pode refletir diferenças no rigor com que os investigadores gerem os seus dados, isto porque os

investigadores que aplicam maior diligência no arquivamento e gestão dos dados tendem a cometer menos erros. Savage and Vickers (2009), mediante as recusas dos investigadores em ceder conjuntos de dados, sob a justificação que preparar os dados daria demasiado trabalho, sugerem que durante a produção não se desenvolvem conjuntos de dados devidamente anotados.

A partilha de dados é benéfica para a comunidade científica, no entanto esses benefícios são menos óbvios para quem produz e torna os dados disponíveis. Uma investigação é um processo que consome vários recursos, à medida que a difusão de conteúdos seja facilitada e os benefícios se tornem verificáveis para o investigador, existe a esperança que os autores se tornem mais aptos a partilhar os dados dos seus estudos e, portanto, a maximizar a sua utilidade para a sociedade.

As condições que à partida os investigadores assumem para a partilha dos seus dados são a correta citação. A maioria dos inquiridos por Tenopir et al. (2011) estão disponíveis a partilhar informação e a reutilizar os dados de outrem, se obtiverem crédito via citação, se obtiverem uma cópia dos artigos que usam os seus dados e o conhecimento de produtos ou publicações que se desenvolveram a partir da sua contribuição. Um dos aspetos que os investigadores mais valorizam é o número de vezes que o seu trabalho é citado, retirando daqui reputação, o que os ajuda a conseguir financiamento para novas investigações e pode implicar um aumento remuneratório. Impulsionar o rácio de citação pode funcionar como uma boa motivação para os autores. Piwowar et al. (2007) procuraram estabelecer uma relação entre a disponibilização de dados e o fator de impacto, chegando à conclusão que os investigadores que divulgam os seus dados são mais citados do que aqueles que não o fazem.

Acreditando que o benefício da partilha de dados prevalece sobre os custos da mesma, muitas iniciativas encorajam ativamente os investigadores a tornarem os dados disponíveis. Alguns periódicos, entre eles a cadeia *PloS ONE*, requerem a submissão de dados detalhados em bases de dados acessíveis publicamente, como condição de publicação (Piwowar et al., 2007). Havendo, também, periódicos que sugerem um guião de incentivo à disponibilização dos conjuntos de dados. Contudo, as políticas explícitas dos periódicos para a partilha dos dados de investigação não leva a que os autores publicados nessas revistas partilhem os conjuntos de dados com outros investigadores independentes (Savage e Vickers, 2009). O mais curioso nesta observação de Savage e Vickers é que foi feita através da análise de publicações em jornais de acesso aberto, o que por si só já revela uma aproximação dos investigadores a políticas de abertura.

1.1.2 O acesso aberto aos dados de investigação

O crescente movimento *Open Access* tem vindo a colocar pressão para a partilha entre pares (Piwowar et al., 2007), funcionando como um incentivo ao acesso aberto à informação científica, está na génese de projetos que se estendem aos dados científicos. Contudo, uma avaliação da Comissão Europeia (2010) revelou uma falta de conhecimento e entendimento dos investigadores relativamente ao acesso aberto. Estes não estão propriamente relacionados com o conceito, ou mesmo que tenham algum conhecimento não demonstram dominar a problemática, especialmente no que diz respeito ao controlo de qualidade, ao papel dos repositórios digitais e aos direitos de autor. São também ainda pouco sensíveis à questão da partilha, alguns demonstram mesmo resistência a esta cultura de abertura. Resistência acentuada quando se trata do protecionismo aos dados de investigação. Os investigadores pouco informados sobre o acesso aberto pensam estar a cometer uma infração aos direitos de autor por auto arquivarem o seu trabalho num repositório, acreditam que o seu trabalho não é exposto à revisão por pares, ou melhor, que não existe controlo de qualidade ou padrões de exigência elevada entre os repositórios. Todavia, sem refutar estas considerações, é preciso ter presente que os investigadores podem estar condicionados por restrições à divulgação dos seus dados em acesso aberto, muitos deles sujeitos a períodos de embargo.

Com efeito, vários cientistas consideram a partilha de dados importante mas, segundo um estudo da *Publishing Research Consortium* (in Tenopir et al., 2011), apenas uma pequena percentagem (38 por cento) considera o acesso aos dados de investigação fácil, uma percentagem que é inferior aos níveis de acesso a outros tipos de comunicação científica, como trabalhos de referência, informação técnica e registo de patentes.

O debate em termos de acesso aberto tem sido feito sobretudo em torno da literatura científica, no entanto à medida que a produção de dados de investigação se tornou intensiva, o conceito *Open Access* desdobrou-se, designando-se por *Open Data* quando corresponde a uma filosofia de abertura de acesso, gratuito e sem restrições de copyright, a um conjunto de dados de investigação. A semelhança do acesso aberto o movimento em prol do *Open Data* estimula a partilha, tendo como princípio o estabelecimento de conexões que fomentem o avanço científico e a aprendizagem⁶. Stevan Harnad, um dos mais conceituados defensores do acesso aberto, discutiu na conferência “*The Open Access Open Data?*” realizada em Colónia no ano de 2010, as divergências e convergências entre os conceitos. Para Harnad (cit. por Giglia, 2011) deve ser conferido aos investigadores o direito à manutenção e exclusividade dos dados até que estes sejam publicados, salientando que o *Open Access* deve ser mandatado enquanto o *Open Data* não, opinião reforçada por Malcolm Read (JISC) na

⁶ Databib; Disponível em: <http://databib.org/connect.php>. Consultado a 2 de janeiro de 2013

mesma conferência. Isto porque apesar do acesso aos dados de investigação ser desejável não é o foco do movimento *Open Access*. Contudo, Harnad, apologista da “*Green Road*” para o auto-arquivamento, assegura que o crescimento da abertura aos dados de investigação será paralelo ao aumento dos mandatos *Open Access*, devido à sua natureza complementar e à dinâmica colaboração global para acelerar o progresso científico (cit. por Giglia, 2011).

As consequências do acentuado interesse no acesso aberto aos dados de investigação começam a sentir-se na prática, com um maior número de projetos, sem fins lucrativos, a registarem-se nos últimos anos. Casos disso são a formalização do “*The Open Data Foundation*”⁷, que confere espaço para a aproximação de membros de diferentes comunidades para o trabalho conjunto no alinhamento de tecnologias standard e *software* para facilitar a visibilidade e utilização dos dados, e da “*Open Knowledge Foundation*”⁸ (OKF), que se apresenta como uma “*globally-located, community-focused organisation*”, dedicada à criação, partilha e aplicação do conhecimento aberto na era digital, com particular enfoque na valorização dos dados de investigação. Para além destes, convém referir o *DataCite*⁹, projeto que arrancou em 2009, tratando-se de um consórcio internacional entre bibliotecas especializadas em investigação e centros de informação conceituados, que procura fornecer um serviço que aumente a aceitação dos dados de investigação como legítimos, através de um processo que procura instituir condições para a correta citação dos dados.

As políticas de acesso aberto aos dados de investigação estão sub-desenvolvidas comparativamente às das publicações, no entanto a previsão é de um alargamento “*of the general concern for unlocking the full value of scientific data*” (Comissão Europeia, 2011). Tal como é referido no Relatório da Comissão Europeia (2011), respeitante às políticas nacionais de acesso aberto, é exetável um crescimento da importância dos dados de investigação motivado pela integração conjunta das publicações e dos dados a si associados em novos serviços de informação, aliada à aplicação de técnicas inovadoras de extração de dados, como por exemplo o *data mining*. Neste Relatório pode ser lido que várias agências de financiamento de investigação estão a definir políticas que cobrem a acessibilidade aos dados criados durante os trabalhos que financiam, continuando a surgir novas abordagens ligadas às infraestruturas *e-science* e às questões relevantes do ponto de vista dos direitos de propriedade intelectual.

Associada a esta preocupação com o enquadramento legal do acesso aberto aos dados de investigação, a OKF administra desde 2009 a iniciativa *Open Data Commons*¹⁰ que providência soluções legais neste domínio. Esta iniciativa reúne uma série de licenças que promovem os direitos de acesso e a reutilização dos dados de investigação, tornando-se pioneira em matéria de licenciamento *Open Data*

⁷ The Open Data Foundation. Disponível em: <http://www.opendatafoundation.org/> Consultado a 03 de janeiro de 2013

⁸ Open Knowledge Foundation. Disponível em: <http://okfn.org/> Consultado a 23 de dezembro de 2012

⁹ DataCite. Disponível em: <http://datacite.org/whatdowedo> Consultado a 03 de janeiro de 2013

¹⁰ Open Data Commons. Disponível em: <http://opendatacommons.org/licenses/>. Consultado a 23 de dezembro de 2012

quando em Março de 2008 lançou a *Public Domain Dedication and License (PDDL)*. Outro passo neste sentido foi dado em 2009, com a introdução das *Panton Principles*¹¹, que reivindicam que o estatuto jurídico dos dados de investigação seja explícito e a instituição dos mesmos ao domínio público.

A legitimidade dos dados de investigação está bem enquadrada na plataforma *figshare*¹², em que todos os conteúdos assentam na licença *Creative Commons*, o que significa que o uso dos dados de outrem implica o reconhecimento a quem os produziu originalmente. O que na opinião de Mark Hahnel (2012)¹³, gestor da *figshare*, faz com que a cedência de dados de investigação num ambiente que promova o acesso, seja uma boa maneira de obter crédito e visibilidade científica.

O acesso e partilha de dados não é apenas uma questão técnica, mas também um processo social complexo no qual os investigadores têm de ponderar entre diversos interesses e pressões. Abordagens puramente regularizadas à partilha de dados não serão bem sucedidas sem atender à consideração desses fatores. Várias abordagens ao acesso e partilha de dados são, portanto, necessárias, incluindo o estabelecimento de regulamentos e incentivos às boas práticas. A reutilização é argumento central do movimento *Open Data*, traduzindo-se no uso dos dados, normalmente sem permissão explícita, noutro estudo previsto ou não pelo autor original. A reutilização têm como atributos a possibilidade de conjugação com dados de outra proveniência para formar novo conhecimento e fomentar a produção de mais e melhores dados, apesar da qualidade e formato dos dados não ser uma prioridade deste movimento (Murray-Rust, 2008). Por isso mesmo a reutilização deve ser promovida e não restrita, sobretudo quando os dados são produzidos sob financiamento público (Arzberger et al., 2004). Na opinião de Malcolm Read (cit. por Giglia, 2011) o futuro da *Open Data* passa pela consideração de conceitos como a integridade, reutilização, preservação a longo prazo, confiança e sustentabilidade. Malcolm Read acrescenta que as principais preocupações manifestadas pelos investigadores, para além das restrições legais e receio do uso inadequado, residem na falta de experiência em curadoria, na criação de metadados e preservação dos dados de investigação.

1.2 A preservação de dados de investigação

Estamos numa era onde os conteúdos digitais são a forma dominante pela qual se cria, molda e partilha informação. Governos, empresas, organizações de investigação, bibliotecas e indivíduos

¹¹ Paton Principles. Disponível em: <http://pantonprinciples.org/> Consultado a 04 de janeiro de 2013

¹² Figshare. Disponível em: <http://figshare.com/> Consultado a 04 de janeiro de 2013

¹³ Data Curation Centre. 2012. Managing Research Data (video) Disponível em: <http://www.dcc.ac.uk/news/managing-research-data-video> Consultado a 04 de janeiro de 2012

tornaram-se dependentes da informação digital. Na sua maioria os ativos da base do conhecimento institucional e intelectual estão em suporte digital. Contudo, existe a plena noção que este conhecimento pode estar em risco, sendo fracas as garantias de acesso futuro devido à dependência para com infraestruturas tecnológicas que tendem a evoluir a um ritmo alucinante, e que fazem dos recursos digitais menos permanentes que os seus equivalentes em papel (Smit et al., 2011). Neste sentido a preservação é um dos problemas imediatos dos recursos digitais.

Os custos da preservação e arquivamento de dados são relativamente menores quando comparados à aquisição de registos científicos através da observação (Lauriault et al., 2007). Anderson e Mandelbaum (cit. por Chowdury, 2010) afirmam que a preservação, outrora uma preocupação exclusiva das bibliotecas, é agora uma preocupação universal. A preservação dos dados é nos dias de hoje uma questão inquietante no seio da comunidade científica, sobretudo devido à constante mutação tecnológica. Paul Conwey (cit. por Li e Banach, 2011) expressa essa preocupação - *"as our capacity to record information has increased exponentially over time, the longevity of the media used to store the information has decreased equivalently"*.

A obsolescência digital é tida como uma séria ameaça, tornando vários documentos ilegíveis devido à desatualização dos formatos que os suportam. Simplesmente, tudo o que é digital é frágil e passível de cair em desuso (Smit et al., 2011). À medida que o mundo da investigação se torna cada mais digital vão ser impostos novos desafios à longevidade e recuperação da informação científica, sendo certo que o advento da ciência está dependente do conhecimento gerado no passado. O que é estranho é a noção geral que, uma vez armazenada num suporte digital, a informação está salvaguardada.

Os dados digitais raramente são arquivados e preservados devidamente, em consequência muitos se perdem, alguns permanentemente (Lauriault et al., 2007). Os cientistas tendem a armazenar os dados com pouca documentação ou informação adicional, fazendo poucos esforços via preservação, para além de *back-ups* nos computadores pessoais, *pen-drives*, *cd-roms*, entre outros suportes. Este tipo de estratégia dificilmente lida eficazmente com a produção de um grande volume de dados, uma vez que se trata de um tipo de abordagem local que não garante o acesso e uso futuro dos dados, comprometendo a interpretação futura, mesmo para quem os produziu. A preservação a longo prazo, quando bem gerida, ajuda a manter a integridade dos dados. A longo prazo sugere uma extensão temporal indefinida, mas que deve ser longa o suficiente para lidar com os problemas que envolvem a evolução tecnológica, incluindo o suporte para novos media e formatos de dados, ou uma mudança na comunidade de utilizadores.

Smit et al. (2011) consideram que entre os periódicos tradicionais a questão da preservação está

devidamente salvaguarda por estratégias bem delineadas, no entanto no contexto de investigação ao nível da preservação de dados a situação não é favorável. O dilúvio de dados, que se espera continuar a emergir, não somente em quantidade mas também em múltiplos formatos representa um grande desafio. A preservação digital é um conceito que gera confusão, mesmo entre pessoas envolvidas na custódia de informação, que geralmente a confundem com conceitos próximos, ou focam-se em aspetos particulares da preservação digital e assumem-nos como um todo (Smith et al., 2011). É o caso de algumas instituições que lidam com o armazenamento a longo prazo que entendem a preservação como armazenamento. Embora o armazenamento seja um elemento de boa preservação, estes conceitos não devem ser entendidos como sinónimos. O armazenamento a longo prazo, por si só, não garante que os dados possam ser lidos e interpretados no futuro. Como os sistemas são renovados e os formatos mudam, os documentos e o conhecimento tácito sobre os mesmos é perdido.

Podem ser verificadas na literatura várias estratégias para lidar com a preservação digital, sobretudo desenvolvidas sob a perspetiva das Tecnologias de Informação. De acordo com Smith et al. (2009), a preservação estará mais perto de ser conseguida se forem atendidos os seguintes aspetos:

- a) Normalização/ Standardização** - assegurar sistematicamente que o formato e estrutura dos objetos é a mesma;
- b) Refrescamento** - proceder à transferência da informação de um suporte de armazenamento físico para outro, antes da sua obsolescência;
- c) Migração (Transformação)** - correspondente à conversão, ou transferência periódica de uma configuração de *hardware/software* para outra, que acompanhe a evolução da tecnologia para as gerações posteriores;
- d) Emulação** - a abordagem ao problema da obsolescência é feita de outra forma, neste caso com a utilização de *software* que reproduza o comportamento de uma plataforma numa outra, que à partida não seria compatível;
- e) Metadados semânticos** - manter a descrição da informação relativa ao objeto e assegurar a sua compreensão pela comunidade alvo no futuro. A semântica é geralmente independente ao formato;
- f) Combinação** - tal como o termo indica, corresponde a conjugação dos esforços já mencionados. No caso dos dados científicos, a emulação terá menor destaque, uma vez que os investigadores preferem usar os sistemas de análise mais recentes.

Paradoxalmente, a preservação pode estar em risco se toda a gente, em todo o lado, fizer exatamente a mesma coisa. Smit et. al. (2011) consideram que para enfrentar o futuro desconhecido da preservação, os vários agentes envolvidos devem adaptar estratégias alternativas, admitindo que não se podem voltar cada um para seu lado.

Tendo em consideração que a problemática da preservação digital será melhor sucedida, se enquadrada num ambiente de cooperação, coordenação e consistência entre os agentes que lidam com esta situação, podem ser identificados alguns programas internacionais a promover essa sinergia. No espaço europeu destaca-se o *Digital Preservation Europe (DPE)*¹⁴, enquanto nos EUA foi lançado, em 2000, o *National Digital Information and Infrastructure Preservation Program (NDIIPP)*. Já o *One Planets Foundation (OPN)*¹⁵ procura oferecer ferramentas e metodologias que prescrevam melhores práticas para o desenvolvimento de uma visão e abordagem holística na preservação de informação digital. Tal iniciativa pode incutir nas universidades a responsabilidade em preservar conjunto de dados no seu raio de atuação. Outro programa, o *Keeping Research Data Safe (KRDS)*, financiado pelo JISC,¹⁶ desenvolve desde 2008 um conjunto de projetos para a preservação a longo termo de dados de investigação ao nível institucional e nacional.

Um dos modelos mais adotados para garantir o sucesso da preservação digital, é o modelo de referência OAIS (*Open Archival Information System*)¹⁷. Publicado pelo *Consultative Committee for Space Data Systems (CCSDS)*, núcleo composto por agências espaciais, incluindo a NASA, o OAIS foi aprovado como modelo *standard* pela *International Organization for Standardization (ISO)*, em 2003 (Ball, 2006). Um dos pontos fortes do OAIS é o facto de fornecer linhas de orientação relativamente a várias funções de preservação, através da representação do *workflow* dos dados a partir do momento em que são inseridos num determinado sistema, o que implica a ingestão, gestão, acesso, armazenamento e disseminação (fig.1). Simultaneamente recomenda algumas sequências de passos a seguir aquando da migração de informação digital para novos formatos ou suportes e recorre a modelos de dados úteis na descrição da informação (CCSDS, 2012), funções referidas como requisito de boa preservação.

¹⁴ Digital Preservation Europe. Disponível em: <http://www.digitalpreservationeurope.eu/> Consultado a 4 de dezembro de 2012

¹⁵ One Planets Foundation. Disponível em: <http://www.openplanetsfoundation.org/about> Consultado a 4 de dezembro de 2012

¹⁶ Keep Research Data Safe. JISC. (2008). Disponível em: <http://www.jisc.ac.uk/publications/reports/2008/keepingresearchdatasafe.aspx> Consultado a 04 de dezembro de 2012

¹⁷ O termo “open” não implica que o acesso ao arquivo seja totalmente aberto. Neste caso significa que as normas e recomendações associadas são desenvolvidas em fóruns abertos.

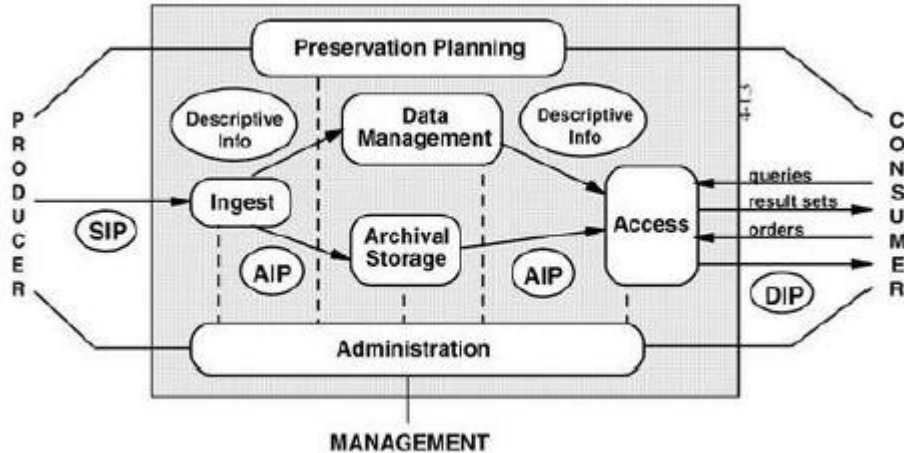


Figura 2: Entidades Funcionais OAIS (CCSDS, 2012)

Este modelo pode ser aplicado a várias realidades em que a tónica seja tornar os dados de investigação acessíveis e compreensíveis a longo prazo, podendo ser encarado como uma estrutura robusta para padronizar as práticas de preservação, uma vez que está pensado para lidar com questões de interoperabilidade entre sistemas.

Apesar do OAIS ser importante do ponto de vista da preservação, é um modelo que não contempla o que acontece antes da ingestão dos dados nos sistemas. É neste sentido que a curadoria de dados fornece um contributo determinante ao *workflow* dos dados digitais, ao introduzir a noção do ciclo de vida dos dados como essencial à gestão dos dados digitais (Ray, 2012).

1.3 Curadoria de dados

A curadoria de dados é uma prática que tem vindo a ser considerada nos últimos anos para responder aos desafios impostos pelo dilúvio de dados, entre os quais a falta de infraestruturas capazes de os armazenar. O termo curadoria digital, aplicado à gestão e preservação de um vasto raio de dados de investigação, surgiu pela primeira vez num relatório de um grupo de trabalho reunido pela JISC, em 2002. Embora a determinação do conceito não tenha sido consensual, em virtude da sua adoção como título profissional pela comunidade bioinformática¹⁸, foi desde então considerado como novo

¹⁸ O termo “curador” é aplicado nesta área referente ao profissional que identifica erros e fornece anotações às bases de dados de sequências de proteínas (*protein sequences databases*), ou outras.

neste contexto e unanimemente reconhecida a existência de áreas suficientemente genéricas e interdisciplinares às quais o fornecimento de um serviço de curadoria seria útil. Na visão da *Digital Data Curation Task Force (DDCTF)*, de 2003, a curadoria de dados devia abordar princípios transversais a várias disciplinas e desempenhar funções para além da preservação, como o planeamento, avaliação, valorização, gestão ativa, acesso, e manutenção da informação de origem dos dados de investigação. A DDCTF preconizava, também, a necessidade de investigação científica em curadoria para o desenvolvimento da prática (cit. por Ray, 2012). Características que se mantêm atualmente.

A curadoria de dados é definida pelo *Digital Curation Centre (DCC)*, centro de investigação líder mundial em curadoria de informação digital, como um processo que “*envolve a manutenção, preservação e enriquecimento dos dados de investigação ao longo do seu ciclo de vida*”¹⁹. A mesma entidade argumenta que a ativa gestão dos dados de investigação reduz o risco de obsolescência digital. Por outro lado, a curadoria de dados está orientada para a redução da duplicação de dados, objetivando uma pesquisa de maior qualidade.

A curadoria digital ocorre em paralelo com a preservação de dados. O próprio ciclo de vida da curadoria de dados engloba um conjunto de procedimentos considerados como adequados à preservação. O DCC explica que o ciclo de vida da curadoria digital inclui as seguintes etapas: **Conceptualização** – conceber e planear a criação de objetos digitais, incluindo métodos de captura e retenção; **Criação** – produzir objetos digitais, atribuindo-lhes metadados estruturais, descritivos e administrativos; **Acesso e uso** – garantir o fácil acesso aos objetos, sejam eles de acesso público ou protegidos por uma *password*; **Avaliar e selecionar** – identificar quais os objetos que requerem curadoria e preservação a longo termo; **Descartar** – limpar dos sistemas objetos tidos como não pertinentes, podendo exigir a destruição segura dos mesmos de acordo com requisitos e políticas legais; **Ingestão** – transferência dos documentos para plataformas de confiança, mais uma vez tendo em conta os requisitos legais; **Ação de preservação** – tomar medidas que assegurem a preservação a longo prazo e a retenção da natureza dos dados.

O ciclo de vida da curadoria de dados prevê ainda a **reavaliação dos dados** que numa primeira instância não foram considerados como válidos, o **armazenamento** seguro dos dados em consonância com *standards* relevantes, a **reutilização** e a **transformação (migração)**, prevista nas estratégias de preservação.

O processo de curadoria implica, preferencialmente, a participação de dois grupos de intervenientes

¹⁹ The Digital Curation Centre (2012). Disponível em: <http://www.dcc.ac.uk/digital-curation/what-digital-curation>
Consultado a: 10 de outubro de 2012

no decurso da submissão de dados num repositório, sendo estes os investigadores e os curadores. Para o processo ter maior probabilidade de sucesso, compete aos investigadores a publicação do seu trabalho, com particular atenção à anotação e organização dos dados. Por outro lado, e como a preservação é da competência dos curadores, estes devem proceder à acomodação dos dados de investigação nos repositórios, garantido a correta indexação e classificação para que os dados sejam facilmente recuperados pelos utilizadores através da Internet (Buneman, 2004). O objetivo desta parceria é minimizar os problemas associados ao processo de curadoria.

Lyz Lyon (2010), investigadora associada ao DCC, fez uma revisão comparativa entre as diferenças disciplinares na área da curadoria, com base em dezasseis estudos de caso, chegando à conclusão que é vital desenvolver estratégias alternativas para agregar as diferentes disciplinas, uma vez que as abordagens tradicionais não vão conseguir responder às necessidades dos investigadores das várias áreas. Esta observação de Lyz Lyon reflete em grande medida um dos desafios apontados às universidades que procuram estabelecer infraestruturas de curadoria de dados de investigação.

Neste contexto o JISC encara a curadoria digital com particular atenção à preservação e interoperabilidade desde o início do ciclo de vida dos dados. Esta abordagem procura contrariar a tendência geral dos criadores de apenas se preocuparem com as questões relacionadas com a preservação dos dados no final do ciclo, o que dificulta o processo de preservação a longo prazo, já que as decisões de maior pertinência na gestão de dados devem ser consideradas atempadamente. Os princípios de integridade, autenticidade e proveniência dos dados têm vindo a ser incorporados nas políticas e práticas dos repositórios digitais (Ray, 2012).

1.4 Repositórios digitais

A produção científica, para além de intensa e colaborativa, está também fortemente instrumentalizada. Os repositórios digitais providenciam oportunidades de acesso, captura, e gestão do dilúvio de dados (Wallis et al., 2010), resultantes do processo de investigação. De acordo com Thomas e McDonald (2007) um repositório digital é um sistema em rede que presta um serviço em torno de uma coleção de objetos digitais. A categorização de um repositório é problemática, uma vez que os contextos em que estes se desenvolvem varia imenso, podendo ser categorizados de acordo com a entidade que os financia ou administra. As duas principais categorias na sua classificação são a instituição ou disciplina.

O projeto *OpenDoar (The Directory of Open Access Repositories)*, atualmente financiado pelo JISC,

cotado pela John Hopkins University como serviço líder no panorama dos repositórios digitais e considerado recurso chave pela comunidade *Open Access*, dispõe uma listagem abrangente de repositórios de todo o mundo permitindo o acesso ao conteúdo dos mesmos, fornecendo simultaneamente ferramentas de apoio à atividade dos responsáveis por repositórios²⁰. Segundo dados do *OpenDoar* os repositórios institucionais dominam na atualidade o panorama de repositórios com política de acesso aberto (82 por cento num total de 2255 repositórios listados).

Um repositório institucional digital, abreviado para repositório institucional, pode ser definido como uma base de dados na Web com material escolar cumulativo, perpétuo, e de acesso aberto para membros da comunidade institucional (Ware, cit. por Asunka et al., 2011). Se as coleções digitais têm o mérito de absorver leitores à medida que se tornam acessíveis on-line em formatos estruturados e pesquisáveis (Huwe, cit. por Asunka et al., 2011), os repositórios institucionais representam uma gestão de trabalhos eficiente em custos. Além disso facilitam o acesso aos recursos através de uma interface na Web, enquanto proporcionam aos autores uma plataforma integrada e robusta para o armazenamento e disseminação do trabalho que produzem. Este tipo de repositório tem suscitado grande interesse, nomeadamente nas universidades, sobretudo devido à proliferação de conhecimento possibilitada pelas tecnologias emergentes. As universidades estão preocupadas com a administração dos ativos que produzem. Os repositórios prometem expandir o acesso a recursos de investigação, ao mesmo tempo que criam os componentes críticos para a reformulação do sistema corrente de comunicação científica dentro das instituições ou na academia em geral (Asunka et al., 2011). Neste contexto um grande número de instituições de ensino superior está a adaptar ou a implementar repositórios.

Crow (2002) adianta que os repositórios podem servir como indicadores tangíveis da qualidade de uma universidade, podendo demonstrar a relevância social, científica e económica das atividades de investigação. Para além disto aumentam a visibilidade, estatuto e valor público das instituições. O interesse da comunidade académica nos repositórios digitais prende-se em grande medida com a lógica de iniciativa do auto-arquivamento para docentes e alunos do ensino superior, e com o descontentamento em relação ao sistema de publicação monopolista. No entanto, estes aspetos não são determinantes para captar o interesse dos investigadores, que encontram no reconhecimento pessoal e no impacto entre pares a motivação para depositar os resultados do seu trabalho (Thomas e Mcdonald, 2007), o que nem sempre é fácil de aferir através de um repositório. No caso dos repositórios institucionais existem duas realidades, os que obrigam os estudantes e docentes a auto arquivar os trabalhos, e aqueles que não o fazem. Independentemente da sua realidade, esperar que os “participantes” colaborarem não pode ser tido como garantido. De facto, ter um rácio de participação

²⁰ OpenDoar. Disponível em: <http://www.opendoar.org/> Consultado a 29 de novembro de 2012.

significativo é um dos maiores desafios dos repositórios institucionais e os resultados do envolvimento voluntário da comunidade acadêmica não se têm revelado satisfatórios (Thomas e McDonald, 2007).

Para Clifford Lynch, um dos primeiros visionários dos repositórios institucionais, o campo de atuação em que os repositórios podem fazer a diferença é na adoção de uma atitude alternativa em relação aos modelos de publicação tradicional. Nas palavras de Lynch - *“the big, important mission for institutional repositories revolves around preserving access to underlying data and things that don't look very much like traditional publishing”* (cit. por Albanese, 2009). Como complemento a esta observação, Lynch acrescenta que as principais agências de financiamento (NSF, Andrew W. Mellon Foundation) estão a reconhecer os dados de investigação como um recurso precioso e estão cada vez mais formais em relação ao destino dos dados, ou seja, vão passar a exigir dos investigadores uma gestão e preservação eficiente dos dados de investigação. Por isso mesmo os investigadores vão necessitar do apoio das suas instituições para satisfazerem os requisitos dos financiadores, e nesta perspectiva os repositórios institucionais podem desempenhar um serviço de apoio à gestão dos dados (cit. por Albanese, 2009). Por outro lado, os repositórios digitais, quer sejam agregadores de dados ou de documentos, são tipicamente utilizados para ingerir conteúdos perto, ou já no fim, do ciclo de vida. Esta não é uma perspectiva desejável para a gestão dos dados de investigação, uma vez que os cientistas precisam de ter acesso aos dados desde uma fase inicial, podendo os repositórios ser uma solução se forem concebidos como sistemas que acompanhem o ciclo de vida dos dados completo.

1.5 Repositórios de dados

Apesar da crescente preocupação em torno da gestão de dados de investigação produzidos a larga escala e da sua importância para o desenvolvimento científico, os repositórios dedicados a conjunto de dados estão atualmente numa posição aquém do desejável. De acordo com estatísticas atuais do *OpenDoar* (fig.3) o número de repositórios dedicados a conjunto de dados de investigação é ainda residual, se comparado com outros conteúdos, suplantando apenas o número de repositórios de patentes e *software*.

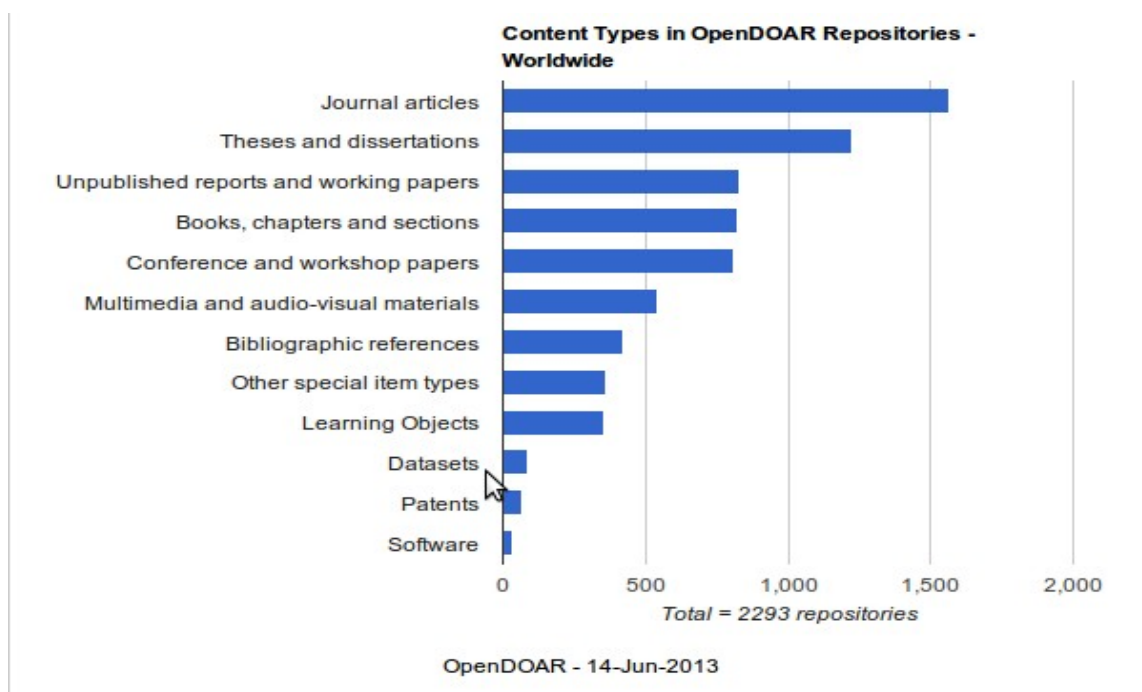


Figura 3: Tipo de conteúdos em repositórios de acesso aberto (OpenDoar, 2013)

Porém, recorrendo a diretórios de repositórios de dados é possível perceber que estão em curso várias iniciativas neste âmbito. Exemplos disso podem ser consultados através do diretório *Databib*²¹. No *Databib* encontra-se um catálogo de pesquisa que ajuda a identificar e localizar repositórios de dados on-line, listando atualmente um total de 86 repositórios de dados por todo o mundo. A comunidade científica envolvida neste projeto, cria e cuida registos que descrevem os repositórios de dados para que estes possam ser pesquisados. O serviço prestado pelo *Databib* é útil aos *stakeholders* envolvidos com os dados de investigação. Através do *Databib*, os investigadores ficam a saber quais os repositórios mais apropriados para submeter dados, os utilizadores encontram repositórios que agreguem dados que coincidam com as suas necessidades, e permite aos bibliotecários ajudar os investigadores na tarefa de localizar ou integrar dados nos seus processos de pesquisa ou de aprendizagem. No fundo o que este projeto tem vindo a promover é o contato entre investigadores com os repositórios de dados, ou com os profissionais envolvidos na gestão de dados.

Outros exemplos de diretórios a atuar neste contexto são o *Open Access Directory* (Simons College, Boston), em funcionamento desde 2008, que fornece *links* para repositórios de dados e bases de dados, segmentado por disciplina, e o *MIT Data Management and Publishing: Sharing Your Data*²², guia multidisciplinar da biblioteca do MIT para a gestão de dados, assim como o já referido *DataCite*.

²¹ Databib; Disponível em: <http://databib.org/> Consultado a 14 de junho de 2013

²² Data Management and Publishing. <http://libraries.mit.edu/guides/subjects/data-management/publishing.html> Consultado a 05 de janeiro de 2013

Também a Amazon está a disponibilizar um serviço centralizado para o *download* de conjunto de dados do domínio público que podem ser integrados em aplicações baseados em nuvem da Amazon Web Services²³. A Google, por sua vez, lançou recentemente o “*Google Public Data Explorer*”²⁴, para a exploração de conjuntos de dados provenientes de vários organismos internacionais, como o World Bank, a Eurostat e a OCDE. Há, ainda, diretórios especializados em disciplinas científicas concretas, sendo esse o caso do *Oceanographic Data Repositories*²⁵, fundado pela NSF, e do *Global Change Master Directory*²⁶, este último tutelado pela NASA.

A nível internacional podem ser identificados alguns projetos relacionados com os repositórios de dados e boas práticas de curadoria, que conjugam os interesses da ciência e da sociedade com o acesso aberto, permanente e seguro a dados científicos devidamente recolhidos e descritos. Exemplos de projetos financiados por organismos oficiais são o *Data Observation Network for Earth (DataONE)*²⁷, que colabora no desenvolvimento de uma organização que procura suportar a preservação e acesso a dados multidisciplinares produzidos a larga escala na área das ciências ambientais, e o *Australian National Data Service (ANDS)*²⁸, que sob o lema “*more researchers reusing data more often*”, é um serviço que tem trabalhado na construção de uma coleção coesa de recursos de investigação na Austrália desde 2009. O ANDS contribui para a formação de especialistas na criação, gestão e partilha da dados, de acordo com políticas de gestão robustas, para o aumento da rotina do depósito de dados em ambientes de gestão e preservação que sejam sustentáveis, acessíveis e estáveis, facilitando a disseminação de dados científicos produzidos na Austrália para dar suporte a equipas de investigação multidisciplinares, tanto na Austrália como a nível internacional (Treloar, 2009).

Outros projetos conjugam esforços de organizações de carácter científico, com destaque para o repositório do *National Center for Biotechnology Information (NCBI)*²⁹ que disponibiliza o acesso (limitado) a um conjunto de bases de dados de informação nas áreas da biomedicina e do genoma, entre as quais a *PubMed*, e para o repositório *Interuniversity Consortium for Political and Social Research (ICPSR)*³⁰, um consórcio com mais de 700 instituições académicas e organizações de investigação, que permite o acesso a uma coleção extensa de dados na área das ciências sociais, desenvolvendo ainda ações de formação em acesso e curadoria de dados.

²³ Amazon Web Services. Disponível em: <http://aws.amazon.com/pt/publicdatasets/> Consultado a 05 de janeiro de 2013

²⁴ Google Public Data Explorer. Disponível em: <http://www.google.com/publicdata/directory?hl=pt&dl=pt> Consultado a 05 de janeiro de 2013

²⁵ Oceanographic Data Repositories. Disponível em: <http://www.bco-dmo.org/data> Consultado a 05 de janeiro de 2013

²⁶ Global Change Master Directory. Disponível em: <http://gcmd.nasa.gov/KeywordSearch/Home.do> Consultado a 05 de janeiro de 2013

²⁷ DataONE. Disponível em: <http://www.dataone.org/> Consultado a 10 de novembro de 2012

²⁸ ANDS. Disponível em: <http://www.ands.org.au/> Consultado a 12 de novembro de 2012

²⁹ NCBI. Disponível em: <http://www.ncbi.nlm.nih.gov/> Consultado a 12 de novembro de 2012

³⁰ ICPSR. Disponível em: <http://www.icpsr.umich.edu/icpsrweb/landing.jsp> Consultado a 12 de novembro de 2012

Por outro lado o *Disk-UK DataShare Project*³¹, impulsionado pelo centro de dados *EDINA*, visa explorar alternativas para investigadores que desejem partilhar dados na Internet. Este projeto que conta com a participação das Universidades de Edimburgo, Oxford e Southampton, conduziu à criação de repositórios institucionais de dados ou serviços relacionados. O desenvolvimento desta iniciativa é consonante com o esforço demonstrado pelas bibliotecas ligadas a organizações de investigação em criar repositórios institucionais que façam a gestão dos recursos internos.

Por fim, deve-se destacar um projeto colateral aos repositórios de dados, o periódico *ESSD (Earth System Scientific Data)*. Idealizado por Hans Pfeiffenberger e David Carlson, o *ESSD* surgiu em 2008 com a ambição de providenciar um espaço para publicação de dados de investigação, tidos como “importantes”. Pfeiffenberger e Carlson (2011) explicam que o *ESSD* não pretende ser um substituto dos repositórios de dados mas sim um complemento - “*since data will not be stored or made accessible by the publisher (not to speak of preservation); rather, the editors require the dataset to reside in a reliable repository and to have a stable way of linking to it*”.

Esta iniciativa procura recompensar os proprietários dos dados através de um mecanismo de citação qualificado, alinhado com a certificação de qualidade idêntica aos jornais revistos por pares. O objetivo da *ESSD* é motivar os investigadores a publicar os seus dados, por outras palavras, atribuir na perceção dos investigadores um sentido prático para a submissão de dados. A recompensa é então vista como essa motivação, devendo o reconhecimento da publicação de dados ser encarado como um feito académico (Pfeiffenberger e Carlson, 2011).

1.6.1 Repositórios de dados e a gestão dos dados de investigação

Na literatura corrente sobre repositórios e curadoria de dados é bastante comum a referência à gestão de dados de investigação. O conceito “*research data management*” emerge, também, associado aos vários projetos que têm aparecido neste domínio, alguns dos quais já referidos. Apesar da tendência para a adoção deste conceito, a sua aplicação prática não está consolidada, estando ainda numa fase exploratória, segmentada em diversas iniciativas.

Inerente a esta tendência está a adoção por parte da NSF de uma abordagem a nível de regulamentos e procedimentos para responder aos desafios que acompanham a proliferação de dados, de forma a ser evitada a perda, o excesso e a dispersão dos dados de investigação. Para contrariar as más práticas de

³¹ Data Information Specialists Committee. United Kingdom. Disponível em: <http://www.disc-uk.org/datashare.html>
Consultado a 16 de novembro de 2012

gestão de dados a NSF promulgou que todas as propostas de financiamento a ela dirigidas devem prever um plano de gestão de dados digitais, com o intuito de os tornar - *“routinely deposited in well-documented form, regularly and easily consulted and analysed by specialist and non-specialists alike, open accessible while suitably protected, and reliably preserved”* (National Science Foundation, 2010)³². Como consequência desta medida o DataONE criou o *Data Management Planning Tool (DPMTool)*³³, em finais de 2011, para ajudar os investigadores a elaborar um plano genérico em conformidade com os requisitos da NSF.

A organização composta por profissionais do ensino superior no Reino Unido, formalizado através do DISK – UK, é uma das parcerias que se tem mostrado mais ativa no apoio concedido aos estudantes e restantes académicos na gestão dos seus dados. Um dos principais contributos desta parceria foi a apresentação numa conferência da IASSIST, em 2009, de um guião que fornece a outras instituições um método que orienta a tomada de decisão na aceitação de conjunto de dados em repositórios. A metodologia DAF (*Data Audit Framework – fig. 4*), desenvolvida pela comunidade DCC, surgiu em consequência das recomendações do JISC, expressas por Lyon (2007) - *“a framework must be conceived to enable all universities and colleges to carry out an audit of departmental data collections, awareness, policies and practice for data curation and preservation”*. Assim, o DISK-UK utiliza o DAF para identificar conjuntos de dados criados dentro das suas universidades, e para responder às preocupações sobre a partilha de dados num repositório de acesso livre (Rice e Haywood, 2011).

O DAF recomenda que as auditorias feitas aos dados de investigação sejam um processo assente em quatro etapas. Na fase de planeamento são definidos os propósitos da auditoria, sendo feito um levantamento preliminar junto dos investigadores. A segunda etapa serve para identificar e classificar os dados existentes. A classificação vai determinar o âmbito das atividades posteriores, isto porque os aspetos seleccionados vão ser tratados com maior detalhe nas fases seguintes. O terceiro ponto vai permitir ao auditor recolher informação para detetar os pontos fracos na gestão dos dados de investigação, permitindo-lhe enumerar recomendações na fase final do processo. O conhecimento adquirido através da auditoria possibilitará às organizações planear melhorias na gestão dos dados de investigação³⁴.

³² NSF. 2010. Scientist Seeking NSF Funding will soon be required to submit Data Management Plans. Disponível em: http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928 Consultado a 25 de novembro de 2012

³³ DataONE. Data Management Planning 2011. Disponível em: <http://www.dataone.org/data-management-planning> Consultado a 02 de janeiro de 2013

³⁴ The Digital Curation Center. Disponível em: <http://www.data-audit.eu/methodology.html> Consultado a 13 de novembro de 2012



Figura 4: Data Audit Framework

Em consequência da aplicação do DAF a casos de estudo tem vindo a ser demonstrada a necessidade em melhorar a prática de gestão de dados. Os principais problemas apontados nestes estudos são a carência de espaço de armazenamento, a fraca perceção dos dados como um valor a reter a longo termo, as práticas de gestão *ad-hoc*, a inexistência de guias e padronização de procedimentos de gestão dos dados. Para além disso verifica-se pouco enriquecimento de metadados, sendo os esforços geralmente canalizados na busca de informação complementar em servidores externos (Rice e Haywood, 2011).

Atendendo à necessidade de incutir nos investigadores uma maior capacidade para lidar com os dados, foi lançada na Universidade de Edimburgo, Escócia, uma iniciativa entre os meses de agosto de 2010 e julho de 2011, que consistia na disponibilização de materiais *e-learning* que refletissem as melhores práticas de gestão de dados de investigação em domínios como as ciências sociais, psicologia clínica e geociência. O designado *Research Data Mantra* foi um programa de treino para jovens cientistas, que incluía entrevistas, gravadas em vídeo, a académicos experientes nos desafios que a gestão de dados de investigação acarreta (Rice e Haywood, 2011).

A Universidade de Oxford, perante tal problema criou uma solução de gestão de dados compatível com todas as etapas do ciclo de vida dos dados (fig.5), suficientemente genérica para poder ser adotada por instituições com diferentes estruturas.

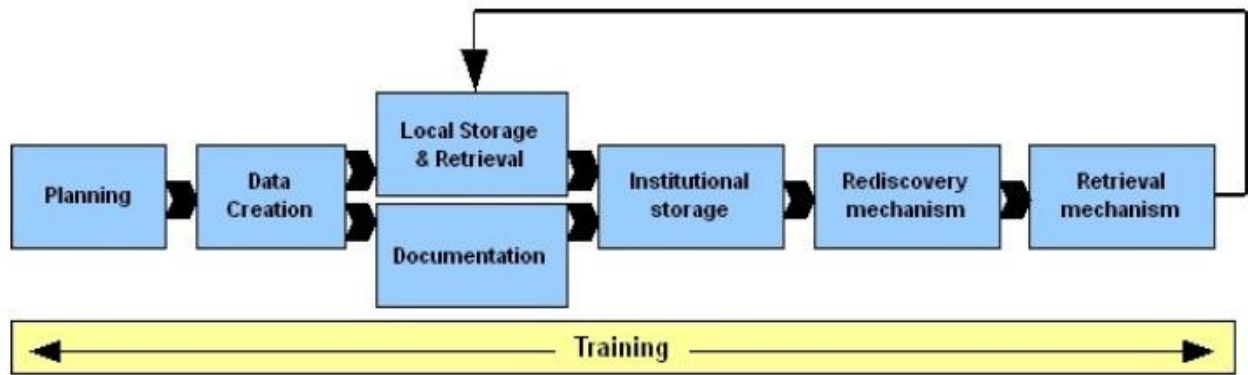


Figura 5: Requisitos para uma infraestrutura de gestão de dados de investigação de U. de Oxford (Wilson et. al., 2011)

Segundo este *framework* a gestão dos dados de investigação pode ser concebida como uma sequência de passos, onde cada um deve ser completado de forma adequada para que os dados tenham condições de prosseguir para o próximo passo, de modo a terem o seu valor maximizado - “Se a infraestrutura não suportar um destes passos, ou os agentes envolvidos não entenderem o que se espera deles, então o valor potencial dos dados não pode ser atingido” (Wilson et al., 2011). Tendo em conta o *framework* proposto, o suporte concedido pela Universidade de Oxford aos projetos de investigação reside na disponibilização de uma equipa de acompanhamento a investigadores em regime de financiamento.

1.6.2 Gestão de dados de investigação: o caso do laboratório Bennet (U. Michigan)

A questão da gestão de dados de investigação está bem documentada num estudo de caso realizado por Akmon et al., (2011), que consistiu na observação de uma equipa de cientistas do ramo dos materiais em contexto de trabalho num laboratório universitário. Este estudo teve como principal objetivo perceber como os dados eram produzidos, quais as práticas utilizadas pelos cientistas na gestão dos dados e quais os desafios que estes tinham de enfrentar.

Os resultados deste estudo mostraram que os cientistas estão envolvidos em atividades de documentação e gestão de um extenso número de dados de investigação, mas que a reutilização dos dados entre eles se demonstrou um processo complicado. Este problema foi particularmente identificado pela líder do laboratório, revelando que a incapacidade para aceder com facilidade aos dados produzidos pela sua equipa afetava a pesquisa. Para além disso, e apesar de compartilharem

motivações comuns nas práticas em relação aos dados, a forma como os dados foram capturados e registados variou consideravelmente entre os cientistas.

Segundo Akmon et al. (2011) a justificação para a falta de coerência no registo de dados reside em dois fatores que se intercetam. Por um lado, a relutância da líder do grupo em impor um sistema comum a todo o grupo e o facto desta nunca ter fornecido à sua equipa um guião adequado à gestão dos dados. São vários os motivos que explicam esta relutância, nomeadamente a não qualificação da mesma como especialista em gestão de dados e o desinteresse em controlar as práticas dos subalternos. Por outro lado, verificou-se que cientistas observados valorizaram a documentação dos dados de acordo com as suas próprias necessidades, não ficando claro o que se pode esperar destes em relação ao tratamento que dão aos dados quando estão em jogo as necessidades de outros cientistas. Mediante esta dedução o que precisa de ser respondido é se a perda de tempo na preparação de anotações detalhadas é suplantada pelos benefícios que se podem tirar desta prática.

A lição que se pode tirar das conclusões deste estudo é o da existência de um espaço para a colaboração entre cientistas e curadores. Primeiro, porque os cientistas tiveram o cuidado de criar sistemas para encontrar os próprios dados no futuro, como a adoção de ferramentas sociais para a documentação dos dados. Sistemas que apesar de idiossincráticos, típicos da organização em rede, não são necessariamente *ad-hoc*. O que na opinião de Akmon et al. (2011) revela uma predisposição dos cientistas em criar sistemas pessoais de gestão e documentação, que uma vez estabelecidos serão usados com frequência. Isto indica o alto valor da gestão e documentação do trabalho científico e que as alterações ao sistema terão efeitos significativos sobre o trabalho que é feito no laboratório. Segundo, porque a conexão entre os dados e os documentos a si associados influenciam a potencial reutilização dos mesmos. A manutenção de uma relação concetual entre os dados de investigação e a documentação é um pré-requisito que poderá garantir o uso dos dados como *input* de novo conhecimento, enfatizando a necessidade da sua manutenção conjunta a partir de um repositório.

Do ponto de vista da curadoria será preciso um entendimento da terminologia técnica utilizada pelos cientistas, para que se percebam os interesses e metodologias de investigação desta comunidade. Só assim será possível aos curadores uma participação efetiva na tarefa de gestão dos dados de investigação. A curadoria de dados proporciona uma oportunidade para os profissionais envolvidos na gestão de dados terem impacto na prática científica. Terminando com uma reflexão de Akmon et al. (2011) – os cientistas à medida que se debatem com as contrariedades do dilúvio de dados, *“along with reports of frustration with data management, suggests that the expertise of non-scientist data curators (...) may indeed be welcome in the lab.”*

Capítulo 2.

Estudos de utilização

2.1 Os estudos de utilização e a Ciência da Informação

Numa era de desenvolvimento tecnológico em conjugação com o excesso de informação, os estudos de utilização são um fator crítico para os profissionais da informação, permitindo-lhes obter um maior entendimento sobre as necessidades dos utilizadores e corresponder com um serviço de maior qualidade. Seja qual for a circunstância em que se aplique um estudo de utilização, deve ser identificado com a maior precisão possível a sub-área em que este ocorre. Os estudos de utilização são uma das áreas mais exploradas em Ciência da Informação, formando um largo campo literário nesta disciplina.

Este tipo de estudo cobre um vasto escopo de potencial investigação, e pode passar pelo modo pelo qual um utilizador escolhe determinado livro, da reação perante *outputs* da pesquisa on-line, até à análise profunda das necessidades subjacentes que levam a uma necessidade de informação.

Um estudo de utilização é uma atividade que exige um significativo consumo de tempo e esforço. No entanto, o entendimento que se pode retirar da observação dos utilizadores em determinado contexto pode ter o valor suficiente para fazer evoluir um serviço de informação, justificando assim o empreendimento (Wildemuth, 2003). O critério de sucesso destes estudos não é o quão bem conseguem descrever o comportamento dos utilizadores, mas sim no impacto que terão na formulação de estratégias de informação (Mick et al., 1980). A ideia inerente a este tipo de estudo é que apenas percebendo os utilizadores se pode inovar em termos de desenvolvimento de novos serviços que acompanhem as necessidades dos mesmos (Snow et al., 2008).

Saber mais sobre os utilizadores permite satisfazer melhor as suas necessidades. Contudo, o desejo vago de conhecer as necessidades dos utilizadores não é suficiente para desenvolver e conduzir um estudo de utilização que produza evidências empíricas válidas. Para cada estudo é necessário formular uma pergunta específica a qual o estudo vai responder, que diga respeito, não somente às necessidades dos utilizadores, mas também ao contexto em que este se vai desenrolar (Wildemuth, 2003).

Na perspectiva de Tom Wilson (2000) o utilizador pode ser enquadrado em diversos contextos e os estudos de utilização, no seu planeamento, devem definir com a maior exatidão possível esse contexto. Os contextos em que tem sido apropriado aplicar este tipo de estudo, e aos quais os investigadores mais se têm dedicado, apresentam o utilizador como: **comunicador** - de que forma as pessoas/organizações utilizam os recursos à sua disposição para comunicar com os seus pares, sobretudo; **pesquisador** - o foco tende a ser na interação das pessoas com os sistemas de informação; **recetor** - nem todos os serviços, ou sistemas de informação são passivos, alguns levam o produto ao utilizador, como é o caso de um SDI (Serviço de Informação). Os estudos neste domínio são mais direcionados para a avaliação do sistema, principalmente com o objetivo de aprimorar o serviço.

2.1.1 Perspetiva Histórica

Os estudos de utilização tiveram a sua origem durante os anos 40, e desde então não têm parado de crescer, com a Ciência da Informação, por tradição, a revelar particular interesse nesta área. A título de exemplo, na *Royal Scientific Information Conference*, de 1948, foram apresentados vários estudos que diziam respeito ao estudo dos utilizadores, e dez anos depois este tipo de estudo intensificou-se (Siatry, 1999).

Anos 60 - Os anos 60 foram, por assim dizer, a rampa de lançamento dos estudos de utilização, duas bibliografias apresentadas por Herbert Menzel, na primeira *Annual Review of Information Science and Technology*, de 1966, revelavam um total 438 e 679 de estudos de utilização realizados no ano de 1964 e 1965, respetivamente. Os estudos de utilização tornaram-se mais sofisticados, na medida em que passaram a procurar respostas para perguntas mais complexas e passaram a focar-se com maior incidência nas comunidades científicas como alvo de avaliação. Neste período as ciências sociais e comportamentais introduziram novos quadros teóricos e princípios que levaram os cientistas da informação a refletir sobre a aplicação dos mesmos aos estudos de utilização, ou seja, houve um despertar para questões de ordem metodológica. Menzel foi um dos primeiros a pensar esta situação, percebendo-se do potencial uso de métodos alternativos, argumentando que as técnicas utilizadas na época eram pouco adequadas à complexidade das tarefas com que os investigadores tinham que lidar (cit. por Siatry, 1999). Mote que foi unanimemente aceite pelo núcleo de investigadores na área, surgindo, assim, novas ideias, *frameworks* conceituais alternativos, metodologias e revisões regulares aos estudos.

Anos 70 - Registou-se uma maior diversidade do público-alvo dos estudos, tendo o foco sido estendido à sociedade em geral. Os anos 70 ficam particularmente marcados pelo estudo de sistemas

de informação concretos, com vista a eficácia e eficiência destes serviços. Não se desenvolveu um *framework* teórico, mas um maior número de investigadores ficou ciente nas debilidades dos métodos aplicados e da fraca concetualização. Podendo-se destacar um marco importante na história do estudo de utilizador, o estabelecimento, na Universidade de Sheffield, do *Center for Research on User Studies* (CRUS).

Anos 80 - Época em que ocorre a intromissão das tecnologias de informação nos estudos de utilização, estabelecendo-se desde então a sua presença como área de interesse dos estudos de utilização, a ênfase nos estudos era dado ao levantamento de requisitos para o desenvolvimento de *software*. Nos anos 80 passaram a ser planeados serviços e sistemas de informação capazes de satisfazer as necessidades dos utilizadores, sem grandes resultados devido à dificuldade em perceber a complexidade do comportamento e necessidades dos mesmos. Mick et al. (1980) observam que os primeiros estudos das necessidades e comportamento informacional tinham uma natureza descritiva. Recorriam a variáveis, para a explicação do comportamento, que pouco contribuíam para o trabalho dos profissionais envolvidos com a gestão de sistemas de informação. Foram, ainda, aplicadas teorias de outras áreas do conhecimento como a Psicologia, Comunicação, Marketing, entre outras. Belkin (1980) produz o *Anomalous State Knowledge*, direcionado para a recuperação de informação, dando nova dimensão ao conceito “necessidade de informação” e à origem dessa necessidade.

2.1.2 O modelo de Tom Wilson (1981)

Em 1981, Tom Wilson escrevia o artigo “*On User Studies and Information Needs*”. Nele pode ser lido que os estudos de utilização combinam alguns dos mais usados e menos detalhadamente definidos conceitos na Ciência da Informação. Conceitos como necessidade de informação, comportamento informacional e canal de comunicação, coexistem num sistema de relações complicadas e interdependentes. O modelo proposto por Tom Wilson (fig.6) sugere que o comportamento informacional resulta da necessidade de informação percebida pelo sujeito, procura estabelecer a coexistência destes conceitos, apresentando as relações dinâmicas entre eles, e tal como o próprio afirma, trata-se de uma forma de “pensar” os estudos de utilização.

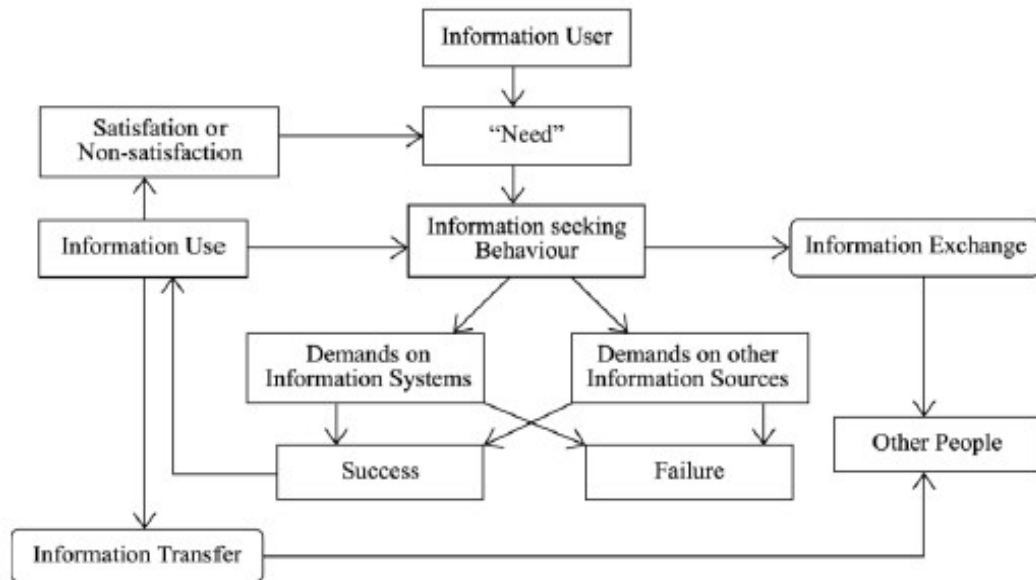


Figura 6: Modelo de Comportamento Informacional (Wilson, 1981)

O conceito “necessidade de informação” é vital para os estudos de utilização em qualquer dos contextos em que estes se desenrolem (Wilson, 2000). Segundo Tom Wilson (1981), “necessidades de informação” é um conceito que provoca muito debate e confusão. Sobretudo na distinção entre o que se entende por informação, dado ou conhecimento. Parte do problema está na dificuldade para detalhar o conceito de informação (tido como vago). Num estudo de utilização a “informação” pode ser interpretada como uma *entidade física* ou *fenómeno* (como o número de livros lidos ou o número de jornais subscritos), *meio de comunicação* (transmissão de dados, seja oral ou escrita), *subject data* (o que está contido num documento ou palavras), *dados concretos* transmitidos com objetividade, ou como uma simples *opinião* (juízo de valor). O múltiplo uso do conceito “informação” ocorre porque os investigadores falham em distinguir entre um sentido ou outro, ou simplesmente porque deixam que seja o leitor a interpretar esse sentido (Wilson, 2000). Na opinião de Bawden (2006) é difícil acreditar que esta situação tenha evoluído favoravelmente nos anos mais recentes, talvez tenha até piorado, tendo em conta que o termo “informação” é usado amplamente como um conceito central em outros ramos científicos, como a Física e a Biologia.

Este é um modelo amplamente citado em publicações mais recentes, contudo não aplicado na prática por outros autores, o que Bawden (2006) interpreta como um problema geral na consolidação teórica da Ciência da Informação. Verificando-se uma maior preocupação em criar modelos do “zero”, ao invés de reproduzir modelos prévios, o que denota uma fraca abordagem científica ao objeto de estudo. Contudo, a relativa simplicidade deste esquema abriu caminho para a intensificação de

tentativas mais sofisticadas de modelar macro-comportamentos (Cronin 2001). Este modelo foi posteriormente alargado em 1994 com a introdução de condicionantes internas e externas.

Como nota Wilson (1981) o comportamento informacional pode destacar-se como um tipo de investigação por si só, quando o motivo da investigação passar pelo desenvolvimento ou design de um sistema. Os processos que envolvem um mediador ou um sistema de informação, ou a identificação de estratégias de busca implementadas por um mediador para satisfazer a demanda de informação do utilizador, podem ser estudados para aferir o quão eficiente é a performance de uma tecnologia.

2.1.3 Desenvolvimento da investigação qualitativa

Após o trabalho de desenvolvimento de *frameworks* e teorias durante os anos 80, os anos subsequentes corresponderam ao crescimento de conceptualizações com base no conhecimento adquirido anteriormente, acompanhado por cada vez mais sofisticados métodos de investigação qualitativa. Os anos 90 testemunharam a explosão do fenómeno Internet, e o seu estabelecimento como fonte de informação para os cientistas da informação. Resultado disso foi o crescente número de trabalhos concentrados na infinidade de assuntos relacionados com o mundo Web, tal como o impacto da Internet nos utilizadores. Assim, uma nova geração de estudos de utilização começou a focar-se mais nos *end-users* e na forma como estes interagem com a informação, a partir do momento em que as necessidades eram identificadas (Tenopir, 2003).

No passado o dominador comum aos estudos de utilização foi o “método científico”, muitas vezes mal-interpretado, com o inquérito a larga escala a ser o método predominante na recolha de dados. A transição para a adoção dos métodos de ordem qualitativa, ocorreu quando se tornou evidente que os estudos quantitativos dificilmente ofereciam respostas para a identificação de necessidades individuais e não contribuíam para o desenvolvimento de sistemas de informação preparados para responder a essas necessidades. Tom Wilson (2000) sintetiza claramente esta observação - *“Quantitative social survey methods are powerful data collection devices, so much so that they are often misused to collect a great deal of data without the researcher having a theoretical (or even descriptive) framework into which to fit the data. In effect, collecting data becomes a substitute for thinking about the problem. This is to criticise the unthinking application of methods, rather than the methods themselves.”*

A pesquisa qualitativa atende a causas de reação dos utilizadores e à resolução do problema informacional, sendo mais apropriada para a identificação das necessidades subjacentes de

informação. Várias teorias e modelos de avaliação qualitativa foram desenvolvidos como suporte à investigação na área dos estudos de utilização. Entre as mais citadas entre os estudos correntes podem ser destacadas os seguintes:

O *user-value* de Taylor (1986) – O propósito do *Value-Added Model* de Taylor é fornecer um quadro que considere a informação e os sistemas de informação na perspectiva do utilizador. Assim, é um modelo assente em três elementos:

Utilizador - Os sistemas existem para ir ao encontro das necessidades dos utilizadores. Estes devem ser vistos como parte do sistema.

Informação - Existe uma hierarquia na informação – *spectrum* da informação. Quando é acrescentado valor, este *spectrum* evolui, de dado para informação, de informação para conhecimento, de conhecimento para ação.

Sistema - O objetivo de um sistema de informação é acrescentar valor para uma maior aproximação às necessidades do utilizador.

O modelo de Kuhlthau (1991) - Assente nos trabalhos de Dervin e Nillan (1986), Kelly (1963), Belkin (1981) e Taylor (1986). O *Information Seeking Process* (ISP) foi desenvolvido após um conjunto de pesquisas sobre a experiência e o comportamento dos utilizadores de bibliotecas, sendo posteriormente estendido a outros ambientes de trabalho, como estudos de utilização em repositórios digitais. O modelo de processo consiste na observação, implicando seis etapas: início, seleção, exploração, formulação, coleta e apresentação. Nas palavras de Carol Kuhlthau o modelo serve para descrever o ISP do ponto de vista do utilizador, sendo tipicamente utilizado para avaliar aspetos cognitivos e afetivos dos utilizadores enquanto interagem com um sistema de informação, permitindo ao investigador suprir as diferenças entre o processo natural do uso de informação pelo utilizador e o proposto pelo sistema.

Contextual Design (1999) - Introduzido por Hugh Beyer e Karen Holtzblatt, é um processo centrado no utilizador, que incorpora métodos etnográficos de recolha de dados para interpretar a forma como os utilizadores trabalham, com o objetivo de desenvolver sistemas e serviços apropriados à realidade dos mesmos. Este método combina entrevistas de fundo e observação contextualizada (*Contextual Inquiry*) seguida da modelação do *workflow* (*Work Modelling*) e da análise de resultados com o uso de um diagrama de afinidades (*Consolidation*). A partir daqui um sistema pode ser remodelado (*Work Redesign and User Environment Design*) e testado por utilizadores antes da sua implementação (*Test with users*).

A introdução de novas tecnologias acelerou o processo de inclusão dos estudos de utilização como um elemento básico em projetos orientados ao utilizador. Uma Ciência da Informação mais focada na compreensão dos utilizadores no seu contexto de trabalho será certamente mais útil a um profissional da informação, ao apontar o caminho da inovação dos serviços de informação.

2.2 Estudos de utilização aplicados a repositórios digitais

Os repositórios digitais implicam vários problemas durante o seu desenvolvimento, devido à larga quantidade de informação envolvida e à dificuldade em perceber o que os utilizadores esperam deste tipo de sistemas. O processo de implementação ou alteração de um repositório digital, qualquer que seja a sua natureza, está sujeito a realização de estudos que validem e garantam a sua utilidade para os utilizadores finais. Estes estudos recorrem geralmente a suporte metodológico de áreas do conhecimento como a *Interação Pessoa Computador* (IPC)³⁵, sobretudo em testes de usabilidade, e à Ciência da Informação, na identificação das necessidades e comportamento informacional do utilizador. Estudos com este foco tornam possível aos profissionais envolvidos com os repositórios digitais perceber os modelos mentais dos utilizadores, facilitando o desenvolvimento de um *design* mais útil e prático de acordo com as necessidades do público-alvo (Ferreira e Pithan, 2005). Tais estudos podem servir vários propósitos, havendo várias técnicas para a recolha de dados.

Decidir o que pode ser estudado é apenas a primeira parte do problema. O que pode ser medido depende, obviamente, da metodologia aplicada. Existem muitas metodologias que podem ser usadas mas há dois métodos principais que ditam o que é medido. Carol Tenopir (2003) agrupa essas metodologias em dois grupos distintos: **Observação** - como técnica exige muito consumo de tempo mas produz conhecimento precioso sobre o modo de utilização dos sistemas. A observação pode incluir diversas metodologias específicas, como por exemplo a verbalização (*thinking out loud*), testes de usabilidade, registos de pesquisa (*query log*). **Questionário** - engloba várias técnicas como inquéritos e entrevistas, com maior ou menor profundidade, grupos-alvo, sondagens de opinião, entre outros. Contudo, estudos que combinem ambos os métodos permitiram obter melhores resultados.

Os testes de usabilidade correspondem à maioria dos estudos feitos com o objetivo de avaliar o desempenho de um repositório digital, e são geralmente implementados com base em três critérios; a avaliação dos conteúdos, da funcionalidade e do interface do ponto de vista do utilizador. A

³⁵ A IPC estuda o contato dos humanos com sistemas computadorizados, mais concretamente a interação que decorre deste contacto. Neste contexto a usabilidade corresponde à extensão a que um produto pode ser utilizado por utilizadores concretos para atingir um objetivo de modo eficaz, eficiente e satisfatório num contexto de utilização específico.

International Organization for Standardization (ISO) define usabilidade como “*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction, in a specified context of use*” (cit. por Kim e Kim, 2008). Shneiderman (cit. por Hariri e Norouzi, 2011), determina que a usabilidade de qualquer sistema informático pode ser avaliada pela conjugação dos seguintes aspectos orientados ao utilizador:

- 1) aprendizagem** - aprender como aceder a um recurso deve ser um processo intuitivo e de fácil entendimento;
- 2) rapidez na execução de tarefas** - encontrar um recurso importante deve ser feito em pouco tempo;
- 3) rácio de erros diminuto** - erros de omissão (não encontrar o pretendido) ou encontrar o que não se pretende deve ser um acontecimento raro;
- 4) satisfação subjetiva** - a pesquisa deve ser uma experiência agradável e compensatória;
- 5) retenção do utilizador** - voltar a utilizar o sistema, após um período sem o fazer, não deve requerer novo processo de aprendizagem.

A questão de usabilidade é fundamental para aferir o sucesso de um repositório digital, porque quando um sistema é difícil de utilizar, os utilizadores podem não conseguir os resultados que pretendem, podem interpretar erradamente aquilo que encontram, podendo mesmo desistir da interação se considerarem que o sistema não os consegue ajudar (Hariri e Norouzi, 2011).

Em 2005, Judy Jeng lançou uma revisão de literatura aos estudos de usabilidade em bibliotecas digitais, com o objetivo de desenvolver um modelo de avaliação para estes sistemas. Nesse sentido a autora chegou à conclusão da existência de um conjunto de elementos comuns a vários estudos para a determinação do sucesso de uma biblioteca digital, sendo estes a *eficiência*, *eficácia* e *satisfação*, conceitos mencionados na definição da ISO, e ainda, a já referida *aprendizagem*.

No contexto da presente dissertação foi feito um levantamento da literatura referente a estudos de usabilidade realizados em repositórios digitais, a partir de 2005. Note-se que a literatura neste domínio é ainda escassa. No entanto, as operações básicas num repositório são similares às realizadas numa biblioteca digital e, portanto, optou-se por expandir a análise aos estudos realizados em bibliotecas digitais. Na tabela (1) estão listados os objetivos dos estudos, enquanto na tabela (2) estão representados os critérios de avaliação, tarefas e técnicas de recolha de dados que têm vindo a ser aplicados para medir a usabilidade dos repositórios digitais.

Tabela 1: Objetivos dos estudos de usabilidade em repositórios digitais

Autor	Objetivo
Ferreira e Pithan, 2005	Testar usabilidade do protótipo da BD <i>InfoHab</i>
Shiri et al., 2009	Avaliar o protótipo da plataforma <i>Searchling</i>
Joo e Lee, 2011	Medir usabilidade de BD's académicas
Zhang et al., 2009	Avaliar a usabilidade de 3 BD's
Kani-Zahibi, 2006	Determinar sugestões de utilizadores para funcionalidades das BD's
Norberg et al., 2005	Processo para a criação de uma BD sustentável
Kim e Kim, 2008	Sugestão de melhorias de usabilidade de um Repositório Institucional

Tabela 2: Técnicas e critérios de usabilidade em repositórios digitais

Autor	Participantes	Critério	Tarefa	Verbalização	Grupo-Alvo	Entrevista	Inquérito
Ferreira e Pithan, 2005	6 académicos	Aprendizagem; eficácia; eficiência gestão de erros; memorização; satisfação.	Pesquisa; salvar documento.	0		0	
Shiri et al., 2009	15 académicos	Utilidade; facilidade de uso; satisfação; eficácia; eficiência	Navegação; pesquisa; pesquisa avançada; exploração livre.	0		0	0
Joo e Lee, 2011	230 utilizadores de BD académica	Eficácia; eficiência aprendizagem; satisfação	Não se aplica				0
Zhang et al., 2009	35 académicos	Eficiência; satisfação; facilidade de uso; utilidade	Pesquisa; navegação.	0			0
Kani-Zahibi, 2006	48 académicos	Levantamento de requisitos: interface, conteúdo e funcionalidades	Pesquisa; acionar alerta (RSS); ver e utilizar conteúdo.				0
Norberg et al., 2005	3 grupos: estudantes e investigadores; educadores; público em geral	Satisfação; design do interface	Pesquisa; navegação.		0	0	
Kim e Kim, 2008	15 participantes inexperientes 15 participantes experientes Grupo-alvo- 4 especialistas	Satisfação; funcionalidade; facilidade de uso; assistência ao utilizador.	Submissão de conteúdo; pesquisa; pesquisa avançada; navegação.	0	0	0	

Recorrendo a tabela 2 pode-se observar que os estudos de usabilidade, para além de recorrerem por norma aos critérios anteriormente identificados por Judy Jeng (2005), combinam várias técnicas de recolha de dados para maximizarem os resultados.

Verbalização - Ferreira e Pithon (2005), Shiri et al. (2009), Zeng et al. (2009), Kim e Kim (2008) e Norberg et al. (2005) pediram aos participantes dos seus estudos para verbalizarem os seus pensamentos, ações e problemas decorrentes da execução das tarefas propostas. Este tipo de abordagem não produz resultados estatísticos significantes, mas pode ser um bom indicador para a identificação de problemas num sistema interativo.

Entrevista - No início do teste de usabilidade realizados por Shiri et al. (2009) e Ferreira e Pithan (2005) os participantes responderam a uma série de questões acerca da natureza das suas investigações e sobre as bases de dados que estes privilegiavam. Na fase complementar à realização das tarefas os autores tiveram a oportunidade de perguntar aos participantes quais os pontos fortes e fracos do sistema. Ferreira e Pithan (2005) recorreram a entrevistas semi-estruturadas para obterem informação sobre a perceção dos participantes em relação ao desempenho e dificuldades na utilização do sistema, pontos fortes e fracos, o nível de satisfação, e sondar os aspetos cognitivos e afetivos gerados durante a interação com o sistema. Para além disso, Ferreira e Pithan questionaram os participantes sobre as suas expectativas, prioridades e sugestões. Norberg et al. (2005) fizeram questões na fase complementar do estudo para obter informação sobre a satisfação em relação aos conteúdos, organização e estética visual da plataforma testada. As entrevistas neste estudo foram abertas para possibilitarem aos participantes desenvolver uma opinião e recomendarem sugestões. Os **inquéritos** seguem a mesma lógica das entrevistas, mas podem ser aplicados para recolher dados em médias (Zhang et al., 2009; Kani-Zahibi, 2006) ou grandes comunidades (Joo e Lee, 2011), sendo geralmente fechados.

Grupo-Alvo - No estudo de Kim e Kim (2008) convidaram quatro especialistas para participarem numa entrevista coletiva. Antes da sessão de discussão em conjunto, foram enviados a estes especialistas os resultados dos testes realizados aos dois grupos de participantes. As entrevistas ao grupo-alvo tiveram duas partes, a primeira para dar a conhecer aos especialistas as condicionantes do projeto, e a segunda para que estes fizessem recomendações balizadas pelos resultados prévios. Já o estudo de Norberg et al. (2005) foi realizado num ambiente controlado. Esse ambiente era um laboratório dividido em duas salas, uma sala de teste com um computador, câmara e microfone, e uma sala de observação, onde o analista procedia ao registo das interações. Antes da realização do estudo de usabilidade os analistas pediram a voluntários que fizessem um teste inicial para estabelecer os tempos médios preferenciais para a execução das tarefas.

Os trabalhos de Kani-Zahibi (2006), Kim e Kim (2008) e Zhang et al. (2009) recorreram ainda a **avaliação heurística** nos seus estudos. A avaliação heurística foi desenvolvida por Jakob Nielsen, e tem sido fortemente aplicada no design de interfaces, sendo uma inspeção sistemática com o objetivo de descobrir quais os problemas de usabilidade. No entanto, para ser colocada em prática implica a participação de um conjunto de analistas, por norma especialistas, para examinar o interface e comparar com reconhecidos princípios de usabilidade (as heurísticas). Estes princípios são dez e incluem a *visibilidade do estado do sistema* (status de carregamento de página por exemplo), a *conformidade com standards* para que o utilizador não se interrogue o que certas palavras, ações e situações significam, a *prevenção do erro*, ou seja, o sistema deve permitir a recuperação de erros e notificar o utilizador quando estes aconteçam. Para além deste permitem avaliar a *flexibilidade e eficiência do uso*, a *simplicidade* e *design minimalista*, entre outras, e podem ser consultadas no site da Nielsen Norman Group³⁶

No estudo de Kani-Zahibi (2006) a avaliação heurística foi aplicada em complemento ao questionário, sendo pedido aos participantes para avaliarem a plataforma segundo os critérios de heurística, contudo os participantes neste estudo foram classificados como “user experts”. No caso de Kim e Kim (2008) foi aplicada na primeira de duas fases do estudo no estabelecimento um quadro de avaliação de usabilidade que pode ser aplicado tanto a avaliação heurística por si só, como também na realização dos testes de utilização em outros repositórios digitais. Por fim, Zhang et al. (2009) recorreram a avaliação heurística, juntamente com o feedback dos utilizadores para identificar problemas para além dos que foi possível encontrar no decorrer dos testes de utilização.

³⁶ Nielsen Norman Group. Disponível em <http://www.nngroup.com/articles/ten-usability-heuristics/> Consultado a 23 de Janeiro de 2013

Capítulo 3.

Desenho de um perfil de aplicação para dados de investigação

A produção de metadados para dados de investigação é uma atividade que tem tanto de importante como de complexa. Enquanto são necessários descritores uniformes para assegurar a interoperabilidade, existe também a necessidade de adotar descritores mais específicos para garantir a abrangência e exatidão das descrições, o que é uma das condições essenciais para a preservação a longo prazo dos dados de investigação.

Os metadados, comumente definidos com “dados sobre dados” por si só não têm grande valor, mas acrescentam valor quando associados a outro tipo de informação. Os metadados para dados de investigação (science metadata) correspondem a toda a informação necessária ao uso e interpretação dos dados recolhidos. Os metadados para efeito de descrição dos dados de investigação podem ser bastantes específicos a um determinado domínio, e por norma requerem conhecimento sobre esse mesmo domínio (Australian National Data Service, 2011).

Este capítulo descreve a experiência de elaboração de perfis de aplicação para domínios específicos da U. Porto, desenvolvidos em circunstâncias distintas. Por um lado a criação do perfil de aplicação para dois dos domínios foi precedida de uma sessão de levantamento dos procedimentos de gestão de dados, só possível devido à proximidade e disponibilidade de horário dos investigadores. Para os restantes domínios não foi possível agendar mais do que uma sessão, mas a proposta de descritores foi devidamente validado por um investigador da área.

3.1 Descrição de dados de investigação

A gestão dos dados de investigação sempre foi vista como uma questão sensível para a comunidade ligada aos metadados, com os metadados a desempenharem um papel central, Enquanto uns defendem a adoção de um certo esquema de metadados normalizado para garantir a interoperabilidade dos registos, outros argumentam que uma abordagem “*mix and match*” é a melhor alternativa - sendo os perfis de aplicação o melhor exemplo da última (Heery e Patel, 2000).

Os projetos correntes de serviços relacionados com repositórios de dados, discutidos anteriormente, estão a convergir as melhores práticas de descrição com o ciclo de vida dos dados de investigação. A iniciativa ANDS, por exemplo adotou a norma ISO RIF-CS como um formato de troca de dados, e o DANS recomenda a adoção de esquemas Dublin Core e DDI (*Data Documentation Initiative*), entre outros, dependendo do domínio. Os repositórios DataONE e EDINA implementaram planos de gestão de dados que enfatizam a necessidade dos investigadores documentarem os dados desde o momento em que os produzem, recomendando também a adoção de esquemas normalizados.

É reconhecido que o processo de criação de dados é altamente específico nos diversos domínios. A necessidade de descrever conjuntos de dados de diversas áreas levou à criação de perfis de aplicação. Estes são mais aptos a capturar o contexto do ciclo de vida de um recurso. Todavia a proliferação de esquemas de metadados específicos a determinados domínios contribui para a criação de barreiras artificiais a uma adoção generalizada de uma semântica de metadados ou à estruturação dos dados, dificultando a investigação inter e transdisciplinar (Willis, Greenberg e White, 2012).

Nos dados de investigação, cujo conteúdo pode ser muito obscuro, os metadados são essenciais para a pesquisa. À medida que mais recursos são criados em diferentes contextos torna-se complicado para um investigador ser muito específico quando procede à pesquisa de conjunto de dados. Pode mesmo dar-se o caso da existência ou localização desse conjunto de dados ser desconhecida, mas corresponder a um resultado interessante, ou mesmo surpreendente, do processo de pesquisa. No caso das publicações a indexação de conteúdo é efetiva, um conjunto de dados pode não deixar pistas em relação à sua natureza. Os metadados são, portanto, importantes para que os dados sejam convenientemente indexados e recuperados.

Para tornar os dados pesquisáveis como documentos comuns, os repositórios ou outras coleções devem prever o uso de descritores normalizados, para que os dados possam ser descritos de forma interoperável. Para além disso os valores associados a esses metadados devem ser, também esses, normalizados. Por outro lado qualquer recurso informativo que apareça numa lista de resultados de pesquisa tem de suscitar a atenção do utilizador, para assim se tornar alvo de inspeção e passível de uso. Esta é outra das funções dos metadados; uma palavra-chave concreta pode ser responsável pela inclusão de um documento numa lista de resultados. Para tal, os metadados não precisam de corresponder a uma caracterização totalmente normalizada, apresentando-se ao utilizador como um substituto do próprio objeto. Mais ainda quando consideramos conjuntos de dados, uma vez que o esforço para tornar o conteúdo perceptível é tipicamente maior que para outro tipo de conteúdos. Esforço que se agudiza quando se espera que os investigadores partilhem os dados, se entendermos

que o montante de esforço exigido para a descrição dos dados é desmotivante para os investigadores (Van House, cit. por Akmon, 2011).

3.2 Proposta de desenho de perfis de aplicação

O desenvolvimento de um perfil de aplicação, que de acordo com a definição de Heery e Pattel (2000) é “*a declaration of which metadata terms (...) user community uses in its metadata*”, que combine elementos normalizados com outros elementos específicos surge como a solução a adotar, na medida em que pode motivar os investigadores a colaborar com a descrição dos seus próprios dados. Ainda, a interoperabilidade dos metadados encoraja as comunidades de investigação à partilhar os dados digitais de um modo mais eficiente (Wira-Alam, Dimitrov e Zenk- Möltgen, 2012).

No plano da descoberta de dados dois passos são precisos: Primeiro, a interoperabilidade tem de ser garantida, e para tal podem ser usados elementos Dublin Core. Segundo, devem ser apurados um conjunto de descritores específicos para irem ao encontro das necessidades concretas da gestão de dados de investigação. Para manter a interoperabilidade do conjunto de dados o segundo ponto pode ser ignorado por soluções de colheita automáticas, ou ser transporto para o elemento «*dc.description*» durante a colheita automática, a título de exemplo.

A descrição formal das experiências é crítica para análise futura. O registo dos métodos e instrumentos utilizados para capturar os dados fornecem o contexto das experiências para terceiros interessados em verificar resultados. De outra forma esta informação contextual apenas reside na mente do(s) investigador(es) original(ais).

Os perfis de aplicação propostos neste trabalho estão elaborados de forma a satisfazer alguns dos requisitos fundamentais avançados por Willis, Greenberg e White (2012), para a documentação de dados de investigação. Estão elaborados para serem *abrangentes*, no sentido de fornecerem todos os descritores necessários, de momento, para o investigador de um dos domínios poder descrever os seus dados - os perfis estão sujeitos a um aperfeiçoamento contínuo, pelo que obviamente esta situação pode mudar. São também *simples*, permitindo a pessoas sem habilidade sofisticada de gestão de dados possam fazer essa descrição. Finalmente promovem a partilha de dados entre o grupo de investigação, fomentando a *documentação dos dados*, assim como a *descoberta e recuperação* para posterior reutilização.

Nos perfis propostos vão ser combinados elementos Dublin Core, simples e qualificados, com alguns

elementos de esquemas de metadados científicos recomendados no diretório de metadados – *Research Data Alliance* (RDA) – da página da comunidade Dublin Core para a descrição de dados de investigação (anexo 1). Por exemplo, recorre-se ao esquema *Ecological Metadata Language* (EML)³⁷, uma vez que esta norma fornece elementos para a descrição do contexto de uma investigação, que no caso particular das ciências experimentais deve incluir os métodos e instrumentos utilizados para produzir dados (Michener, 2006; Willis, Greenberg e White, 2012).

3.3 Gestão do workflow dos dados de investigação

Em 2011 uma experiência de auditoria de dados conduzida na U. Porto permitiu levantar requisitos para a gestão de dados de investigação, assim como recolher amostras de conjunto de dados de diversos domínios (Ribeiro e Fernandes; 2011). Alguns deste dados foram usados para testar o protótipo do Repositório de dados e, portanto encontram-se genericamente descritos.

O repositório encontra-se atualmente na sua fase de protótipo, estando a ser feitos esforços para encorajar os investigadores a depositar os seus dados para posterior disseminação entre parceiros. Esse processo de depósito vai ser suportado por duas plataformas integradas. A UPBOX, com um interface similar à *Dropbox*, procura facilitar o armazenamento dos dados, enquanto o DataNotes está pensado para permitir o processo de anotação de dados desde o início do ciclo de vida dos dados. Pretende-se assim conferir uma maior autonomia aos investigadores para gerirem conteúdos, sem contudo negligenciar o apoio de um curador de dados a esta atividade³⁸.

Entretanto, e no contexto deste trabalho procedeu-se a um levantamento dos processos de gestão de dados em dois grupos de investigação da FEUP. Um desses grupos está ligado ao Departamento de Engenharia Mecânica, e dedica a sua atividade à avaliação do comportamento da fratura de materiais. O outro grupo é do Departamento de Engenharia Química, e desenvolve investigação na área da química analítica.

A primeira abordagem junto dos investigadores correspondeu à recolha de informação sobre as práticas correntes do grupo de investigação. Para suportar esta atividade foi desenvolvido um guião, adaptado do *Data Curation Profile Toolkit*³⁹, desenvolvido em conjunto pelas Universidade de Perdue

³⁷ Ecological Metadata Language. Disponível em: <http://knb.ecoinformatics.org/software/eml/eml-2.1.1/index.html>
Consultado a 03 de fevereiro de 2013

³⁸ Mais detalhes sobre as plataformas UPBox e DataNotes são fornecidos no capítulo 4, relativo aos testes de utilização.

³⁹ Disponível em: <http://datacurationprofiles.org>

e Illinois (anexo 2). Durante as sessões foi discutida a possibilidade de aplicação de elementos simples e qualificados do esquema Dublin Core, em virtude do carácter documentado e normalizado do mesmo, uma vez que os investigadores se demonstraram pouco familiarizados com conceitos como “metadados” e descritores”. O resultado prático desta colaboração foi o desenvolvimento de um perfil de aplicação para cada um dos domínios, combinando elementos de descrição genéricos com outros específicos que retratam com maior exatidão o contexto de recolha de dados de cada um dos grupos.

3.4 Estudo de caso do grupo de investigação no domínio da fratura de materiais.

Neste domínio os dados são coletados através de ensaios de fratura de uma determinada amostra. A amostra é colocada numa viga dupla que exerce pressão, propagando-se uma fenda em consequência da pressão aplicada. A partir deste ponto é feita a medição da energia associada à propagação dessa fenda. O que é registado é a evolução da força e do deslocamento (fig. 7). Estes dados são tidos como importantes pois fornecem conjuntos de teste para avaliar a precisão de modelos numéricos que têm como objetivo representar o comportamento do material na realidade.

De acordo com o investigador responsável a coleta de dados pode ser dividida em duas fases. Primeiro, os dados são recolhidos por *software* apropriado, que produz um ficheiro em formato *excel*, com os valores da evolução da força e deslocamento. Posteriormente os dados são processados através de métodos analíticos próprios da área e os valores de força e de deslocamento são convertidos no valor de energia.

	A	B	C	D	E	F
1	Provete 1		Provete 2		Provete 3	
2	Força	Desloc	Força	Desloc	Força	Desloc
3	(N)	(mm)	(N)	(mm)	(N)	(mm)
4	1.2	0.003	0.4	0.003	0.92	0.003
5	1.56	0.003	1.04	0.003	1.4	0.003
6	1.2	0.003	0.88	0.003	1.2	0.003

Figura 7: Excerto de dados de um ensaio em fratura de materiais

Os dados de cada uma das etapas circulam sem qualquer limitação dentro do grupo de investigação, sem que contudo exista um protocolo de partilha formalizado. O grupo de trabalho inclui investigadores da Universidade do Porto e da Universidade de Trás-os-Montes e Alto Douro, havendo a necessidade da partilha entre os elementos ser feita de forma controlada. As limitações de acesso a

elementos externos são justificadas pela dificuldade inerentes ao processo, quer estes sejam de ordem financeira ou de ordem temporal, tendo em conta que é um tipo de atividade que consome muito tempo e recursos.

Relativamente à descrição de dados um dos pontos que foi possível aferir foi a pouca informação detalhada para o efeito. A explicação prende-se com a facilidade de interpretação por parte dos elementos que integram o grupo devido à pouca complexidade dos mesmos para pessoas com formação neste domínio. Contudo os dados, quando processados, requerem informação adicional que diz respeito às condições particulares da recolha de modo a serem percebidos na totalidade, e tal informação de contexto apenas surge descrita nos artigos publicados.

Independentemente da facilidade de interpretação e pouca informação requerida, ficou patente a pouca noção em relação à importância do uso de elementos de descrição para a preservação a longo prazo, assim como para efeitos de recuperação da informação, mesmo tendo em consideração que os dados não são do interesse de pessoas externas à área de investigação. Não existe também a adoção de esquemas de metadados, nem qualquer mandato das agências financiadoras ou editoras para um plano de gestão formal ou ingestão num repositório.

Por fim, e quando confrontado com elementos básicos e qualificados Dublin Core, o investigador responsável afirmou que apenas um número limitado de elementos seriam precisos para a documentação dos dados, interessando-lhe apenas um número restrito de descritores, como por exemplo «*title*»; «*creator*»; «*date*». Contudo, e quando questionado sobre as principais dificuldades na gestão dos dados produzidos admitiu que encontrar determinado conteúdo era uma tarefa por vezes complicada. No entanto mostrou disponibilidade para adotar outros elementos de descrição à medida que tomou consciência do problema identificado e se apercebeu das oportunidades concedidas por melhores práticas de descrição de dados.

3.4.1 Perfil de aplicação para experiências de fratura de materiais

O perfil de aplicação para experiência de fratura de materiais (prefixo fm) combina elementos simples e qualificados do Dublin Core, juntamente com elementos do esquema EML, este apropriado para a descrição das ciências experimentais, apesar de ser desenvolvido para a comunidade da Ecologia, e com elementos específicos das experiências DCB criados para o efeito e devidamente aprovados pelos investigadores da área em questão.

A título de exemplo o termo «specimen» foi adotado em vez do termo «sample» porque é um termo mais próximo ao grupo de trabalho, sendo assim mais confortável usar este conceito. De seguida foram estabelecidos elementos que cobrem as propriedades relacionadas com este «specimen» - «height»; «width»; «length»; «initialCrackLength» - o campo «specimenProperties» engloba outras propriedades complexas que devem ser preenchidas em texto livre.

O perfil inclui ainda descritores para as condições que exercem influência sobre os resultados finais, estes são relativos ao ambiente - «temperature», «moisture» - e a velocidade à qual a experiência ocorreu - «testVelocity».

Tabela 3: Perfil de aplicação “fm.”

Namespace	Description	Example
dc.creator	An entity primarily responsible for making the resource	-
dc.title	A name given to the resource	-
dc.subject	The topic of the resource	-
dc.description	An account of the resource	-
dc.date	A point of period of time associated with an event in the lifecycle of the resource	-
dc.type	The nature or genre of the resource	-
dc.medium	The material or physical carrier of the resource	-
dc.identifier	An unambiguous reference to the resource within a given context	-
dc.isReferencedBy	A resource that references, cites, or otherwise points to the described resource	-
dc.references	A related resource that is referenced, cited, or otherwise pointed to by the described resource	-
eml.methods	Actual procedures that are use in the creation or the subsequent processing of a dataset	Free text
eml.instrumentation	The Instrumentation field allows the description of any instruments used in the data collection or quality control and quality assurance.	INSTROM-1125
eml.specimen	A word or phrase describing the type of specimen collected	Wood (pinus pinaster) – a clear wood sawn from a single maritime pine
fm.specimenLength	Speciment geometric length	400 mm

fm.specimenWidth	Speciment geometric width	20mm
fm.specimenHeight	Speciment geometric height	20 mm
fm.specimenInitialCrack Lenght	The inicial crack of the DCB specimen prior to the fracture test	150mm
fm.specimenProperties	Specific properties related to the specimen, such as elastic, stress properties	Free text
fm.temperature	The ambient temperature of the location where the experiment was performed, in degrees Celsius	18°
fm.moisture	The moisture percentage at the location where the experiment was performed.	55%
fm.testVelocity	Velocity at which the experiment was performed (millimeters per minute)	3mm/m

3.5 Gestão de dados de investigação: Estudo de caso do domínio da química analítica

A sessão realizada no grupo de investigação em química analítica teve como pano de fundo os dados recolhidos durante uma campanha que teve como objetivo proceder à análise da concentrações de determinados poluentes. Estas análises são realizadas sobre diversas amostras de água e sedimentos coletadas em diversos pontos do Porto de Leixões. Através destas análises pode ser identificado o grau de contaminação dos sedimentos dragados no fundo das águas, de acordo com a legislação Portuguesa vigente. Este tipo de dados pode ser categorizado como dados de observação e experimentais, pois envolve uma forte componente de análise de laboratório a partir das amostras.

Estes são dados com um grau de importância elevado porque fornecem indicações úteis para determinar as melhores estratégias de tratamento de resíduos, atividade que pode acarretar custos elevados.

1	Amostra	Ensaio	Peak Nam	HCB	Soma PCB	Naphthalen
2	1	959	Conc. (ng/l)	nd	nd	419
3	2	960	Conc. (ng/l)	nd	nd	576
4	3	961	Conc. (ng/l)	nd	nd	438
5	4	962	Conc. (ng/l)	nd	nd	63
6	5	963	Conc. (ng/l)	nd	nd	618
7	6	964	Conc. (ng/l)	nd	nd	522
8	7	965	Conc. (ng/l)	nd	nd	733
9	8	966	Conc. (ng/l)	nd	nd	748
10	9	967	Conc. (ng/l)	nd	nd	772
11	10	968	Conc. (ng/l)	nd	426	nd

Figura 8: Excerto dados ensaio de química analítica

Segundo o entrevistado estes dados podem ser reutilizados pelos próprios produtores, que revelam interesse em manter a produção de séries temporais a partir dos mesmos, o que pode ajudar a tirar conclusões sobre a evolução da concentração de poluentes nos diferentes locais de recolha de amostras. No entanto, não vê pertinência na reutilização dos dados por parte de outros indivíduos e existem constrangimentos legais para que estes dados possam ser disponibilizados a terceiros, uma vez que foram produzidos no contexto de um acordo de prestação de serviços estabelecido com uma empresa, à qual os dados pertencem.

O ciclo de vida da produção destes dados pode ser dividida em quatro etapas. A primeira etapa corresponde à recolha dos dados sendo produzido um ficheiro por amostra com recurso a *software* apropriado (cromatógrafo). O investigador prossegue com a criação de um ficheiro único para cada relatório que pretenda desenvolver, neste ficheiro em forma de folha de cálculo são criadas duas folhas, uma correspondente aos dados em bruto e outra relativa aos dados tratados. Na quarta e última fase o investigador produz um relatório que funde os dados das fases intermédias.

Em cada uma das fases os dados ficam na posse do elemento que os produz podendo ser partilhados com outros elementos do grupo em qualquer fase, mas por norma essa partilha é feita entre o técnico de laboratório e o responsável pelo laboratório através do correio eletrónico.

A manutenção é feita com recurso aos computadores de trabalho ou unidades de armazenamento pessoal, sendo que o instrumento de análise de amostras também mantém uma cópia de segurança dos dados em bruto, existindo cópias de segurança que são atualizadas mensalmente. O investigador salientou que os dados de cada uma das fases devem ser preservados sem exceção para análises posteriores com o objetivo de detetar eventuais inconsistências nas análises originais.

No que diz respeito à anotação dos dados, esta é feita através de um relatório associado a cada uma das folhas de cálculo criadas, com base em campos pré-definidos pelo *software* utilizado, adaptados às

necessidades de anotação deste grupo de investigação. A descrição de dados é, portanto, feita sem recurso a um esquema de metadados concreto.

Os dados não foram produzidos com a finalidade de serem publicados e as entidades que financiaram a produção não determinam a necessidade de execução de um plano formal para a gestão dos mesmos, e não exigem o armazenamento dos mesmos num repositório. No entanto os investigadores estão sujeitos a proteger a confidencialidade dos dados, embora essa questão não esteja formalizada.

3.5.1 Perfil de aplicação para experiências de química analítica

O perfil de aplicação para descrição de experiências realizados no contexto de química analítica, (prefixo qa) foi elaborado para corresponder às necessidades de anotação previstas nos relatórios resultantes do processo de investigação, em que deve constar a amostra analisada, o número de amostras recolhidas, os compostos analisados e os limites legais para cada um dos poluentes. Os limites legais dos poluentes correspondem, de acordo com o anexo XX do Decreto-Lei nº236/96, à atribuição de um valor máximo admissível à qualidade de águas superficiais.

Para além da data de produção de ficheiros o perfil de aplicação deve incluir um campo para o registo da data de recolha da amostra (*sampleCollectionDate*), tido como importante pelo investigador para a leitura dos dados. Do Dublin Core qualificado adotam-se os elementos “*spatialCoverage*” para a delimitação do espaço de recolha de amostras, e “*rights*” tendo em conta as condicionantes de acesso a estes dados. Tal como no caso anterior, os metadados relativos à metodologia e instrumentos utilizados foram adaptados do esquema EML, tendo em vista a uniformidade nas descrições entre domínios.

Tabela 4: Perfil de aplicação “qa.”

Namespace	Description	Example
dc. creator	An entity primarily responsible for making the resource	-
dc.title	A name given to the resource	-
dc. subject	The topic of the resource	-
dc. description	An account of the resource	-
dc.date	A point of period of time associated with an event in the lifecycle of the resource	-
dc.type	The nature or genre of the resource	-

dc.spatialCoverage	Spatial characteristics of the resource	Porto de Leixões, Matosinhos
dc.identifier	An unambiguous reference to the resource within a given context	-
dc.isReferencedBy	A replaced resource that references, cites, or otherwise points to the described resource	
dc.references	A related resource that is referenced, cited, or otherwise pointed to by the described resource	-
dc.rights	Information about rights held in and over the resource.	
eml.methods	Actual procedures that are use in the creation or the subsequent processing of a dataset	Free text
eml.instrumentation	The Instrumentation field allows the description of any instruments used in the data collection or quality control and quality assurance.	Varian CP- 3800
qa.sample	A word or phrase describing the type of sample collected	Water
qa.compound	A word or phrase describing the elements analyzed	PCB's (bifelinos policlorados)
qa.sampleCollectionDate	Exact date in which the sample was collected	23-01-11
qa.sampleCount	Number of used samples	10
qa.compoundLegalLimit	Legal value related to sample quality	100

3.6 Descrição de dados do domínio da Dinâmica de Estruturas

A investigação no domínio da dinâmica de estruturas estuda o comportamento de uma estrutura quando sujeito a alguma ação. O conjunto de dados deste grupo de investigação foi recolhido numa campanha experimental assente em conhecimento adquirido em campanhas preliminares, em que foram feitas algumas medições com o objetivo de identificar os locais da via com uma resposta dinâmica mais significativa e estudar a melhor forma de instalar os dispositivo de medição a utilizar.

A aquisição de dados foi efetuada para a passagem de tráfego ferroviário, sendo a via instrumentada usado acelerómetros, extensómetros e dispositivos de medição de deslocamentos sem contato em diferentes pontos da via organizados por *setups*. A campanha decorreu em dois dias, o primeiro

dedicado à montagem de dispositivos. Os resultados obtidos são úteis para validar as ferramentas numéricas já desenvolvidas e para esclarecer algumas dúvidas relativamente aos elementos de projeto facultados.

08-10-2009

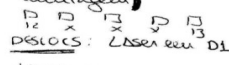
SETUP / OBS.	FICHEIRO	HORA	CATEGORIA / VELOCIDADE
~ SETUP 1 ACEL: Montagem 1 sem Δ14, Δ15 e Δ16 (por questões de tempo de montagem)  DISCOS: Laser em D1 Laser em D2 Cambria PH EXT: sim (6)	- aq_1 (Nuno) - Alfa (Diogo)	≈ 18:15	Alfa Pendular v = 225 Km/h
~ SETUP 1 ~ (igual ao anterior)	- aq_2 (Nuno) - sub urbano (Diogo)	≈ 18:20	Sub urbano v = 80 Km/h (Pára em Vila Nova de Gaia)

Figura 9: Excerto do registo de dados na campanha experimental, domínio da dinâmica de estruturas

Os dados produzidos enquadram-se na categoria de observação e experimentais, já que envolvem a recolha de amostras que posteriormente são analisadas com recurso a modelos próprios, e a preservação dos mesmos pode ser útil para a reutilização tanto por quem os produziu como por investigadores da mesma área.

O perfil de aplicação para o domínio da dinâmica de estruturas (prefixo de), utiliza elementos Dublin Core qualificado e simples, adota os elementos do esquema EML, “method” e “instrumentation”, e ainda o elemento “software” que no caso deste tipo de projeto é vital para a leitura dos dados, e portanto o grupo de investigação pode anotar, em formato de texto livre, elementos que digam respeito ao tipo, modelo e características do software utilizado. De notar que este perfil de aplicação pode evoluir para a descrição em campos separados dos atributos relativos ao software. Este perfil apresenta ainda os campos para descrição, em texto livre da configuração e posicionamento (setup) dos instrumentos utilizados.

Este perfil aproveita ainda dois elementos da especificação da *Data Documentation Initiative*

(*DDI Lifecycle – version 3.1.*)⁴⁰. Esta especificação é apropriada para a preservação de dados na áreas das ciências sociais e comportamentais, mas podem ser adotados aqui elementos para a determinação dos dias de início e fim de uma campanha de recolha de dados, informação tida como muito relevante quando se trata de anotar os dados neste domínio. Por fim foi criado o elemento “*samplingFrequency*” que tal como foi explicado por um dos elementos a trabalhar este conjunto de dados é uma informação que permite compreender o intervalo de tempo para aquisição dos dados.

Tabela 5: Perfil de aplicação “de.”

Namespace	Description	Example
dc.creator	An entity primarily responsible for making the resource	-
dc.title	A name given to the resource	-
dc.subject	The topic of the resource	-
dc.description	An account of the resource	-
dc.date	A point of period of time associated with an event in the lifecycle of the resource	-
dc.type	The nature or genre of the resource	-
dc.spatialCoverage	Spatial characteristics of the resource	Vila Franca de Xira, Azambuja
dc.identifier	An unambiguous reference to the resource within a given context	-
dc.isReferecedBy	A replaced resource that references, cites, or otherwise points to the described resource	
dc.references	A related resource that is referenced, cited, or otherwise pointed to by the described resource	-
eml.methods	Actual procedures that are use in the creation or the subsequent processing of a dataset	Free text
eml.instrumentation	The Instrumentation field allows the description of any instruments used in the data collection or quality control and quality assurance.	Acelerómetros, extensómetros
de.configuration	Defines how the instruments were configurated	Free text
de.positioning	Defines where the instruments were placed	Free text
eml.software	The software element contains general information about a software resource that is being documented.	PSD module C 10442/C10443 series

⁴⁰ Data Documentation Initiative. Disponível em: <http://www.ddialliance.org/>, Consultado a 25 de fevereiro de 2013

ddi.startDate	Start of a date range	10-08-09
ddi.endDate	End of a date range. If range is continuing on, end date can be blank	09/10/09
de.samplingFrequency	Defines the number of samples collected per unit of time (usually seconds).	2000 Hz

3.7 Descrição de dados do domínio da Ecologia Georreferenciada

Os dados produzidos pelo grupo de investigação na área da ecologia têm uma forte incidência na georreferenciação, e foram recolhidos na zona Norte de Portugal Continental. Durante a sua produção dos mesmos os investigadores recolheram amostras que posteriormente foram analisadas com recurso a modelos próprios, e por isso mesmo são dados de observação e experimentais.

Um tipo de informação comum neste grupo de investigação são os inventários de espécies, que são feitos através da identificação de amostras. A partir da dessa informação podem ser elaborados mapas de distribuição de espécies ou superfícies que representam a probabilidade de ocorrência de uma dada espécie numa certa localização.

Os dados ecológicos georreferenciados podem ser sobrepostos a outro dados com as mesmas características para estabelecer correlações, servindo os propósitos da investigação secundária em áreas como a geologia, meteorologia ou ciências sócioeconómicas.

Os investigadores nesta área são um exemplo de um grupo de investigadores com sensibilidade para a questão dos metadados. Os investigadores neste domínio têm a preocupação de catalogar e organizar os dados de forma interoperável decorrente da sua participação em projetos que têm como principal objetivo a produção e análise de dados. De momento recorrem à norma INSPIRE⁴¹ para a anotação de conjuntos de dados, uma vez que esta foi concebida para descrever conjuntos de dados no domínio geoespacial. A norma INSPIRE, parte da ISO 19115, é um perfil de metadados completo e bastante extenso. Esta norma tem um forte suporte institucional no sentido da sua adoção por parte de todas as infraestruturas da União Europeia que produzem e armazenam informação geoespacial.

Os elementos propostos para a descrição dos dados deste domínio seguem as recomendações da norma INSPIRE, e são entendidos como obrigatórios. Na tabela 6 são listados os metadados

⁴¹ INSPIRE. Disponível em: <http://inspire.jrc.ec.europa.eu/>. Consultada a 14 de março de 2013

selecionados que podem ser considerados abrangentes o suficiente para fazer essa descrição. Note-se que os campos apresentados na tabela 6 podem ser estendidos para uma descrição com maior profundidade. Por exemplo o *namespace* “*ResponsibleParty*” pode ser desdobrado em “*RespParty.OrganizationName*”; “*RespParty.Email*”; *RespParty.RespPartyRole*”.

Tabela 6: Perfil de aplicação do domínio da ecologia georreferenciada (norma INSPIRE - ISO 19-115)⁴²

Namespace	Description	Example
ResponsibleParty	Description of the organization responsible for the establishment, management, maintenance and distribution of the resource	-
TemporalExtent	The temporal extent defines the time period covered by the content of the resource	-
TempExtent.StartDate	-	2009
TempExtent.EndDate	-	2009
DateOfCreation	Date of the creation of the resource	2009
ConditionsAccessAndUse	Defines the conditions for access and use of spatial data sets and services	Non-Commercial
LimitationsPublicAccess	Provide information on any fees necessary to access and use the resource	Available to use without licensing
ResourceTitle	Characteristic, and often unique, name by which the resource is known	Geomalacus maculosus species distribution data
Keyword	An INSPIRE thematic category from a controlled list	Elevation
KeywordValue	The keyword value is a commonly used word, formalised word or phrase used to describe the subject	Elevation Solar orientation Land Use
GeographicBoundingBox	The bounding box should be expressed with westbound and eastbound longitudes, and southbound and northbound latitudes in decimal degrees	523541 - 733300- 464500 - 4513000
Lineage	Statement on process history and/or overall quality of the spatial dataset	This data is based on modelled species distribution using Maxent algorithm and software

⁴² A regulamentação de metadados INSPIRE pode ser consultada em: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32008R1205:EN:NOT>

ReferenceSystem	Description of the spatial reference systems used in the dataset	-
RefSyst.Identifier	Name of reference system	WGS 84
SpatialRepresentation	Digital mechanism used to represent spatial information	vector
GeographicExtentCode	Description of geographic area using identifiers	PT11
DistributionFormatName	Description of the format of the data to be distributed	ESRI Shapefile

3.8 Descrição de dados do domínio da Gravimetria

Entende-se como gravimetria o estudo e determinação do valor exato da aceleração da gravidade num dado ponto do planeta. Neste contexto existem vários métodos para a produção de dados, entre os quais a gravimetria aérea. O conjunto de dados aqui descrito foi recolhida numa investigação neste domínio em particular e foi produzido pelo Observatório Astronómico da Faculdade de Ciências da Universidade do Porto. O objetivo final da investigação foi a determinação de uma superfície de referência denominada geóide, assim como a produção de mapas contendo as chamadas *anomalias da gravidade*.

O processo consiste na montagem de diversos sensores numa aeronave que percorre um trajeto sobre a área determinada para a recolha de dados. Esses sensores incluem diversos sistemas inerciais, um giroscópio e recetores de GPS colocados em diferentes partes da aeronave, que registam a sua posição exata. A informação gerada é associada aos valores lidos pelo giroscópio para os tornar georreferenciados. Como o dispositivo utilizado não foi um gravímetro o grupo de investigação teve de desenvolver programas informáticos capazes de interpretar os sinais produzidos pelos sensores e convertê-los em valores numéricos passíveis de serem usados neste contexto. Para tal são usadas curvas de calibração dependentes do sistema inercial considerado.

Estes tipo de dados pode ser classificados como experimentais e de observação, uma vez que a análise envolveu o uso de *software* construído à medida pelo próprio grupo.

488688.000000	38.76030000	-27.08409800	110.767	-53.750371
488689.000000	38.76029600	-27.08410600	110.813	-67.168032
488690.000000	38.76029500	-27.08410900	110.846	-80.584969
488691.000000	38.76029300	-27.08410800	110.871	-93.994527
488692.000000	38.76029200	-27.08410600	110.892	-107.391006
488693.000000	38.76029100	-27.08410500	110.911	-120.765010
488694.000000	38.76029000	-27.08410400	110.929	-134.106571
488695.000000	38.76029000	-27.08410300	110.944	-147.404566
488696.000000	38.76028900	-27.08410300	110.958	-160.644978
488697.000000	38.76028900	-27.08410300	110.969	-173.814814
488698.000000	38.76028900	-27.08410300	110.979	-186.901223

Figura 10: Excerto de dados do domínio Gravimetria

Os cabeçalhos dos valores representados em cada uma das colunas da figura 10 são relativos ao “tempo em GPS”, a “latitude”, a “longitude”, a “altitude” e a “gravidade específica” em determinado momento, respetivamente. Desta forma os descritores a aplicar devem incluir esta informação contextual.

Os dados de gravimetria podem servir os interesses de diversas comunidades, nomeadamente informáticos interessados em utilizar os dados para a construção de soluções de navegação, geólogos que requeiram informação sobre a constituição do solo num ponto concreto e investigadores na área da tectónica que pretendam estudar zonas de falhas. Os próprios investigadores do Observatório tendem a usar os dados para comparar e validar resultados, que podem servir como teste a novos algoritmos. Para além disso, como muitos destes dados são georreferenciados podem ser cruzados outros dados georreferenciados para deduzir correlações entre grandezas de diferentes domínios.

O perfil de aplicação proposto para o domínio da Gravimetria (prefixo grav.), agrega sobretudo elementos do *Content Standard for Geospatial Metadata* (CSDGM)⁴³, esquema que fornece a terminologia e as definições para a documentação de dados geoespaciais, sendo um esquema recomendado no diretório de metadados da comunidade Dublin Core. Os elementos escolhidos surgem como limitadores de espaço e tempo. Assim para determinar os limites de cobertura em termos de latitude, longitude, recorreu-se aos campos “*BoundingCoordinate*”, e ao registo dos valores mínimos e máximos de altitude. No caso da cobertura temporal, e como neste domínio as anotações temporais podem ser detalhadas ao segundo, aplica-se o “*TimePeriodInformation*”, onde se pode anotar desde do ano ao momento exato do início e fim do evento.

⁴³ Content Standard for Geospatial Metadata. Disponível em: <http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata> Consultado a 14 de março de 2013

Tabela 7: Perfil de aplicação “grav.”⁴⁴

Namespace	Description	Example
dc.creator	An entity primarily responsible for making the resource	-
dc.title	A name given to the resource	-
dc.subject	The topic of the resource	-
dc.description	An account of the resource	-
dc.date	A point of period of time associated with an event in the lifecycle of the resource	-
dc.type	The nature or genre of the resource	-
dc.spatialCoverage	Spatial characteristics of the resource	Açores
dc.identifier	An unambiguous reference to the resource within a given context	-
dc.isReferencedBy	A replaced resource that references, cites, or otherwise points to the described resource	-
dc.references	A related resource that is referenced, cited, or otherwise pointed to by the described resource	-
eml.methods	Actual procedures that are use in the creation or the subsequent processing of a dataset	Free text
eml.instrumentation	The Instrumentation field allows the description of any instruments used in the data collection or quality control and quality assurance.	GPS
csgdm.northBoundingCoordinate	Northern-most coordinate of the limit of coverage expressed in latitude	30.54
csgdm.southBoundingCoordinate	Southern-most coordinate of the limit of coverage expressed in latitude	30856
csgdm.eastBoundingCoordinate	Eastern-most coordinate of the limit of coverage expressed in longitude	-75,63
csgdm.westBoundingCoordinate	Western-most coordinate of the limit of coverage expressed in longitude	-75625
csgdm.altitudeMaximum	The maximum elevation extent of coverage	11004
csgdm.altitudeMinimum	The minimum elevation of coverage	110767
csgdm.timePeriodInformation	Information about the date of time of an event	-

⁴⁴ O perfil de aplicação proposto para a área da Gravimetria, embora validado por um investigador da área, encontra-se numa fase preliminar e requer um aprofundamento ao nível dos descritores a aplicar, em conformidade com os requerimentos expressos na GSGDM, para a criação de perfis para um domínio ou organização específica: http://www.fgdc.gov/metadata/documents/workbook_0501_bmk.pdf

csgdm.begginingDate	The first year (and optionally month, or month and day) of the event	23-09-92
csgdm.endingdate	The last year (and optionally the month, or month or day) of the event	13-11-92
csgdm.begginingTime	The first hour (an optionally minute, or minute and second) of the day for the event	09:21:45
csgdm.endingTime	The last hour (an optionally minute, or minute and second) of the day for the event	11-22-32
grav.gpsTime	GPS time in seconds of week	448710.0
grav.specifGravity	Specific gravity in MilliGal = 10^{-5} m.s ⁻² . Differential to the specific gravity field.	-356.189613
grav.samplingFrequency	Defines the number of samples collected per unit of time (usually seconds).	50 hz

Capítulo 4.

Testes de utilização do Repositório de dados da Universidade do Porto e das plataformas UPBox e DataNotes

Para a realização do estudo do comportamento dos utilizadores ao usar as ferramentas Repositório de dados da Universidade do Porto, UPBox e DataNotes, foram feitos dois testes de utilização. O primeiro teste serviu para avaliar a interação dos investigadores com o protótipo do Repositório de dados da Universidade do Porto. O segundo foi realizado nas plataformas UPBox e DataNotes e teve como desígnio identificar possíveis limitações no processo de depósito de conjuntos de dados, assim como para apresentar estas ferramentas como uma solução integrada para a gestão dos dados de investigação. Para cada um dos testes de utilização foi concebido um guião de apoio, que podem ser consultados nos anexos 3 e 4.

De seguida é descrita a metodologia que justifica as opções tomadas para levar a cabo estes testes.

Participantes

No total colaboraram nestes testes sete investigadores. O painel de participantes foi composto por um investigador de cada um dos Departamentos de Engenharia Mecânica, Química, Civil e Informática da FEUP, um investigador do Observatório Astronómico da U. Porto, e outro do Departamento de Biologia da Faculdade de Ciências da U. Porto. O painel de participantes ficou completo com a colaboração de um investigador da Universidade de Valência, Espanha, com notório interesse na investigação ligada à temática dos repositórios digitais.

De acordo com Jakob Nielsen (2003) quanto maior for o número de participantes maior é a percentagem de erros encontrados, no entanto, segundo o mesmo, para um teste de utilização ter real valor terá de ser efetuado com um mínimo de cinco participantes. Além disso é importante realizar os testes individualmente e deixar os participantes resolver as tarefas por conta própria de modo a não condicionar os resultados (Nielsen, 2003).

Recolha de dados

A recolha de dados foi dividida em três momentos. Na primeira fase foram apresentados aos participantes os objetivos do estudo, sendo-lhes pedido que explorassem livremente o Repositório de dados (aproximadamente cinco minutos), com o intuito de se familiarizarem com o protótipo. Nesta fase recolheram-se as primeiras impressões gerais, com a colocação de questões preliminares para aferir as rotinas de procura de informação científica dos participantes e se estes se considerariam como utilizadores experientes em ferramentas similares. No caso do teste das plataformas UPBox e DataNotes não se procedeu à exploração livre, uma vez que os processos nestas plataformas, particularmente no caso da UPBox, se desenrolam de forma linear.

A segunda fase implicou a realização de tarefas pré-definidas de acordo com os casos de uso de cada uma das plataformas. Durante a execução de tarefas foi pedido aos participantes para verbalizarem as suas ações e pensamentos (Ferreira e Pithan, 2005; Shiri et al., 2009; Zhang et al., 2009; Kim e Kim, 2008), para registo da experiência. Aos participantes foi transmitida a noção de que o teste se destinava a avaliar o desempenho das plataformas e não o desempenho dos próprios. Para além disso definiram-se indicadores de performance (*KPI – key performance indicator*) como parâmetros de avaliação para cada uma das tarefas a realizar. Por fim, procedeu-se a uma entrevista complementar com o objetivo obter impressões gerais dos participantes em relação à experiência, pontos fortes e fracos do sistema, identificar prioridades e recolher sugestões de funcionalidade a desenvolver nas plataformas.

A avaliação heurística embora considerada como abordagem complementar aos testes de utilização, não pôde ser aplicada a este trabalho porque, tal como é referido no fim do capítulo dedicado aos estudos de utilização, implicaria recorrer a especialistas na Interação Humano-Computador, o que não foi possível durante este processo. No entanto alguns dos princípios da avaliação heurística são aplicados de forma indireta, nomeadamente, na conversa pós-teste, e da própria observação do analista (pessoa encarregada do teste de utilização). A recolha dos dados foi feita manualmente, mesmo tendo em consideração a possibilidade de recurso a *software* apropriado para este tipo de teste. No entanto, e como os testes foram realizados no local de trabalho e computadores pessoais dos investigadores, para conforto dos mesmos, a instalação de *software* em cada uma das máquinas revelar-se-ia inconveniente. O registo de dados foi feito diretamente nas tabelas presentes no guião de teste, sendo feita a gravação em áudio, com o devido consentimento dos participantes, e o uso de um cronómetro.

Variáveis do estudo

A literatura referente aos testes de utilização no contexto de bibliotecas digitais tende a avaliar o desempenho deste tipo de plataforma sob o ponto de vista da usabilidade. Neste contexto têm sido aplicados a larga escala os cinco atributos de qualidade propostos por Jakob Nielsen (2003),⁴⁵ originalmente concebidos para o teste de interfaces de *Websites*: aprendizagem, eficiência, erro, satisfação e memorização.

Para o desenvolvimento do guião de apoio à realização dos testes de utilização aplicaram-se os atributos de qualidade listados (tabela 8), excluindo a memorização, uma vez que esta implica a repetição do teste num período de tempo alargado após a realização do teste original, para o qual não existe disponibilidade no âmbito deste trabalho (limitação temporal).

Tabela 8: Descrição dos atributos de qualidade de Nielsen (2003)

Aprendizagem	A facilidade com que os utilizadores resolvem as tarefas na primeira vez que utilizam o sistema.
Eficiência	A rapidez de execução da tarefa.
Erro	Número de erros cometidos pelo utilizador, gravidade, e superação de problema.
Satisfação	O quão agradável é usar o sistema.

Complementarmente é utilizado como modelo de referência a proposta de Judy Jeng (2005), para a avaliação da usabilidade de bibliotecas digitais (fig. 11). Este modelo, desenvolvido após a revisão de estudos de utilização entre 1989 e 2005, está construído a partir da definição de usabilidade da norma ISO 9241-11, com a inclusão do conceito “aprendizagem”.

O modelo de avaliação de estudos de usabilidade fornece as pistas para o estabelecimento de critérios, ou indicadores de performance para os testes a desenvolver. De acordo com Jeng (2005) a *eficácia* avalia se no seu todo o sistema fornece informação e funcionalidades de um modo eficaz. A *eficiência* avalia se o sistema permite resolver um problema com desenvoltura, podendo ser medida, por exemplo, através do tempo despendido a realizar determinada tarefa (Pithan e Ferreira, 2005; Shiri et al, 2009; Zhang et al., 200).

⁴⁵ Fonte: Nielsen, Jakob. 2003. “Usability 101: Introduction to usability”, Useit.com: Usable Information Technology. Disponível em <http://www.useit.com/alertbox/20030825.html> (Consultado a 03 de dezembro de 2012)

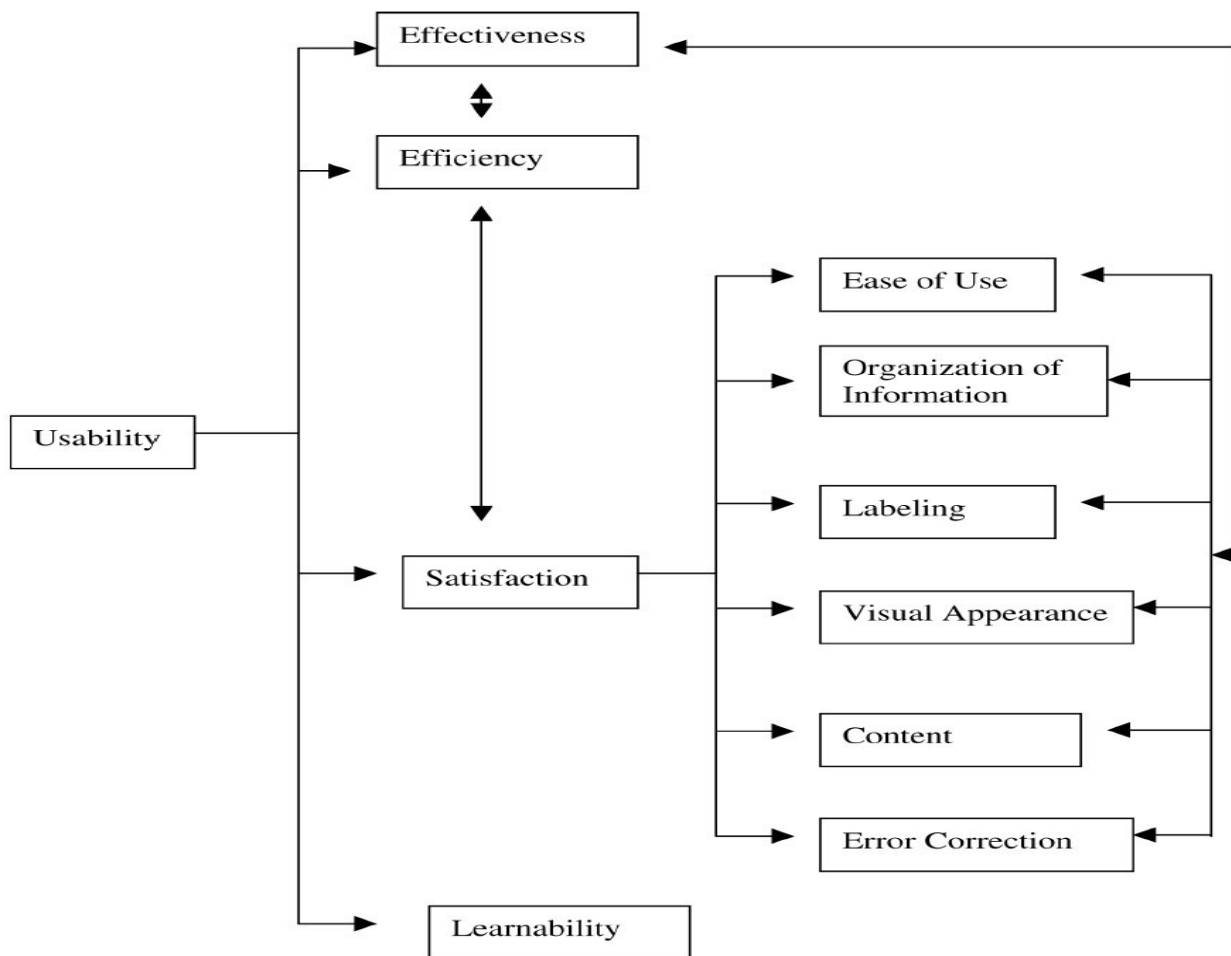


Figura 11: Modelo de avaliação da usabilidade de bibliotecas digitais (Jeng, 2005).

A *satisfação*, como variável subjetiva, é mensurável de preferência através de entrevistas (Norberg et al., 2005; Kim e Kim, 2008; Shiri et al., 2009; Ferreira e Pithan, 2005) ou com recurso a uma escala de *Likert* (Zhang et al., 2009; Shiri et al., 2009) e avalia, entre outros aspetos, a facilidade de uso percebida pelo utilizador e o desenho amigável de um interface. A *aprendizagem* diz respeito ao esforço despendido pelo utilizador para dominar o sistema e pode ser medida através da facilidade com que executa uma tarefa, da comparação com outros sistemas semelhantes, ou na execução dos procedimentos numa experiência posterior ultrapassando as dificuldades sentidas na primeira interação.

4.1 Teste de utilização do Repositório de dados da Universidade do Porto

O Repositório de dados da Universidade do Porto foi desenhado e implementado como uma extensão da aplicação em código aberto *Dspace*⁴⁶, uma das plataformas mais populares para a criação de repositórios institucionais⁴⁷. Um dos argumentos favoráveis à adoção do Dspace como solução para o protótipo deste repositório é a adoção da mesma ferramenta nos repositórios já implementados na Universidade do Porto.



Figura 12: Interface inicial do Repositório de dados da Universidade do Porto

O objetivo deste repositório é providenciar funcionalidades de preservação de dados relevantes – armazenamento, descrição, facilidade de partilha e pesquisa, interrogação de subconjuntos de dados e o *download* total ou parcial dos mesmos. No âmbito deste trabalho não se procede a uma descrição especificada desta plataforma, para tal pode ser consultada a documentação disponibilizada pelo *Dspace*⁴⁸.

⁴⁶ Disponível em: <http://www.dspace.org/>; Consultado 23 de abril de 2013

⁴⁷ Segundo dados da OpenDoar de junho de 2013, cerca de 41,3% de um total de 2293 Repositórios com software Open Access usam o DSpace como solução.

⁴⁸ Dspace 1.7.2 Documentation. Disponível em: <https://wiki.duraspace.org/display/DSDOC17/DSpace+Documentation>, consultado a 18 de abril de 2013

O Repositório de dados da Universidade do Porto apresenta, enquanto protótipo, um conjunto de funcionalidades do ponto de vista do utilizador que permitem dar resposta aos seguintes casos de uso:

1. **Explorar dados** – O utilizador pode pré-visualizar os dados contidos num conjunto de dados sob a forma de tabela interativa. Esta tabela listará uma ou mais tabelas contidas no ficheiro e oferecerá funções de filtragem de dados.

2. **Descarregar dados** – Esta funcionalidade permite ao utilizador fazer o *download* da totalidade ou de uma parcela de dados contidos no ficheiro original, sob a forma uma folha de cálculo.

3. **Filtrar dados** – O utilizador tem a possibilidade de filtrar os dados que lhe são apresentados na tabela interativa de exploração de dados.

4. **Obter *link* para dados** – O investigador poderá ter o interesse em enviar um subconjunto de dados a um terceiro. Esta funcionalidade permite ao investigador executar uma consulta sobre um conjunto de dados do qual seja autor e produzir um identificador (*link*) que pode ser partilhado por email.

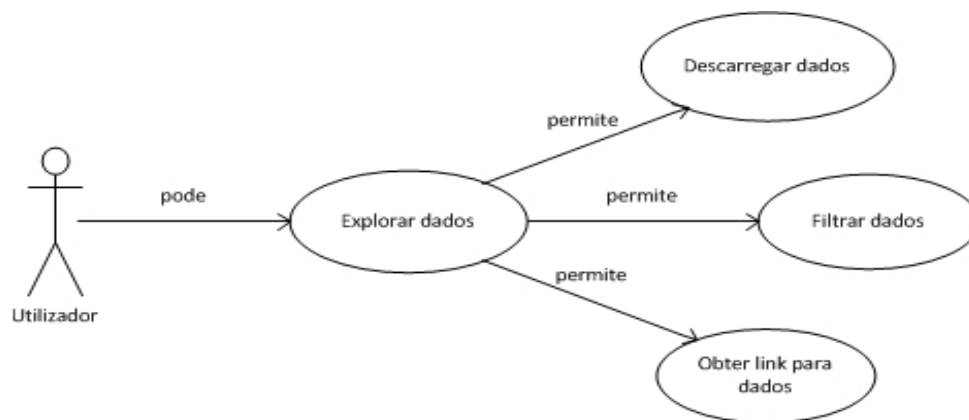


Figura 13: Casos de uso do Repositório de dados da Universidade do Porto

Tal como foi referido na descrição do processo de recolha de dados as tarefas propostas têm como base os casos de uso do ponto de vista do utilizador. Todavia a possibilidade de obter *link* para dados não figura como uma das tarefas propostas. O motivo dessa exclusão é a necessidade de proceder à exploração do repositório através das credenciais do analista, como utilizador previamente registado no sistema, mas não proprietário dos dados.

Outra das funcionalidades promovidas nesta extensão do *DSpace* para a exploração de dados é a página de pesquisa por conjunto de dados e pode ser utilizada para recuperar tabelas de dados que

tenham colunas ou metadados específicos. O acesso a esta página é feito no canto superior esquerdo da página inicial do Repositório de dados, através do botão “*Search Dataset*”. Esta funcionalidade não foi incluída nas tarefas propostas, optando-se por se fazer uma demonstração desta funcionalidade junto dos participantes.

Na próxima secção são descritos os resultados dos testes realizados nesta plataforma.

4.1.1 Resultados do teste de utilização do Repositório de dados da Universidade do Porto

Antes de proceder às tarefas propostas para o teste de utilização do protótipo do Repositório de dados, foram colocadas breves questões aos participantes de modo a aferir as suas rotinas de pesquisa de informação e o grau de experiência no uso deste tipo de repositórios. Aos mesmos foi pedido que fizessem uma exploração livre da plataforma para que fossem identificadas as suas estratégias de pesquisa.

O **participante A** usa como fonte de referência exclusiva o *Science Direct*. Este participante não recorre por norma a repositórios para efeito de recuperação de dados ou armazenamento, e portanto não é um utilizador experiente deste tipo de ferramenta. Este destacou o carácter *amigável* do Repositório de dados, que para si engloba o essencial para uma pesquisa. O **participante B** satisfaz as suas necessidades de informação através do Google e não utiliza repositórios de dados, não sendo um utilizador experiente. Após a primeira interação com a plataforma avaliou-a como simples, contudo criticou o facto de não haver destaque para os botões mais relevantes, o que condiciona a navegação. Mencionou ainda que por não saber o que podia encontrar no repositório seria interessantes ser apresentada uma lista de conteúdos genérica para ir aprofundando a pesquisa. Já o **participante C** não tem por hábito recorrer a serviços de informação científica, nem com a intenção de recuperar nem no sentido de armazenar dados de investigação, sendo um utilizador sem experiência. No geral acha que a plataforma é de fácil navegação e tem uma boa estrutura apesar de não ter ficado totalmente convencido com a relação entre comunidades e coleções.

O **participante D** utiliza habitualmente o Google Académico e o site da biblioteca da FEUP como fontes primordiais de pesquisa. Este participante é responsável pelo serviço de gestão de dados de investigação do grupo no qual está inserido, razão pela qual se considera um utilizador experiente quando se trata de repositórios digitais. Após a abordagem livre do protótipo a testar indicou que se trata de uma ferramenta que lhe é muito familiar, sendo uma rotina para si lidar com metadados e

funcionalidades de pesquisa. Por isso mesmo avaliou a plataforma como bastante simples e fácil de utilizar. No caso do **participante E** as necessidades de informação podem ser colmatadas numa primeira fase a partir de uma pesquisa no Google para identificar artigos especializados na área de investigação, apesar de já ter consultado repositórios para recolher dados, e por isso considera-se um utilizador com um grau de experiência normal neste tipo de ferramentas. Na sua perspetiva o protótipo é de utilização simples e semelhante a algumas ferramentas anteriormente consultadas.

O **participante F** é um caso de utilizador com elevada experiência a nível de repositórios de dados, tratando-se de um investigador com particular interesse neste domínio. Na opinião do mesmo o Repositório de dados está construído de acordo com os padrões para este tipo de ferramentas. As rotinas de pesquisa do **participante G** envolvem bases de dados especializadas, como o *Scopus* e o Google Académico. Atualmente não tem dados publicados em nenhum repositório nem utiliza nenhum com o objetivo de os recolher. Por estes motivos é um utilizador inexperiente nestas ferramentas. A primeira impressão sobre este protótipo é a noção da utilidade desta para tornar dinâmica a troca de dados entre comunidades.

Durante a exploração livre do Repositório os vários participantes adotaram estratégias diferentes. O **participante A** optou por começar a interrogar o sistema através da pesquisa avançada, enquanto o **participante G** o fez pela caixa de pesquisa simples. Os restantes optaram pela navegação de conteúdos. Apesar da exploração ao Repositório ter sido completa no geral, nenhum dos participantes procedeu à exploração dos dados, ora por não entrarem na opção “*Search Dataset*”, ora por não entrarem na exploração dos dados quando o nível de profundidade da navegação os fez chegar a ficheiros que tinham essa opção ativa.

Tendo em conta que a exploração dos dados é uma funcionalidade importante nesta ferramenta, será conveniente destacar essa oportunidade numa posição central da página inicial do Repositório de dados.

Outro aspeto a desenvolver é a capacidade do sistema sugerir termos indexados no momento em que os utilizadores começam a introduzir uma expressão de pesquisa. A verificação desta necessidade surgiu durante a interação do **participante G**, cuja estratégia de pesquisa, quando não conhece a coleção, passa pela interrogação do sistema com a introdução das primeiras letras de um termo.

Após a primeira fase de exploração livre do Repositório de dados foram propostos aos participantes objetivos de pesquisa concretos, sob a forma de tarefas, que passam a ser apresentados.

Tarefa 1 – Visualizar coleção de dados

A primeira tarefa proposta aos participantes consistiu na visualização de uma das coleções criadas na plataforma. O Repositório de dados da Universidade do Porto, de acordo com a estruturação do *Dspace*, está organizado em comunidades e coleções, correspondendo as comunidades aos diferentes domínios de investigação da U. Porto e as coleções à produção de conteúdos produzidos numa determinada comunidade.

Para executar a tarefa os participantes deveriam navegar através de duas opções presentes na página inicial do Repositório de dados, pela exploração do botão “Communities & Collections”, ou entrar diretamente na comunidade “Communities in Dspace” (fig 14). Após este passo surge uma página onde o utilizador pode escolher qual das coleções pretende explorar. A tarefa seria considerada como concluída quando visualizada a *home page* de uma coleção.

The screenshot shows the 'Aereal Gravimetry Runs' collection page. At the top left, there are navigation links: 'U.PORTO ScienceData Repository >' and 'Astronomic Observatory (FCUP) >'. The main heading is 'Aereal Gravimetry Runs' with the subtitle 'Collection home page'. Below this is a search bar with a dropdown menu set to 'Aereal Gravimetry Runs', a 'Search for' input field, and a 'Go' button. There are also buttons for 'Issue Date', 'Author', 'Title', and 'Subject', with an 'or browse' option. A 'Subscribe' button is present for email notifications, and a 'View Statistics' button is at the bottom of the search area. The main text describes the collection, mentioning it was created by Prof. Luísa Bastos and is used for determining a surface called geoid. At the bottom right, there are 'RSS Feeds' for versions 1.0 and 2.0. The page is marked as 'Open Access'.

Figura 14: Página da coleção “Aereal Gravimetry Runs” (Repositório de dados)

O **participante A** procurou chegar à coleção que lhe foi proposta através da pesquisa pelos termos exatos e invariavelmente a pesquisa não lhe produziu resultados, o que levou este utilizador a regressar à página inicial por três vezes para retomar a tentativa. Durante a tarefa admitiu não estar a ver - “*não está óbvio*” - e formulou um pedido de ajuda - “*devia aparecer aqui certo?*”. No fim desabafou que encontrar a coleção - “*não foi fácil*”.

O **participante B** também começou por interrogar a caixa de pesquisa, regressando de imediato à página inicial para completar a tarefa. No entanto comentou que a relação entre comunidades e coleções não lhe pareceu óbvia e que ficou com a sensação que depois de entrar na coleção esta estaria vazia. Os **participantes F e G** também ficaram com a sensação da falta de conteúdos da coleção. O **participante C** cometeu um erro do qual recuperou imediatamente, o que não comprometeu o sucesso da execução da tarefa.

Tabela 9: Resultados tarefa 1 (Repositório de dados)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	3	1	9	-	143 s
B	1	-	1		23 s
C	-	-	1	-	18 s
D	-	-	-	-	14 s
E	-	-	-	-	08 s
F	-	-	-	-	16 s
G	-	-	-	-	45 s

média- 38.14 s

Face aos resultados pode-se concluir que chegar à página de uma coleção é uma tarefa que pode ser executada de forma eficiente, como demonstra o tempo médio para terminar a tarefa, inflacionado pelo tempo despendido pelo **participante A**. Contudo a eficácia da mesma pode ser questionada uma vez que alguns participantes ficaram com a impressão que a coleção estava vazia. Por esta razão a página de coleção deve ser editada para permitir a visualização de conteúdos.

Tarefa 2 – Visualizar ficheiros de um autor

A segunda tarefa consistiu na pesquisa de ficheiros de acordo com a autoria. Note-se que a proposta de tarefa determinou qual o autor do qual os participantes deveriam recuperar ficheiros, implicando que os participantes sabiam de antemão que o autor estava na base de dados do Repositório. Assim os participantes podiam utilizar a caixa de pesquisa para interrogar o sistema ou então recorrer à navegação por autoria.

A tarefa foi dada como terminada quando os participantes recuperaram todos os ficheiros sugeridos pelo analista.

Browsing by Author

Jump to: [0-9](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)
 or enter first few letters:

Order: Results/Page

Showing results 1 to 6 of 6

[Alves, Arminda](#)
[Alves, Moreira, José Luís Arminha](#)
[Bastos, Luísa](#)
[Deurloo, Richard](#)
[Moreira, José Luís](#)
[Moura, Marcelo](#)

Showing results 1 to 6 of 6

Issue Date	Title	Author(s)
4-Oct-2011	Base Data (Inertial Navigation System) from the Aereal Gravimetry Run over the Azores (1992)	Bastos, Luísa ; Deurloo, Richard
4-Oct-2011	Base Data (Reference Stations) from the Aereal Gravimetry Run over the Azores (1992)	Bastos, Luísa ; Deurloo, Richard
4-Oct-2011	Configuration files for the data processing program	Bastos, Luísa ; Deurloo, Richard
4-Oct-2011	Final Results of the gravimetry runs	Bastos, Luísa ; Deurloo, Richard
4-Oct-2011	GPS Solutions used in data processing	Bastos, Luísa ; Deurloo, Richard
4-Oct-2011	Processed GPS data used during the Aereal Gravimetry Run (1992)	Bastos, Luísa ; Deurloo, Richard

Showing results 1 to 6 of 6

Figura 15: Navegação por autor (Repositório de dados)

Os resultados confirmam que esta tarefa foi executada sem indicadores de dificuldade. O participante B comentou que a pesquisa por autor é fácil.

Tabela 10: Resultados tarefa 2 (Repositório de dados)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	-	-	-	05 s
B	-	-	-	-	04 s
C	-	-	-	-	17 s
D	-	-	-	-	12 s
E	-	-	-	-	14 s
F	-	-	-	-	10 s
G	-	-	-	-	19 s

média- 11.57 s

Tarefa 3 -Visualizar ficheiros de um assunto

Na tarefa 3 foi pedido que os participantes recuperassem os ficheiros relativos a um assunto. As condições de realização desta tarefa foram em tudo semelhantes ao descrito na tarefa anterior, sendo que a navegação desta vez devia ser feita por assunto (*subject*), um dos classificadores previstos no repositório.

Browsing by Subject

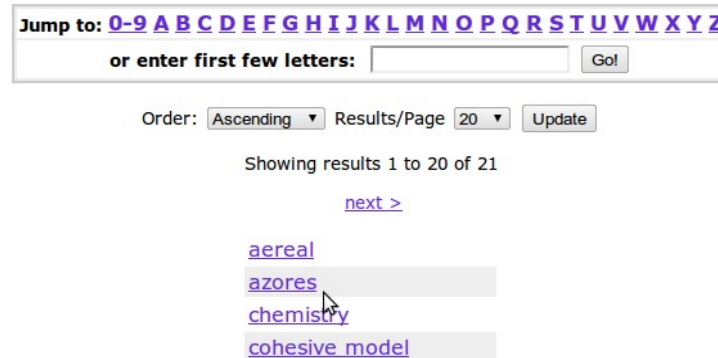


Figura 16: Navegação por assunto (Repositório de dados)

Tal como na tarefa anterior não foram registadas dificuldades para visualizar os ficheiros do assunto proposto.

Tabela 11: Resultados tarefa 3 (Repositório de dados)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	-	-	-	07 s
B	-	-	-	-	09 s
C	-	-	-	-	11 s
D	-	-	-	-	09 s
E	-	-	-	-	13 s
F	-	-	-	-	12 s
G	-	-	-	-	18 s

média- 10 s

Tarefa 4 – Ordenação de ficheiros de coleção por ordem descendente de acordo com a data de emissão

A tarefa 4 tinha como objetivo a pesquisa de ficheiros de acordo com a data de emissão (*issue date*), seguindo a lógica de navegação referidas nas tarefas 2 e 3. Contudo foi pedido ao participantes que ordenassem a lista de resultados por ordem descendente, isto é, do ficheiro mais recente para o mais antigo, ao invés da configuração definida do mais antigo para o mais recente. Os participantes deveriam manipular a barra de edição de preferências e selecionar “*In order: Descending*” e terminar ao carregar no botão “*Update*”.

Sort by: In order: Results/Page Authors/Record:

Showing results 1 to 10 of 10

Issue Date	Title	Author(s)
11-Jun-2013	Dados DCB Madeira	Moura, Marcelo
4-Jun-2013	Relatório de análises de PCBs, PAHs e HCB de amostras de águas provenientes do Porto de Leixões – 4ª Campanha – Janeiro de 2011	Alves, Moreira, José Luís Arminha
3-Jun-2013	Relatório de análises de PCBs, PAHs e HCB de amostras de águas provenientes do Porto de Leixões – 4ª Campanha – Janeiro de 2011	Alves, Arminda; Moreira, José Luís
3-Jun-2013	Resultados amostras janeiro 2011	Alves, Arminda; Moreira, José Luís
4-Oct-2011	Configuration files for the data processing program	Bastos, Luísa; Deurloo, Richard
4-Oct-2011	Final Results of the gravimetry runs	Bastos, Luísa; Deurloo, Richard
4-Oct-2011	Base Data (Reference Stations) from the Aereal Gravimetry Run over the Azores (1992)	Bastos, Luísa; Deurloo, Richard

Figura 17: Ordenação dos ficheiros por ordem descendente (Repositório de dados)

Os **participantes B** e **D** regressaram à página inicial da plataforma porque não visualizaram quaisquer ficheiro na página da coleção proposta, o que os levou a pensar que a coleção estava vazia. O erro registado pelo **participante D** foi ter selecionado na caixa “*sort by:*” a opção de data de submissão.

Tabela 12: Resultados tarefa 4 (Repositório de dados)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	-	-	-	42 s
B	2	-	-	-	59 s
C	-	-	-	-	12 s
D	3	-	1	-	31 s
E	-	-	-	-	10 s
F	-	-	-	-	12 s
G	-	-	-	-	26 s

média- 27.42 s

Após avaliar os resultados da tarefa 4, não podem ser atribuídos problemas à ordenação dos conteúdos. Os registos efetuados, exetando o erro do **participante D**, e a maior duração para terminar a tarefa pelos **participantes A e B**, reforçam a identificação do problema da tarefa 1.

Tarefa 5 - Visualizar ficheiros relativos a um item, mostrar registo de metadados completo e abrir um dos ficheiros.

Esta tarefa constituiu um desafio mais complexo, estando desenhada em três momentos. Primeiro foi pedido aos participantes que seleccionassem um item concreto, previamente definido pelo analista. Os participantes tiveram a liberdade de escolher a estratégia de recuperação do objeto, tanto utilizando o motor de busca ou por navegação no *browser*. Em seguida teriam de visualizar o registo completo de metadados (*Show full item record*), tendo em conta que o interface apresenta por omissão uma visão parcial dos mesmos. A tarefa finalizava com a abertura de um dos ficheiros associados ao item, ao carregar no botão *View/Open*.

Please use this identifier to cite or link to this item: <http://hdl.handle.net/123456789/8>

Title: Final Results of the gravimetry runs

Authors: [Bastos, Luísa](#)
[Deurloo, Richard](#)

Keywords: aerial
gravimetry
final
results

Issue Date: 4-Oct-2011

Abstract: Final Georeferenced Data, combining specific gravity values with GPS coordinates of the gathering location.

URI: <http://hdl.handle.net/123456789/8>

Appears in Collections: [Aereal Gravimetry Runs](#)

Files in This Item:

File	Description	Size	Format	
grv03.dat	Gravity Values for each Instant and coordinates of the measuring locations	1.06 MB	Text	View/Open Explore Data
Readme.txt	Description of the method followed during the combination of the specific gravity values and the processed GPS coordinates	1.85 kB	Text	View/Open

Show full item record

Figura 18: Visualização de ficheiros associados a um item (Repositório de dados)

Full metadata record

DC Field	Value	Language
dc.contributor.author	Bastos, Luísa	-
dc.contributor.author	Deurloo, Richard	-
dc.date.accessioned	2011-10-04T15:54:05Z	-
dc.date.available	2011-10-04T15:54:05Z	-
dc.date.created	1997-10-03	-
dc.date.issued	2011-10-04	-
dc.identifier.uri	http://hdl.handle.net/123456789/8	-
dc.description.abstract	Final Georeferenced Data, combining specific gravity values with GPS coordinates of the gathering location.	en_US
dc.language.iso	other	en_US
dc.subject	aereal	en_US
dc.subject	gravimetry	en_US
dc.subject	final	en_US
dc.subject	results	en_US
dc.title	Final Results of the gravimetry runs	en_US
dc.type	Dataset	en_US

Appears in Collections: [Aereal Gravimetry Runs](#)

Files in This Item:

File	Description	Size	Format	
grv03.dat	Gravity Values for each Instant and coordinates of the measuring locations	1.06 MB	Text	View/Open Explore Data
Readme.txt	Description of the method followed during the combination of the specific gravity values and the processed GPS coordinates	1.85 kB	Text	View/Open

Show simple item record

Figura 19: Registo de metadados completo (Repositório de dados)

Os **participantes B, C, D e G** cometeram erros menores. No caso do **participante C** deu a entender que o erro se deveu à indevida interpretação sobre o que lhe tinha sido pedido para fazer, tendo

perguntado - “é isto?”, quando clicou sobre um botão irrelevante. No caso dos **participantes B e D** o erro ficou registado porque seleccionaram a opção “*view statistics*”, que surge por baixo da opção que lhes permitia ver o registo completo de metadados. O **participante E**, embora sem cometer erros, demorou a abrir o registo de metadados completo (86 segundos para completar a tarefa), afirmando não ter percebido bem onde executar essa ação. O **participante G** afirmou que o botão - “*não está evidente.*”

Tabela 13: Resultados tarefa 5 (Repositório de dados)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A					59 s
B	-	-	1	-	39 s
C	-	-	1	-	13 s
D	-	-	1	-	20 s
E	-	-	-	-	86 s
F	-	-	-	-	23 s
G	-	-	2	-	74 s

média- 44.85 s

Os **participantes B e G** sugeriram que a opção “*show full item record*” fosse destacada. Isto porque se trata de uma opção relevante, sendo do interesse do utilizador ter acesso a toda a informação associada a determinado conteúdo. Para o **participante E** que não estava familiarizado com este tipo de funcionalidade e caso não fosse o objetivo da tarefa - “*o comando teria passado despercebido*”. O **participante F** propôs a inversão entre o quadro que dispõe os ficheiros e o registo de metadados para ser mais fácil visualizar os ficheiros no caso da informação detalhada ser demasiado extensa.

Como resultado dos comentários feitos pelos participantes em relação ao destaque que tem de ser dado à opção de visualizar o registo completo de metadados, esta funcionalidade deverá ser incluída nas recomendações de melhoria deste protótipo.

Tarefa 6 - Explorar dados de um dos ficheiros, visualizar metadados associados e terminar com o download dos dados

A tarefa 6 surge como consequência da tarefa 5. Aqui os participantes teriam de seleccionar o botão “*Explore Data*” na linha da tabela correspondente ao ficheiro cujos dados pretendessem explorar. Em consequência o sistema mostra uma secção que contém diferentes tabelas dentro do ficheiro. O

segundo passo foi a visualização dos metadados associados a cada tabela, que por definição não aparece, tendo o utilizador de carregar no botão “Click to show metadata for current sheet” disponibilizado imediatamente acima da tabela. A figura 20 representa a página com o registo de metadados fechado e após o momento da abertura. A conclusão da tarefa ocorria com o *download* dos dados, “Get Current Data”.

Data Explorer

The screenshot shows the Data Explorer interface. At the top, there are three buttons: "Get Current Data", "Get Sheet", and "Get Workbook". Below these is a link: "Click to show metadata for current sheet". The main area displays a table titled "Sheet 0" with the following data:

grav.gpstime	grav.latitude	grav.longitude	grav.altitude	grav.gravity.specific
488688.0	38.7603	-27.084098	110.767	-53.750371
488689.0	38.760296	-27.084106	110.813	-67.168032
488690.0	38.760295	-27.084109	110.846	-80.584969
488691.0	38.760293	-27.084108	110.871	-93.994527
488692.0	38.760292	-27.084106	110.892	-107.391006
488693.0	38.760291	-27.084105	110.911	-120.76501

Below the table, there are three buttons: "Get Current Data", "Get Sheet", and "Get Workbook". A mouse cursor is pointing at the "Get Sheet" button. Below these buttons is a link: "Click to show metadata for current sheet". The metadata is displayed in a table below:

dc.creator	Bastos, Luísa; Deurloo, Richard
dc.title	Aereal Gravimetry Run (Processed Data)
dc.rights	open access
dc.type	Numerical Data
dc.date.issued	1992.0
dc.description	Final Results, combining the determined gravity values from the INS system with the GPS coordinates of the airplane in each moment

Figura 20: Explorar e descarregar dados (Repositório de dados)

Os **participantes A, B e G** não identificaram o campo “Click to show metadata for current sheet” como relevante, afirmando que o viram mas pensaram tratar-se do título da tabela. O **participante A** cometeu dois erros devido à exploração de outros botões e afirmou - “*não sei.*” - consequentemente pediu que lhe fosse confirmado o botão a carregar - “*É isto não é?*”. Já o **participante B** reiniciou a tarefa por duas vezes. O **participante C** cometeu um erro antes de entrar no botão “Explore Data” clicou na opção “View/Open”. No caso do **participante G** os erros foram contabilizados em função da insistência na manipulação do cabeçalho da tabela.

Tabela 14: Resultados tarefa 6 (Repositório de dados)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	1	2	-	72 s
B	2	-	-	-	61 s
C	-	-	1	-	43 s
D	-	-	-	-	24 s
E	-	-	-	-	16 s
F	-	-	-	-	27 s
G	-	-	4	-	53 s

média- 42.28 s

Os **participante A e B** indicaram que talvez fosse melhor reforçar a opção de ver o registo dos metadados como um *link* que possa ser explorado, enquanto o **participante G** sugeriu a adoção de uma terminologia mais concreta, como por exemplo “*Description*” ou “*Metadata*”.

Em virtude das ocorrências registadas durante a realização desta tarefa deve-se proceder a alterações na disposição no botão “*Click to show metadata for current sheet*” que aparentemente passou despercebida para alguns dos participantes, que o interpretaram como o título da tabela.

Tarefa 7 - Filtrar e descarregar uma versão parcial dos dados

A tarefa 7 propôs aos participantes experimentar uma das funcionalidades peculiares desta extensão do *Dspace*, a filtragem dos dados. Desta forma e seguindo o fluxo da tarefa anterior, o participante pode visualizar os dados numa tabela dinâmica, que para além de proporcionar a ordenação das linhas do conjunto de dados por qualquer um dos cabeçalhos, proporciona também a especificação de um conjunto de filtros a aplicar aos dados armazenados no sistema. Cada filtro é composto pelo nome da coluna a filtrar, pelo operador a aplicar - >; <; =; inclusão (*strings*), exclusão (*strings*) – e por outros operadores relevantes. O utilizador poderá seleccionar um operador lógico a utilizar na composição do filtro a partir das suas parcelas. Estão disponíveis os operadores AND e OR - no caso de o utilizador seleccionar o operador AND, apenas as linhas que satisfaçam todas parcelas do filtro serão mostradas. No caso do utilizador seleccionar o operador OR, as linhas que satisfaçam qualquer uma das condições serão incluídas nos resultados da pesquisa.

A filtragem de dados é desencadeada assim que o utilizador clique no ícone em forma de lupa, no canto inferior esquerdo da tabela. A simulação desta funcionalidade pedia aos participantes, por exemplo, que filtrassem os valores da “altitude” para um intervalo igual ou maior (*greater or equal*) a “100.846” e inferior (*less*) a “110.911”. Uma vez filtrados os dados os participantes deveriam descarrega-los a partir do botão “Get Current Data”.

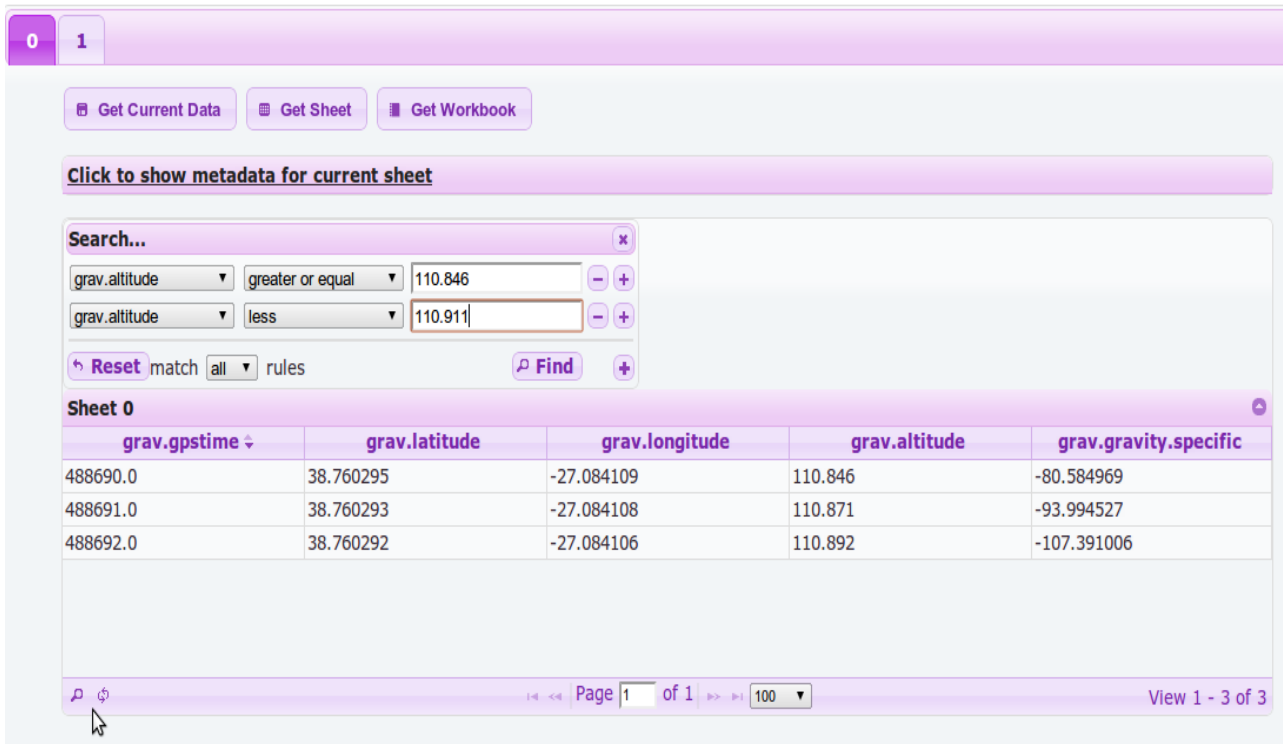


Figura 21: Filtragem de dados (Repositório de dados)

O **participante A** não terminou a tarefa porque não foi capaz de encontrar a opção de filtro, o que o levou a desistir após de cometer dez erros ao explorar várias alternativas em redor da tabela. Os dois erros dos **participantes B, D e F** devem-se à exploração dos diversos cabeçalhos da tabela e o terceiro erro do **participante B** ocorreu na manipulação da caixa de operadores de filtragem. Durante a tarefa este participante foi revelando indícios de aborrecimento por não estar a conseguir filtrar os dados.

O **participante C** procurou extrair os dados diretamente da tabela através do contorno do intervalo de dados pedido. A sua insistência nesta ação elevou a contagem de erros para cinco registos. A tarefa só foi concluída após este participante solicitar ajuda e ser-lhe indicado a presença de um ícone que permitiria proceder à filtragem. O **participante E** durante a manipulação dos operadores comentou que esta atividade - “tem que se lhe diga!” - sem que o tom indicasse uma manifestação de desconforto. O **participante F** disse que - “não tinha ideia” de como proceder à filtragem pedindo

ajuda por três vezes. O erro do **participante G** ocorreu enquanto editava operadores, carregando no botão (-), o que apagou um registo já feito. Este participante afirmou que os símbolos (+) e (-) induziam em erro porque pensou que o propósito destes seria o de aumentar ou diminuir os valores inseridos.

Tabela 15: Resultados tarefa 7 (Repositório de dados)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	-	10	X	-
B	-	-	3	-	116 s
C	-	1	5	-	144 s
D	-	-	2	-	77 s
E	-	-	-	-	48 s
F	-	3	2	-	94 s
G	-	-	1	-	123 s

média- 86 s

Os **participantes A, B, C, F e G** apelaram à substituição do ícone “lupa”, que sugere pesquisa por um elemento que seja mais fácil de associar, como um “funil”. O ícone deve estar também em maior destaque e a informação “*find records*” também não é a mais apropriada. O **participante C** acredita que não conseguiria fazer a filtragem se não lhe tivesse sido solicitada como tarefa. Outra das sugestões feitas por este participante foi a capacidade de o utilizador conseguir manipular diretamente na tabela. O **participante B** é da opinião que a caixa de operadores pode ser melhorada para se tornar mais convenientemente interpretável e que a disposição da tabela não é amigável, aconselhando o ajustamento da tabela à dimensão global do interface para uma leitura mais agradável.

O **participante G** em virtude da confusão com os botões (+) e (-) recomendou que o botão (+) imediatamente acima da tabela fosse removido por ser redundante.

Os resultados desta tarefa mostram claramente que há aspetos críticos a mudar no interface de exploração de dados, para além do que foi indentificado na tarefa anterior. Em primeiro lugar os participantes não se aperceberam que tinham a possibilidade de filtrar os dados se esta funcionalidade não lhes tivesse sido sugerido, por isso mesmo é recomendada a inclusão de notificação para esse efeito. Em segundo lugar a funcionalidade de começar o processo de filtragem não é eficaz e intuitiva. Os participantes não relacionaram a “lupa” e a designação “*find records*” à oportunidade de filtrar os dados. Assim, este botão deve ser corrigido. Por último devem ser eliminados da caixa de operadores elementos redundantes como é o caso do botão (+) junto ao botão “Find”.

Tarefa 8 - Subscrever uma coleção

A última tarefa a executar no Repositório de dados consistia na subscrição de uma das coleções. O que significa que um utilizador no caso de ter o interesse em acompanhar as atualizações de uma determinada coleção pode receber uma notificação no correio eletrónico assim que sejam acrescentados novos itens. A subscrição pode ser feita quando se acede à página inicial de uma coleção. Contudo a opção “*Receive email updates*” está prevista na página inicial do sistema.

Aereal Gravimetry Runs

Collection home page



Figura 22: Subscrição de coleção (Repositório de dados)

Os erros cometidos pelo **participante C** foram os de carregar em botões associados à navegação que não o encaminharam para a subscrição. O **participante F** regressou à página inicial depois de ter acedido à página de comunidades e não ter selecionado a coleção pedida. O **participante G** selecionou a opção “*Receive email updates*” que no entanto não disponibiliza uma lista de coleções a subscrever, aconselhando o utilizador a visitar a página da coleção, por esse motivo este utilizador teve de regressar à página inicial.

Tabela 16: Resultados tarefa 8 (Repositório de dados)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	-	-	-	48 s
B	-	-	-	-	33 s
C	-	-	2	-	40 s
D	-	-	-	-	26 s
E	-	-	-	-	09 s
F	1	-	-	-	26 s
G	1	-	-	-	38 s

média- 31.42 s

Os resultados apontam que a subscrição da coleção pode ser feita sem problemas de maior. No entanto o que a experiência do **participante G** mostra é que se o utilizador recorrer à opção “*Receive e-mail updates*”, opção lógica para fazer cumprir este objetivo, é-lhe sugerido um caminho alternativo. Desta forma é pertinente editar esta opção para que exiba as coleções passíveis de serem subscritas.

4.1.2 Impressões gerais Repositório de dados da Universidade do Porto

Para o **participante A** a possibilidade de extrair conteúdo é um ponto a favor desta ferramenta, para além da partilha de dados com elementos do grupo de investigação que estão à distância. A função de pesquisa avançada precisa, na opinião do mesmo, de ser afinada e ter um maior destaque para ser acedida de modo mais intuitivo. Isto porque a primeira coisa que faz quando usa este tipo de plataforma e não sabe o que pode encontrar é recorrer a essa funcionalidade para colmatar uma necessidade de informação. Não ficou agradado com o ícone da lupa porque lhe sugeria pesquisa, para além de não estar muito visível. A linguagem foi entendida como acessível e acredita que o sistema na globalidade permite uma adaptação rápida, e que uma vez superadas as dificuldades não teria problemas em saber como explorar as funcionalidades.

O **participante B** apontou a boa organização da informação, a rapidez e a simplicidade da plataforma, no entanto gostava de ver o interface mais limpo. Como aspeto negativo salientou que os diferentes botões devem estar identificados de acordo com a sua relevância, porque ao navegar pelas coleções foi percebendo que todos os botões que lhe foram surgindo estavam sinalizados da mesma maneira independentemente da função que desempenham, o que o obrigou a ler um por um. Para exemplificar lembrou o caso dos botões “*show full item record*” e “*view statistics*” que na opinião dele

têm diferentes graus de importância e surgem aos olhos do utilizador sem qualquer diferenciação, podendo-se perder algum conteúdo de interesse. Para além deste aspeto reforçou a ideia da dificuldade em perceber que existe a possibilidade de filtrar os dados.

A linguagem e a aprendizagem de processo foram classificadas por este utilizador como adequadas. Apesar de reconhecer o sentido prático de ter os dados disponíveis *on-line*, e o que isso representa para a partilha entre investigadores, este participante afirmou que a opção de filtrar os dados no interface não era por si só relevante para o trabalho que desenvolve. Justificou esta posição com o facto de sentir maior conforto em descarregar os dados e proceder à filtragem através de uma folha de cálculo, porque para ele é um procedimento de rotina.

Segundo o **participante C** este tipo de plataforma é inexistente no seu domínio de investigação, e já se sentiu confrontado com a necessidade de trabalhar com algo assim para conferir maior dinâmica aos projetos em que se encontra envolvido. Nesse sentido relevou a utilidade da ferramenta, utilizando o termo - “*fantástica*”, para a descrever. Os pontos fortes que identifica no Repositório de dados são a possibilidade de submissão e filtragem de um subconjunto de dados, que no seu entendimento potencia a partilha, dando como exemplo a reprodução de experiências científicas. Como sentiu dificuldades em filtrar é da opinião que este é um dos aspetos a rever no desenvolvimento da plataforma. Apesar disso considera a linguagem acessível e que a aprendizagem é imediata.

O **participante D** indicou que esta plataforma é bastante completa e que a representação das funcionalidades está no geral bem conseguida. Na opinião deste a capacidade de explorar os dados surge como um aspeto inovador. Não reforçou a ideia de pontos negativos mas gostava de ver a pesquisa avançada num posicionamento mais próximo à atividade do utilizador, ao invés do posicionamento discreto no canto superior esquerdo. A linguagem foi tida como simples e acredita que não teria dificuldades em superar os poucos problemas que sentiu numa segunda interação.

Já o **participante E** mostrou-se impressionado com o desempenho do Repositório de dados - “*acho que é enorme o tipo de pesquisa que permite em tempo real*”, e que vai facilitar imenso a partilha de dados entre os investigadores da área porque disponibiliza todos os campos relevantes a essa partilha, complementada pela possibilidade de filtrar os dados. Comentou ainda que a possibilidade de ter os dados armazenados num serviço *on-line* lhe conferia maior segurança do que nas ferramentas nas quais estes se encontram (*pen-drives*, registos anotados em papel, entre outros), em que a possibilidade destes se perderem e não poderem ser recuperados é elevada. Para este participante uma das limitações era não saber logo à partida quais os conteúdos que estão disponíveis, mas que fazer essa exploração é simples. Para além disso recomendou a adoção da língua portuguesa à ferramenta,

apesar de se sentir muito à vontade com o inglês, e da linguagem ser acessível. Segundo este participante a plataforma requer um período de habituação, sobretudo para dominar as funcionalidades paralelas.

O **participante F** revelou-se bastante satisfeito com a interação realizada. Realçou como um dos pontos fortes da plataforma a possibilidade de consultar os dados sem a necessidade de os ter de descarregar, o que na opinião do mesmo poupa trabalho aos investigadores. Tendo em conta o seu grau de especialização no uso do *Dspace* não teve dúvidas em relação à linguagem utilizada e na execução dos desafios propostos, exceto na funcionalidade estendida à filtragem de dados, que segundo o mesmo deve ser acompanhada de informação adicional que ajude o utilizador a perceber que existe essa possibilidade, para além do destaque que tem de ser dado ao ícone de filtro. Como sugestão propôs que na página de coleção fosse implementada uma lista de conteúdos, com uma indicação do número de ficheiros associados, para que um utilizador que não conheça a coleção saiba o que possa encontrar. Este participante pesquisa os conteúdos por tipo, e nesse sentido indicou que a navegação pudesse ser feita por tipologia de ficheiros. Deixou ainda a ideia de que no caso de os dados estarem com o acesso protegido, essa mensagem fosse representada através de um elemento iconográfico (por exemplo: chave), que pudesse ser pesquisável e transmitisse ao utilizador informação suplementar, como o contato dos investigadores proprietários dos dados.

Por fim o **participante G** afirmou que existem botões que não são visíveis e prejudicam a usabilidade e outros que são redundantes e produzem ruído. Para este participante a linguagem não é exata em alguns aspetos, contudo no seu todo é sintética e imediata. Explicou ainda que “metadados” pode ser uma terminologia que alguns elementos da U. Porto podem não compreender, havendo a necessidade de dar formação a alguns dos utilizadores antes de os colocar a utilizar o Repositório de dados. Salientou ainda que um serviço como este é fundamental para dinamizar a investigação porque a descrição associada às tabelas permite a quem queira reutilizar os dados obter informação contextual e - *“a oportunidade de absorver apenas o intervalo de dados que nos interessa é também muito interessante, ainda mais se o ficheiro for muito grande”*.

No final do teste foi perguntado aos participantes como avaliavam o desempenho da plataforma e o seu nível de satisfação pessoal após a experimentarem. Para responder a esta questão foi lhes pedido que atribuíssem um valor de zero a 10, numa escala de *Linkert* (zero para cotação negativa e 10 para positiva). O valor médio atribuído foi de 7,8 para ambas as avaliações. Este valor transmite uma resposta positiva por parte dos participantes à experiência realizada mas deixa em aberto o espaço para melhorias, tal como referiu um dos participantes - *“não fosse aquela questão...e o valor atribuído seria 10”*.

4.1.3 Recomendações Repositório de dados da Universidade do Porto

Tendo sido concluído o teste de utilização do Repositório de dados da Universidade do Porto, e identificados os pontos fortes e fracos do mesmo, pode ser consultada na tabela 17 o conjunto de recomendações para melhoria desta ferramenta.

Tabela 17: Recomendações de melhoria Repositório de dados da Universidade do Porto

1	Apresentar na página inicial do Repositório uma lista genérica de conteúdos que possa ser explorada.
2	Destacar a pesquisa direta nos conjuntos de dados numa posição central da página inicial do Repositório.
3	Permitir a pesquisa direta nas coleções, sem ter de percorrer comunidades.
4	Enquadrar a pesquisa avançada com a pesquisa simples, ao invés de apenas estar disponível no canto superior esquerdo do interface.
5	Sugerir termos indexados quando o utilizador começar a introduzir uma interrogação na caixa de pesquisa.
6	Mostrar conteúdos na página de coleção, com a referência ao número de documentos associados a cada conteúdo.
7	Diferenciar botões de acordo com a relevância da atividade a desenvolver, por exemplo: botão <i>“view full item record”</i> visualmente mais representativo que o botão <i>“view statistics”</i> .
8	Inverter a ordem entre a apresentação dos ficheiros e o registo completo de metadados na página relativa aos documentos associados a um item, para os ficheiros se manterem visíveis quando a descrição for muito extensa.
9	Destacar a opção de visualizar metadados na página de exploração de dados, procurando uma alternativa mais perceptível do que <i>“Click to show metadata for current sheet”</i> - ex: <i>“Dataset Description”</i> .
10	Substituir e aumentar ícone “lupa” por “funil” para dar início à filtragem de dados, e corrigir mensagem <i>“find records”</i> por <i>“filter data”</i> .
11	Introduzir informação adicional que alerte o utilizador para a possibilidade de poder filtrar os dados.
12	Eliminar o botão (+) redundante junto à opção <i>“Find”</i> da caixa de operadores.
13	Mostrar coleções que podem ser subscritas na opção <i>“Receive email updates”</i> .
14	Adaptar (aumentar) o tamanho da tabela da página de exploração de dados ao tamanho da página.
15	Disponibilizar uma versão em língua portuguesa.
16	Implementar a navegação por tipologia dos ficheiros.
17	Introduzir elemento iconográfico que simbolize o carácter protegido dos conteúdos (por exemplo: chave) que possa ser pesquisável e forneça informações como o contato dos proprietários dos dados.

4.2 Teste de utilização das plataformas UPBox e DataNotes

No caso do processo de depósito de dados está previsto um conjunto de passos a desempenhar quer pelo investigador quer pelo curador de dados. Na fase inicial, onde se desencadeia o processo de submissão, será da responsabilidade do investigador preparar e descrever genericamente o conjunto de dados a submeter. Isto para que na fase complementar, a desempenhar pelo curador, existam as condições necessárias para uma descrição específica dos dados necessária para a futura recuperação e interpretação dos mesmos.

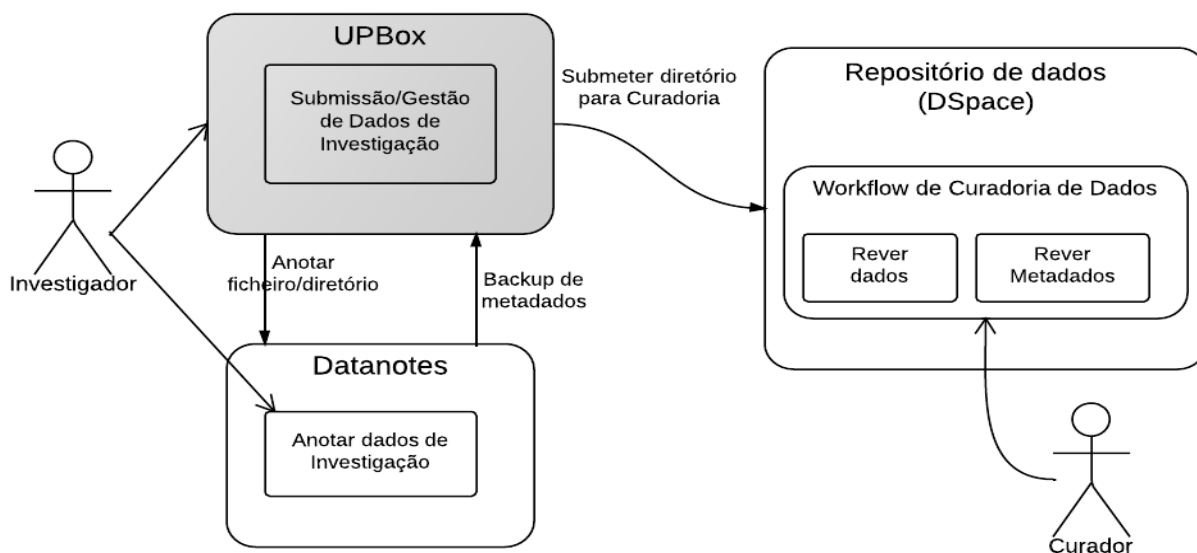


Figura 23: Workflow do processo de submissão da dados no Repositório de dados da Universidade do Porto (Barbosa, 2013)

A solução avançada consiste na integração de duas plataformas que propõem uma nova abordagem à curadoria de dados, tendo como objetivo auxiliar a gestão e a anotação de dados de investigação desde o início do ciclo de vida dos mesmos. Com isto é de esperar que no momento da submissão desses dados no Repositório de dados da Universidade do Porto não seja necessário proceder a uma organização e anotação “intensiva” por parte do curador, agilizando todo o processo.

As plataformas propostas para apoiar o processo de depósito e anotação de dados são a UPBox e o DataNotes, desenvolvidas por Pedro Barbosa (2013) e Mariana Gouveia (2013) respetivamente, no contexto de dissertação académica.

Com a plataforma UPBox pretende-se estabelecer um serviço de armazenamento de dados de investigação na nuvem, dados esses que oportunamente possam ser submetidos para curadoria e consequentemente disponibilizados no Repositório de dados da Universidade do Porto. A utilização da

UPBox é semelhante à popular solução na nuvem, *Dropbox*, comumente usada pelos investigadores, e visa funcionar como uma alternativa às já referidas práticas *ad-hoc* de gestão de dados entre os investigadores.

Esta plataforma apresenta diversas funcionalidades que permitem a gestão remota de ficheiros expressa nos casos de uso da figura 24.

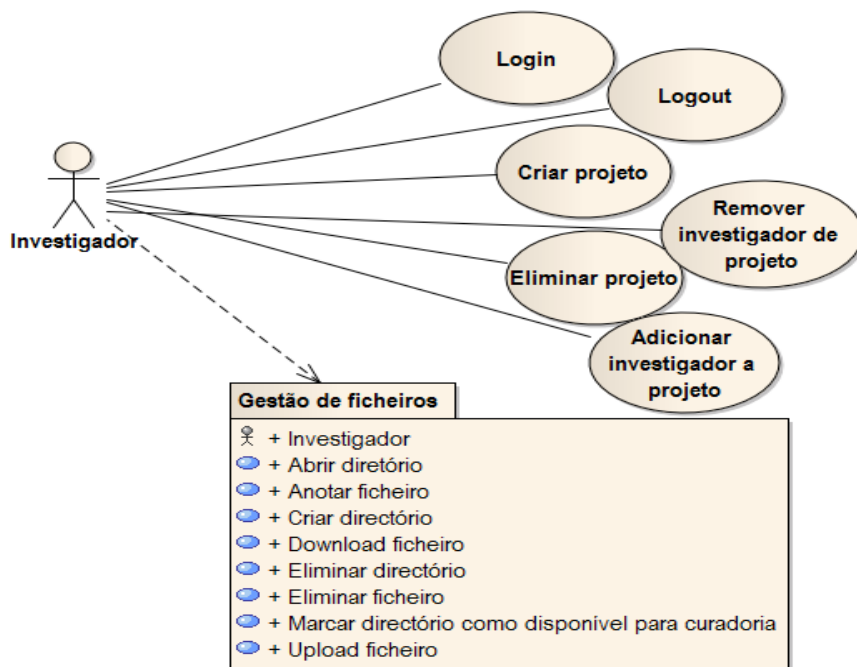


Figura 24: Casos de uso UPBox (Barbosa, 2013)

Assim durante a interação com a UPBox o investigador, como utilizador, deve efetuar a sua autenticação com as credenciais da Universidade do Porto, para não ter de criar um registo específico. Após o *login* o investigador pode criar um novo projeto, associar directórios e ficheiros ao mesmo, fazer o *download* ou eliminar um ficheiro previamente carregado, assim como, associar ou remover outro investigador do projeto. Note-se que a anotação do ficheiro, propriamente dita, é feita na aplicação DataNotes, mas o processo é desencadeado na UPBox através de um botão que serve para fazer a transição entre as plataformas. Na UPBox será possível a marcação do directório como disponível para curadoria, mas à data da realização desta dissertação essa funcionalidade não se encontra disponível, não sendo considerada nas tarefas propostas para o teste de utilização.

No caso do DataNotes pretende-se estabelecer uma plataforma colaborativa, com base numa *wiki* semântica, que permite aos investigadores anotar os dados associados a um determinado projeto. O processo de anotação é suportado por descritores previamente introduzidos, que tal como foi

explicado no capítulo 3, tanto podem pertencer a esquemas de metadados formais (estabelecidos, documentados), multidisciplinares, ou criados para um domínio concreto. A descrição dos dados permitirá que estes se tornem interpretáveis para outros investigadores, tendo em conta os princípios da criação de projeto e associação de colaboradores ao mesmo.

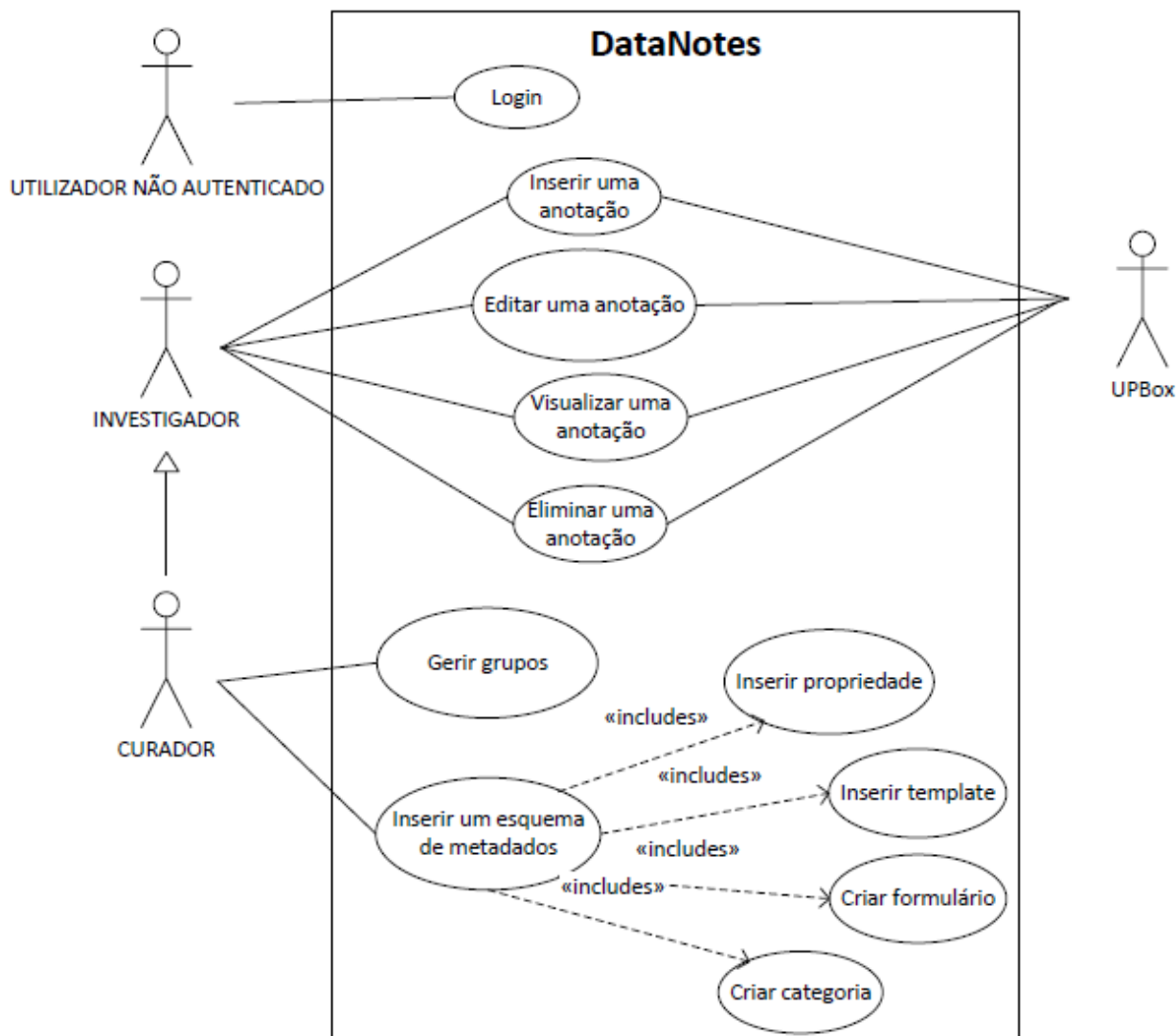


Figura 25: Casos de uso DataNotes (Gouveia, 2013)

Do ponto de vista do utilizador o DataNotes apresenta os seguintes casos de uso: após realizar o *login* na *wiki*, o utilizador pode seleccionar qual dos projetos pretende visualizar na estrutura de directórios e ficheiros. Assim que essa selecção é feita surge uma estrutura onde o utilizador pode indicar o que pretende anotar. No caso dessa anotação já estar feita em determinado ficheiro ou pasta os metadados podem ser verificados, de outra forma surge uma nova página para iniciar a anotação. A página de anotação apresenta por definição alguns campos do esquema de metadados Dublin Core para orientação do utilizador, contudo o utilizador pode escolher entre outros descritores, usar os mais pertinentes e eliminar aqueles que não precisa.

De seguida são apresentadas as tarefas a executar e os resultados dos testes de utilização das plataformas UPBox e DataNotes. Os resultados são acompanhados por sugestões de melhoria, caso estas se justifiquem.

4.2.1 Resultados do teste de utilização das plataformas UPBox e DataNotes

Tarefa 1 – Criação de um projeto

Para a iniciar o teste da plataforma UPBox foi pedido aos participantes que criassem um novo projeto. A tarefa deveria ser completada através do clique no botão “*New Project*” e consequente preenchimento dos campos “*Project name*” e “*Project description*”, terminando com “*Create Project*”. O nome e descrição do projeto ficaram ao critério dos participantes.

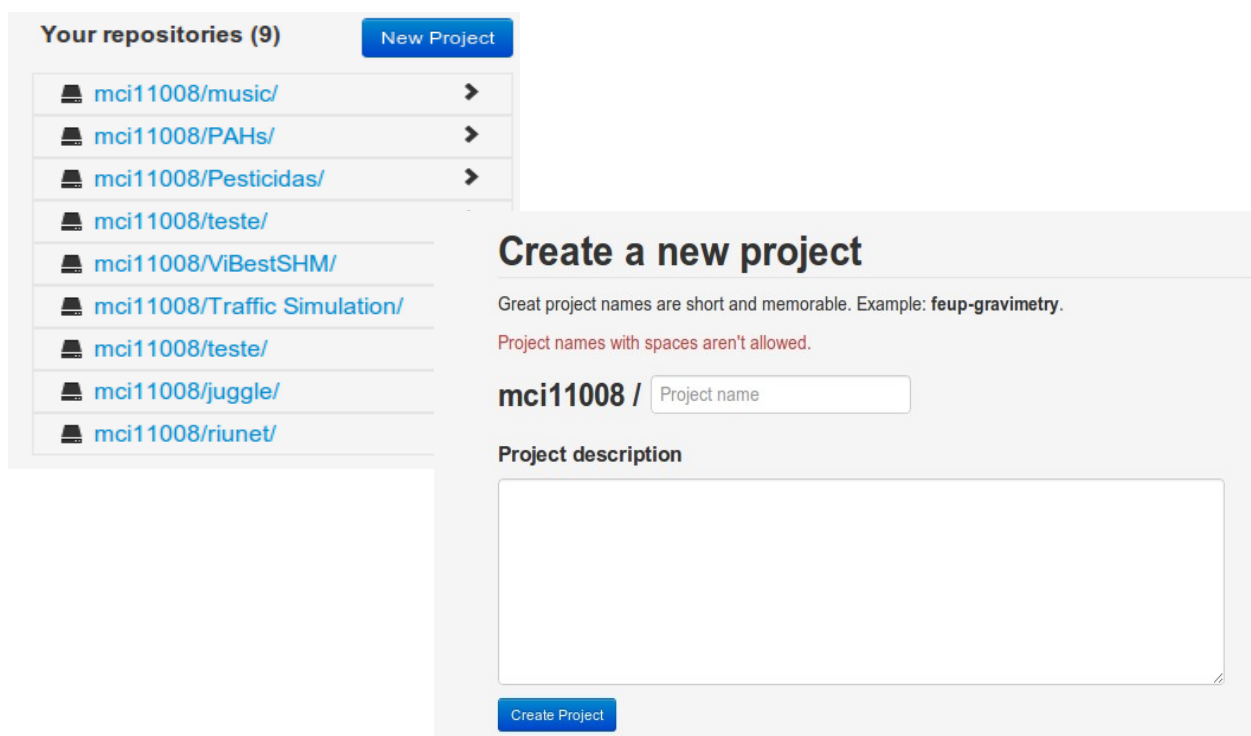


Figura 26: Processo de criação de um projeto (UPBox)

A criação de um projeto revelou-se relativamente simples de executar como comprovam os resultados obtidos.

Tabela 18: Resultados tarefa 1 (UPBox)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	-	-	-	38 s
B	-	-	-	-	19 s
C	-	-	-	-	24 s
D	-	-	-	-	50 s
E	-	-	-	-	13 s
F	-	-	-	-	45 s
G	-	-	-	-	41 s

média- 32.85 s

No término da tarefa o **participante D** sugeriu que a disposição dos projetos que vão sendo criados fosse ordenada cronologicamente, do mais recente para o mais antigo, argumentando que seria a forma lógica de encontrar os projetos em futuras utilizações. O **participante F** recomendou a aplicação de um sistema de controlo de atualização de projetos.

No seguimento da realização desta tarefa com os vários participantes, e à medida que os projetos se foram acumulando, tornou-se complicado identificar qual o projeto criado. Esta circunstância numa fase adiantada de utilização pode impedir a visualização integral da lista de projetos. Assim, será útil estabelecer a possibilidade de eliminação de projetos.

Tarefa 2 - Partilhar ficheiro com um colaborador

Após a criação do projeto foi pedido aos participantes que o partilhassem com um colaborador. Esta funcionalidade tem como fim desenvolver o trabalho colaborativo, permitindo o acesso e a edição de conteúdos a outros elementos. Para executar esta tarefa os participantes tinham de inserir as próprias credenciais na caixa existente para esse efeito, uma vez que o registo foi feito com as credenciais do analista. Desta forma, foi possível verificar em tempo real se o botão “*Add a collaborator*” estaria a funcionar corretamente, através da abertura da caixa de correio do participante para verificar a receção de uma notificação de adição ao projeto.

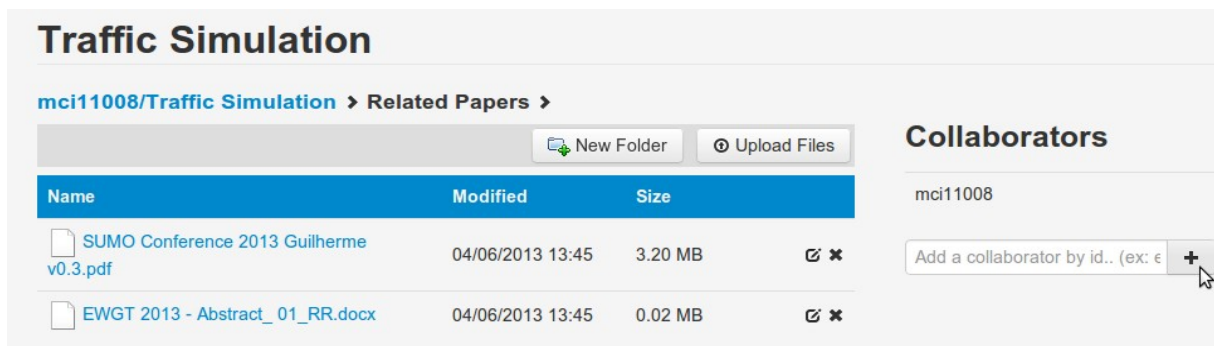


Figura 27: Caixa de adição de colaborador (UPBox)

O **participante A** revelou alguma dificuldade em executar esta tarefa, nomeadamente na identificação do campo para adicionar o colaborador. Durante a realização da mesma levantou as seguintes questões - “Onde está o partilhar aqui?”; “Clico aqui?” - sendo feita a contabilização como dois pedidos de ajuda. O mesmo participante afirmou - “Não sei como fazer” - e comentou que esta funcionalidade devia ser afinada. Por estes motivos a duração da tarefa foi de 54 segundos.

Os **participantes B, C, D, E, F e G** apesar de terem percecionado com facilidade como adicionar um colaborador ao projeto, sugeriram que o sistema transmitisse *feedback* ao utilizador com a confirmação do envio do pedido de colaboração. Nas palavras do **participante D** - “devia sair uma notícia” - a confirmar a adição do colaborador. De acordo com os **participantes E e F** não se percebe que tipo de informação se deve inserir, ficando na dúvida se se trata de um *e-mail* ou o nome de qualquer outro utilizador registado. No entendimento do **participante G** esta é uma funcionalidade essencial para a partilha de dados e como está desenvolvida pode levar ao bloqueio do utilizador.

Tabela 19: Resultados tarefa 2 (UPBox)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	2	-	-	54 s
B	-	-	-	-	22 s
C	-	-	-	-	35 s
D	-	-	-	-	45 s
E	-	-	-	-	17 s
F	-	-	-	-	24 s
G	-	-	1	-	33 s

média- 32.85 s

Um dos problemas da caixa que permite adicionar o colaborador é não permitir ler na totalidade o exemplo que indica que o utilizador apenas tem de fazer o registo parcial da credencial. O mesmo será dizer que um utilizador cuja credencial institucional seja “mci11008@fe.up.pt” apenas deve inserir “mci11008”. Não existe qualquer mensagem de erro ou sugestão de correção na eventualidade do utilizador escrever a credencial completa.

O **participante A** afirmou que esta funcionalidade lhe seria mais útil se fosse permitida a adição de colaboradores externos à Universidade do Porto (condicionante no momento da realização do teste de utilização), porque o trabalho que desenvolve é feito em colaboração com elementos de outras instituições. O **participante F** afirmou que seria interessante se o sistema antecipasse o que o utilizador está a digitar, apresentando uma lista de colaboradores que correspondessem ao pedido do utilizador, a partir da terceira letra.

Tendo em conta o problema identificado esta funcionalidade não pode ser classificada como eficaz, já que não transmite a informação suficiente para ser executada e superar possíveis erros e, portanto devem ser promovidas alterações.

Tarefa 3 - Criar uma pasta e adicionar ficheiros

Na tarefa 3 foi pedido aos participantes que criassem uma pasta e em seguida fizessem o *upload* de ficheiros para a mesma. O nome atribuído à pasta ficou ao critério de cada um, o mesmo acontecendo em relação à tipologia e localização do ficheiro no computador dos participantes.

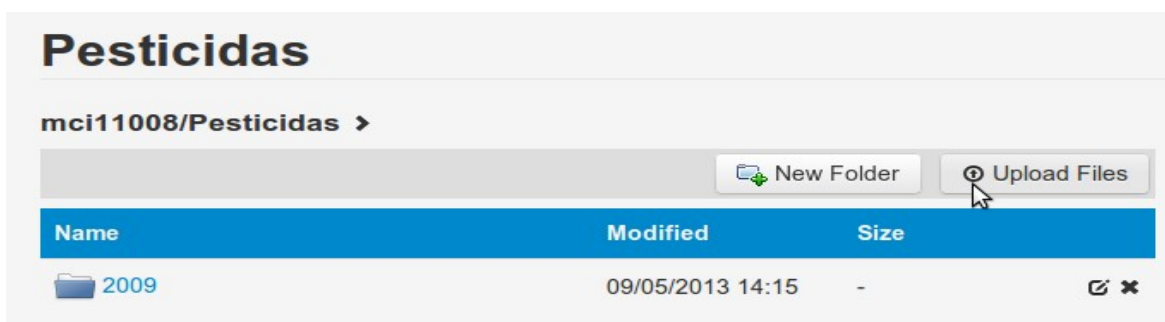


Figura 28: Criação de pastas e *upload* de ficheiros (UPBox)

O **participante A** comentou a facilidade em adicionar as pastas, dizendo que está claro, assim como o *upload* e escolha de ficheiro foram consideradas como atividades óbvias do ponto de vista do utilizador. O tempo de execução da tarefa está relacionado com a seleção dos ficheiros no disco rígido. O **participante E** destacou a simplicidade da tarefa.

Tabela 20: Resultados tarefa 3 (UPBox)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	-	-	-	42 s
B	-	-	-	-	47 s
C	-	-	-	-	37 s
D	-	-	-	-	27 s
E	-	-	-	-	18 s
F	-	-	-	-	31 s
G	-	-	-	-	46 s

média- 35.84 s

A limitação encontrada na realização desta tarefa foi a incompatibilidade com o *web browser Mozilla FireFox*. Ao utilizar este *browser* o **participante A** não conseguiu fazer o carregamento de ficheiros, passando a ser sugerido aos outros participantes que utilizassem o *Google Chrome*.

Paralelamente foi sugerido o *upload* de mais do que um ficheiro em simultâneo, para que pudesse ser testada essa funcionalidade.



Figura 29: Visualização do estado de carregamento de ficheiros (UPBox)

O desempenho do carregamento foi testado com ficheiros de maior peso, no caso aproximadamente 100 *megabyte*, tendo o carregamento sido feito com uma velocidade que agradou aos participantes. O **participante F** expressou a sua satisfação pela rapidez com que o sistema carregou vários ficheiros em simultâneo e com um peso considerável.

A sinalização do estado de carregamento de ficheiro executa bem a sua função. Em aberto fica a possibilidade de implementação da funcionalidade de carregamento de pastas.

Tarefa 4 – Efetuar o *download* de um ficheiro

Para a execução desta tarefa os participantes deveriam proceder ao *download* de um ficheiro. Para tal deviam clicar sobre a hiperligação correspondente.



Figura 30: Hiperligação para *download* de ficheiro (UPBox)

Os resultados indicam que os participantes não tiveram dificuldades em executar esta tarefa.

Tabela 21: Resultados tarefa 4 (UPBox)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	-	-	-	06 s
B	-	-	-	-	16 s
C	-	-	-	-	09 s
D	-	-	-	-	07 s
E	-	-	-	-	04 s
F	-	-	-	-	06 s
G	-	-	-	-	06 s

média- 7.7 s

De acordo com o **participante D** o *download* dos ficheiros deve ser precedido de uma opção de visualização dos conteúdos, para evitar estar a descarregar ficheiros que não sejam relevantes. Outro aspeto a melhorar na opinião deste participante é tornar mais perceptivo o momento do *download*. Nesta fase o sistema procede imediatamente ao *download* assim que se clica no *link*, sem que haja um pedido de confirmação ou do destino de armazenamento. Para o **participante F** a opção de *download* deveria estar mais bem sinalizada e não apenas funcionar através do *link*. Enquanto o **participante G** comentou não se ter apercebido de ter feito o *download*.

Por estes motivos a execução do *download* deve ser revista para transmitir maior controlo dos utilizadores sobre a ação. Outa das opções a considerar é a possibilidade de descarregar diretamente os diretórios, mediante a compressão num único ficheiro.

Tarefa 5 – Remover um ficheiro do projeto

Para executar a tarefa 5 foi pedido aos participantes que eliminassem um dos ficheiros. Para o fazer deviam carregar no botão (x) - correspondente ao ficheiro selecionado.

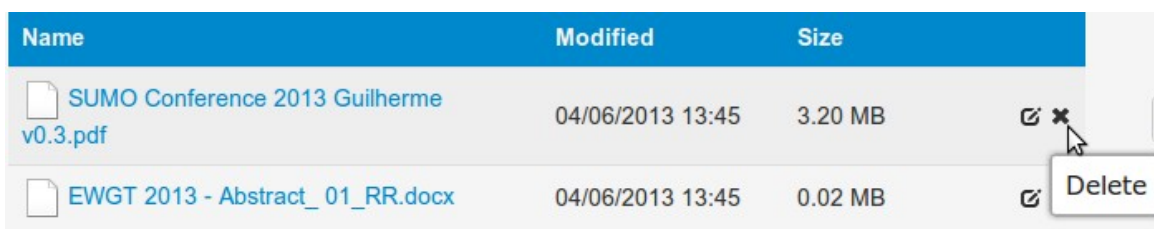


Figura 31: Eliminação de ficheiros (UPBox)

Durante a eliminação de ficheiros não foram feitos registos de problemas em executar a tarefa e o tempo médio para a execução da tarefa está dentro dos valores esperados.

Tabela 22: Resultados tarefa 5 (UPBox)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	-	-	-	09 s
B	-	-	-	-	09 s
C	-	-	-	-	12 s
D	-	-	-	-	11 s
E	-	-	-	-	08 s
F	-	-	-	-	07 s
G	-	-	-	-	18 s

média- 10.57 s

O **participante F** recomendou que fosse alargado o tamanho do ícone para facilitar a visualização a utilizadores com eventuais problemas de visão, considerando que as dimensões dos ícones são muito pequenas. Para além disso afirmou que o ícone podia ser reforçado com uma cor mais viva, semelhante ao ícone de eliminação do DataNotes (ver fig. 37), como numa fase mais adiantada do teste

lembrou. O **participante G** recomendou que o cabeçalho fosse completo com mais informação relativa às funções a desempenhar neste interface, dando o exemplo de - “*Gestão de ficheiros*”.

Tarefa 6 – Iniciar processo de anotação

A última tarefa a executar no UPBox teve como objetivo iniciar o processo de anotação. Através do ícone destacado na figura 32, o participante faz a transição do UPBox para o DataNotes.



Figura 32: Iniciar processo de anotação (UPBox)

Esta tarefa revelou-se a de maior complexidade na UPBox. O **participante A** ficou confuso com a forma pela qual se faz a transição para o DataNotes, ou seja, a forma de começar o processo de anotação. Os comentários feitos durante a tarefa foram os seguintes: “*Não estou a ver*”; “*É isto aqui, UPBox?*”; “*Daqui não consigo sair*”; “*Há algum link para lá?*”.

O **participante B** também teve algumas dificuldades em executar a tarefa. Não compreendendo imediatamente qual era o ícone que lhe permitiria dar início à anotação retrocedeu à página inicial e perguntou - “*aqui?*”. Durante a tarefa este participante e o **participante D** carregaram inadvertidamente no *link* do ficheiro (*download*). O **participante E** definiu o ícone de anotação como pouco intuitivo, tendo cometido um erro e regressado à página inicial da UPBox para procurar a solução, comentando - “*não é fácil ver como isso se faz*”. O **participante F** demonstrou uma certa hesitação em identificar o botão de iniciar a anotação, apesar de não cometer erros ou retroceder de página, procurou durante 32 segundos essa opção.

Tabela 23: Resultados tarefa 6 (UPBox)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	4	2	-	66 s
B	1	1	1	-	45 s
C	-	-	-	-	15 s
D	-	-	1	-	12 s
E	1	-	1	-	26 s
F	-	-	-	-	32 s
G	-	-	-	-	5 s

média- 32.42 s

Foi opinião unânime entre os participantes que a funcionalidade de prosseguir para a anotação deva estar sinalizada com maior nitidez ou ser acompanhada de informação complementar, nomeadamente no cabeçalho. Entre os participantes houve quem sugerisse que o ícone podia ser mais representativo da função, dando como exemplo um lápis, mas em maiores dimensões. O **participante F** recomendou o aproveitamento do muito espaço livre do interface para ser introduzida uma caixa de ajuda ao utilizador, que pudesse resolver este problema.

O tempo médio de execução desta tarefa demonstra que não se trata de uma funcionalidade eficaz se comparado com o tempo médio da tarefa anterior (10.57 segundos), em que o modo para atingir o objetivo era idêntico. Isto, somado aos comentários dos participantes e a incidência de erros e pedidos de ajuda indicam que uma alteração deve ser feita na representação do ícone de anotação.

Tarefa 7 - Selecionar ficheiro a anotar

A primeira tarefa pedida para os participantes executarem no DataNotes foi a seleção de um ficheiro entre os anteriormente submetidos na UPBox. Para isso os participantes deveriam clicar no botão “*Select*” no canto inferior direito, que dá acesso a uma estrutura de ficheiros UPBox.

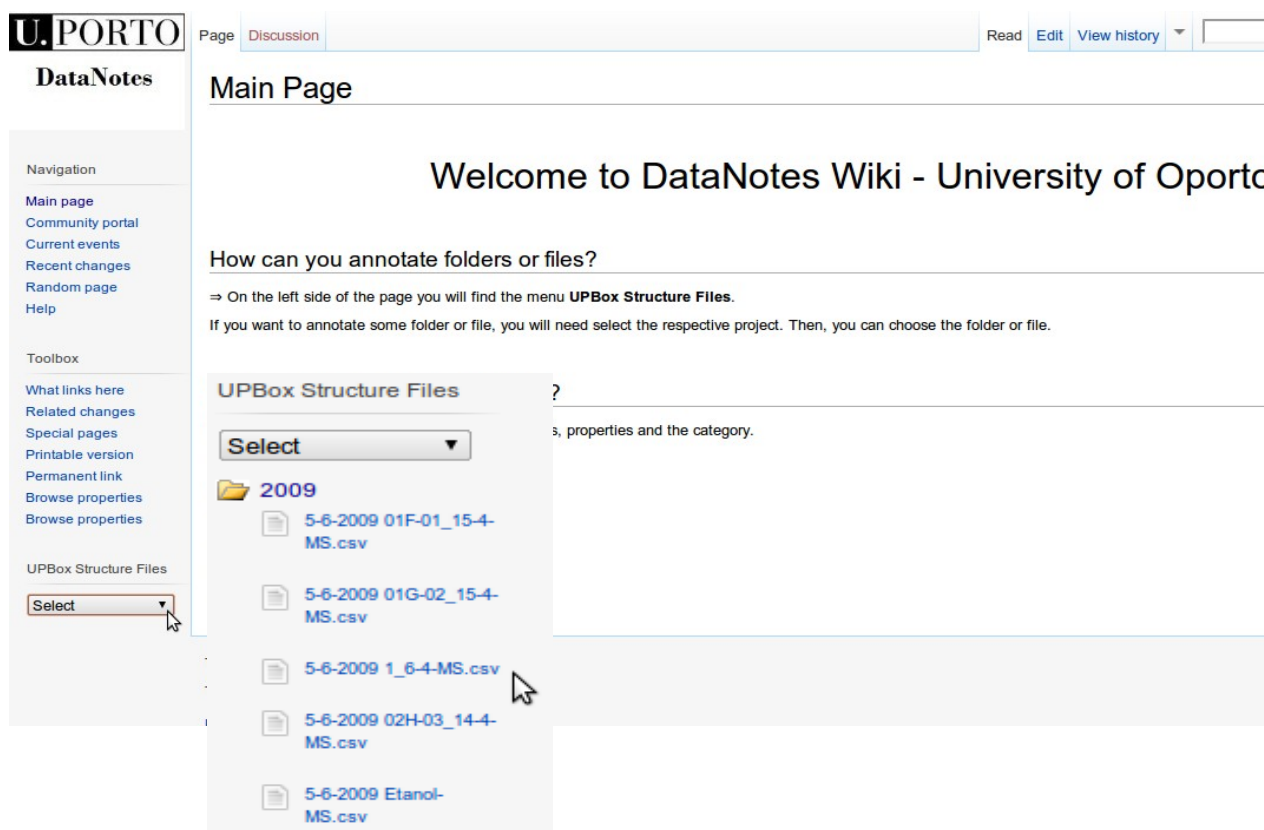


Figura 33: Seleção de ficheiro a anotar (DataNotes)

O **participante B** teve dificuldades em encontrar o botão “Select”, cometendo dois erros ao clicar em opções irrelevantes para ser bem-sucedido nesta tarefa, confirmando ainda com o analista se o botão “Select” era o pretendido antes de pôr fim à tarefa. Após terminar referiu que a opção para seleccionar o ficheiro devia estar em destaque - “*devia aparecer maior, fiquei um pouco na dúvida*”.

Tal como o **participante B**, o **participante C** não conseguiu encontrar a caixa de seleção do ficheiro com facilidade. Os erros cometidos devem-se à exploração de vários botões no interface inicial do DataNotes, e por consequência o **participante C** teve de voltar três vezes à página inicial. Durante a realização da tarefa comentou não encontrar a opção para seleccionar o ficheiro a editar.

O **participante D** hesitou um pouco e procurou confirmar junto do analista se o seu procedimento era o correto antes de escolher definitivamente o ficheiro - “*Chegar aqui, não?*”; “*é isto?*” - o que foi entendido como dois pedidos de ajuda.

Tabela 24: Resultados tarefa 7 (DataNotes)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	-	-	-	37 s
B	-	1	2	-	59 s
C	3	1	7	-	175 s
D	-	2	-	-	54 s
E*	-	-	-	-	-
F	-	-	-	-	77 s
G	-	-	-	-	47 s

média- 64.14 s

* O registo de dados relativos ao **participante E** não foi contabilizado por motivos de uma falha momentânea no sistema que implicou que o analista procedesse à resolução do problema, o que em último caso condicionou os resultados.

À luz dos resultados, tempo médio despendido (a duração prevista para a realização da tarefa era de 30 segundos) e ocorrências de erros, a opção de seleção de ficheiro não pode ser considerada como eficaz e eficiente. Para o **participante F** a caixa de seleção de ficheiros deve ser colocada numa posição cimeira, invertendo a ordem com todas as opções que surgem na coluna da esquerda, que na sua opinião produzem ruído. Já o **participante G** queixou-se da omissão do nome do projeto enquanto navegava pela estrutura de ficheiros, o que lhe causou uma sensação de confusão.

Tendo em conta o carácter central da seleção do ficheiro a anotar para o processo a decorrer no DataNotes este problema de usabilidade deve ser corrigido.

Tarefa 8 – Editar anotação

Após selecionar o ficheiro a anotar, o DataNotes conduz o utilizador para a página de criação de metadados associada ao documento. Para a realização da tarefa 8 foi pedido aos participantes que simulassem a anotação de um ficheiro escolhendo os seguintes campos: *dc.title*; *dc.subject*; *dc.creator*; *dc.date*; *dc.description*.

Tal como é sugerido na figura 34, o campo *dc.title* aparece automaticamente na abertura desta página, o mesmo acontece nos casos do *dc.description* e do *dc.subject*. Os restantes campos teriam de ser selecionados a partir do botão “Select” que dá acesso a uma lista de descritores pré-definida. O preenchimento dos campos, ou seja o grau de detalhe da anotação, ficou ao critério de cada um. Para

terminar a tarefa foi pedido aos participantes que gravassem a edição. Foi também sugerido aos participantes que memorizassem a anotação feita no campo *dc.creator* para executarem uma das tarefas posteriores (tarefa 10).

Create metadata: Mci11008/Pesticidas/2009/5-6-2009 01G-02 15-4-MS.csv

Figura 34: Página de edição de metadados (DataNotes)

Apesar da demora para executar a tarefa, os **participantes A, E, F e G** não tiveram problemas em chegar à página de edição. O tempo despendido deveu-se ao grau de detalhe conferido à descrição (opção facultativa). O **participante C** referiu - “*não encontro o link*” - enquanto procurava o botão que lhe daria acesso à seleção dos campos, solicitando ainda a ajuda do analista em dois momentos. O **participante E** reconheceu a simplicidade em seleccionar os campos.

Tabela 25: Resultados tarefa 8 (DataNotes)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	-	-	-	236 s
B	-	-	-	-	78 s
C	-	2	1	-	214 s
D	-	-	-	-	52 s
E	-	-	-	-	207 s
F	-	-	-	-	159 s
G	-	-	-	-	126 s

média- 130.42 s

O **participante E** após terminar a tarefa referiu que entendia como desvantagem o facto de não estarem definidos metadados como obrigatórios, clarificando que no seu ramo de investigação essa é uma questão essencial para criar as condições ideais de partilha. O **participante F** partilhou este ponto de vista, aconselhando a adoção de um sinal tipográfico para identificar os campos que devem ser de preenchimento obrigatório.

O **participante E** acrescentou que devia ser utilizado um esquema de cores para guiar o processo de anotação. Vermelho para indicar que o processo de edição está por iniciar, amarelo para uma anotação incompleta e verde como certificado de validação. Desta forma seria retirado o desconforto de ter de preencher tudo de uma só vez, tornando o processo gradual e mais fácil de acompanhar.

Outro dos aspetos mencionados pelo **participante F** foi a discrepância entre os campos de título, com uma barra de preenchimento curta, e de assunto, com uma barra longa, propondo uma inversão desta lógica. Para além disso, e como o campo de assunto por norma é preenchido mais do que uma vez pode ser aborrecido para o utilizador ter de ir buscar o campo à caixa de seleção sempre que o desejar repetir. Este constrangimento pode ser ultrapassado segundo o **participante F** se o sistema permitir acrescentar campos simplesmente por clicar num botão (+) colocado à frente dos campos. Referiu que o botão “*Save page*” deveria ocupar uma posição de maior destaque. Por fim aconselhou a adoção de um calendário durante o preenchimento da data, para não tornar ambígua essa anotação (por exemplo: **01**, june, 2013 / **1**, june, 2013).

Uma possibilidade para agilizar o processo de anotação, para não obrigar o utilizador a preencher vários ficheiros com as mesmas descrições, é permitir que o sistema associe automaticamente as descrições da pasta “mãe” aos ficheiros incluídos, salvaguardando, porém, que a anotação possa ser feita particularmente num determinado ficheiro caso haja essa necessidade. Por outro lado a caixa de seleção de metadados deve ser organizada de acordo com o domínio para que o número elevado de elementos a percorrer não torne a tarefa penosa.

O que também foi possível concluir pelo analista durante os testes prévios é que o sistema não grava as edições se o utilizador abandonar a página sem carregar em “*Save page*”, o que atribui um valor relevante a este botão, que como foi adiantado pelo **participante F** está discreto. Por isso será incluída nas recomendações de melhoria desta ferramenta uma mensagem de confirmação do abandono da página editada quando forem feitas edições que não tenham sido gravadas.

Tarefa 9 – Eliminar uma das anotações

A tarefa 9 consistia em eliminar uma das anotações previamente anotadas. A tarefa era concluída no momento em que o utilizador clicasse no botão (**x**) associado a um descritor à sua escolha (ver figura 34). No entanto os participantes não fizeram a eliminação do ficheiro instantaneamente porque assim que é guardada a edição feita, a exemplo da tarefa anterior, o sistema conduz o utilizador para uma página onde é apresentada uma tabela com os campos preenchidos. A opção do participante para voltar à página de edição requer carregar na opção “*Edit with form*” ou “*Edit*”, nesta última os campos a editar aparecem em linguagem XML.

O **participante A** teve alguns problemas em perceber como poderia voltar à página de edição para eliminar uma anotação. Alguns dos comentários feitos por este participante foram os seguintes: “Não sei o que quer dizer “edit with form””; “sei lá!”; “qual a diferença entre “edit” e “edit with form” - contabilizando um erro por explorar o botão “edit” sem que daí tivesse conseguido fazer a eliminação do campo. Após a tarefa seguiu-se uma explicação sobre a diferença entre os botões, o que nas palavras do participante significa que o botão “edit” - “não presta!”.

O **participante B** teve dificuldade em executar a tarefa. Não viu o botão “*edit with form*”, o que o levou à exploração de opções irrelevantes para a tarefa. Os comentários deste participante revelaram um sentimento de aparente frustração - “*não aparece nada para apagar*”. A tarefa foi concluída após o **participante B** ter feito a sugestão de a abandonar. No caso do **participante C** a tarefa só foi terminada após uma exploração intensa dos botões presentes no interface o que levou à contagem de onze erros, e o regresso à página inicial do DataNotes por três ocasiões. Este participante referiu - “*não estou a ver*” - e perguntou - “*como se faz?*”.

O **participante G** teve o mesmo tipo de problemas que os outros participantes com a particularidade de ter insistido com o botão “*browse properties*” e consultado o “*help*” que no entanto apenas apresenta uma mensagem da não existência de texto. Os comentários feitos por este participante foram - “*não estou a ver*”; “*está complicado*”; “*estou a andar em círculos*”.

Apesar dos problemas em aceder à página de edição de anotação, uma vez dentro dessa página nenhum dos participantes teve dúvidas em relação ao botão que permite eliminar um registo.

Tabela 26: Resultados tarefa 9 (DataNotes)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	2	1	-	39 s
B	-	-	2	X	140 s
C	3	2	11	-	215 s
D	-	-	-	-	17 s
E	-	-	-	-	18 s
F	-	-	-	-	23 s
G	4	3	8	-	265 s

média- 102.42 s

No final da tarefa os **participantes A, B, C, D e G** referiram que a solução para resolver o problema de voltar à edição dos ficheiros, seria acrescentar essas opções à tabela apresentada (fig. 35). Tanto o **participante B** como **G** concluíram que os elementos de interação estão muito distantes, a caixa de seleção no canto inferior esquerdo e a opção de editar no canto superior direito. Ambos recomendaram uma aproximação entre as funcionalidades chave.

Mci11008/ViBestSHM/SHM1/Grelha.pdf

DC

Title	Test Grid
Subject	Testing Software

Figura 35: Tabela de descrições (DataNotes)

O que estes resultados traduzem não é um problema na eliminação de campos mas sim no processo de edição, portanto as modificações devem minimizar a dispersão das opções ao dispor do utilizador, assim como flexibilizar a edição na tabela de dados descritos. Por fim a página de ajuda deve ser editada.

Tarefa 10 – Recuperar ficheiro pela autoria (*dc. creator*)

Para executar a tarefa 10 foi pedido aos participantes que recuperassem um dos documentos editados. Essa recuperação teria de ser feita com recurso à caixa de pesquisa no canto superior esquerdo do

interface do DataNotes, ao colocar como interregação ao sistema o nome que atribuíram ao *dc.creator* no momento da edição dos ficheiros (tarefa 8). Note-se que no momento da edição de campos foi pedido aos participantes que relembassem a autoria do ficheiro.



Figura 36: Caixa de pesquisa (DataNotes)

O **participante B** não identificou a caixa de pesquisa prevista para a execução com sucesso desta tarefa. Após voltar à página inicial este utilizador percorreu outros dois botões “*view history*” e “*recent changes*”, o que implicou a contagem de dois erros. O **participante G** cometeu um lapso que não deve ser tido como relevante.

Tabela 27: Resultados tarefa 10 (DataNotes)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	-	-	-	-	16 s
B	-	-	2	-	58 s
C	-	-	-	-	13 s
D	-	-	-	-	10 s
E	-	-	-	-	11 s
F	-	-	-	-	13 s
G	-	-	1	-	17 s

média- 19.71 s

O **participante B** recomendou que na página inicial deveria estar presente uma lista de documentos relativos ao utilizador registado, enquadrada no centro da página e devidamente estruturada no caso do número de ficheiros atingir uma dimensão considerável, que pudesse ser navegada até ser encontrado o ficheiro de interesse.

Tarefa 11 – Voltar à página de anotação e adicionar campo

A última tarefa proposta para realizar na plataforma DataNotes determinou que os participantes voltassem à página de anotação e adicionassem o campo *dc.type* aos campos anteriormente anotados. Tal como na tarefa 9 os participantes teriam de recorrer à opção “*Edit with form*” após o surgimento da tabela com as descrições efetuadas.

Edit metadata: Mci11008/Traffic Simulation/Related Papers/EWGT 2013 - Abstract 01 RR.docx

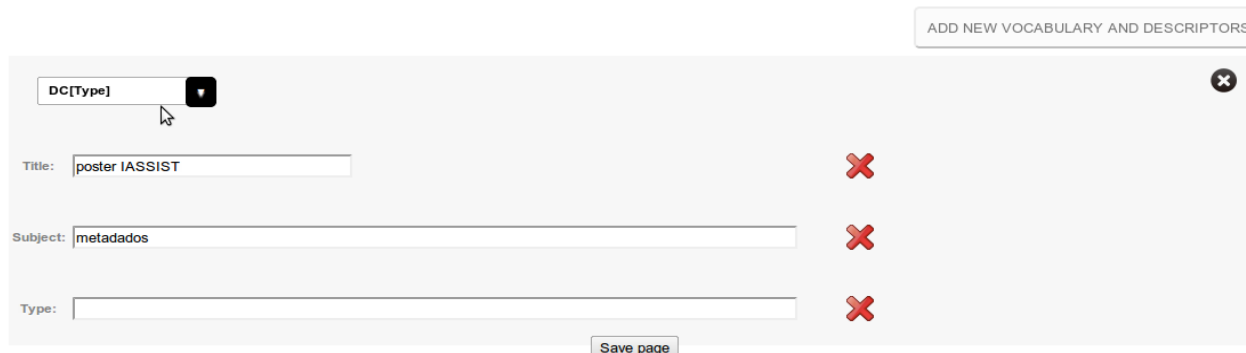


Figura 37: Seleção de metadados (DataNotes)

O **participante A** revelou dificuldades em levar a cabo esta tarefa, particularmente em voltar ao campo de edição de dados, mesmo tendo em consideração o procedimento da tarefa 9. Os comentários tecidos por este participante foram - “*não está evidente*”, “*New page?*”, “*não consigo lá chegar*”, “*não sei*”, “*será isto?*”. Decorrente dos erros cometidos e do facto de ter reiniciado a tarefa por três vezes, somados aos pedidos de ajuda, este participante decidiu-se pela desistência da tarefa. O **participante B** também revelou dificuldades no processo de adição de um novo elemento à descrição do ficheiro, navegou por oito botões de funcionalidade irrelevantes para o utilizador e fez duas tentativas de voltar ao início da tarefa.

Tabela 28: Resultados tarefa 11 (DataNotes)

Participante	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo despendido
A	3	2	4	X	-
B	2	-	8	-	176 s
C	-	-	-	-	33 s
D	-	-	-	-	18 s
E	-	-	-	-	21 s
F	-	-	-	-	27 s
G	-	-	-	-	43 s

média- 53.42 s

Com esta tarefa foi possível identificar um erro no serviço: a incapacidade aparente do sistema em mostrar os elementos anteriormente descritos. Os utilizadores ficam assim sem saber quais os campos já preenchidos, o que se agrava se o projeto estiver a ser partilhado entre vários investigadores. Na figura 38, por exemplo, são apresentadas anotações feitas previamente nos campos “title” e “subject”, e no campo “type” adicionado na execução desta tarefa, omitindo o campo “description”, que também já se encontra anotado (tarefa 8). Essa anotação é recuperada quando o utilizador voltar a seleccionar o campo “description” na caixa de seleção de metadados, como se pode comprovar na figura 37.

A screenshot of a web form interface. At the top, there is a dropdown menu with 'DC[Description]' selected and a mouse cursor pointing at it. Below the dropdown are four input fields: 'Title:' containing 'poster IASSIST', 'Subject:' containing 'metadados', 'Type:' which is empty, and 'Description:' containing the text 'este documento descreve um processo...'. At the bottom right of the form, there is a 'Save name' button.

Figura 38: Recuperação de um campo previamente descrito (DataNotes)

As sugestões de funcionalidade para a opção de voltar a editar ficheiros foram previamente identificadas na tarefa 9. Curiosamente o **participante A** abandonou esta tarefa quando na tarefa 9 não tinha tido problemas, o que reforça o problema deste processo.

4.2.2 Impressões gerais UPBox e DataNotes

O **participante A** considerou o interface do UPBox como amigável e semelhante à ferramenta que utiliza normalmente, a *Dropbox*. Para além do problema em localizar o botão de anotação, a dificuldade sentida foi em relação à funcionalidade de adicionar um colaborador, considerando que é uma função que não é assimilada intuitivamente, sugerindo para este caso a necessidade de afinação. No entanto considera que após saber como o fazer não terá dificuldades em repetir o procedimento.

No caso do DataNotes este participante considerou que apesar de um ou dois passos a experiência foi positiva, afirmando estar satisfeito mas acrescentando que há coisas difíceis de perceber, principalmente quando se trata de voltar a editar um ficheiro ou uma pasta. Afirmou nunca ter

trabalhado com nenhuma ferramenta deste género, acreditando que é um serviço útil porque a informação fica organizada, registada, garantindo a autonomia necessária para desencadear o processo de anotação de dados. Não lhe ocorreram sugestões de melhoria mas mostrou desagrado em relação ao excesso de informação, porque segundo o mesmo a maioria dos campos a descrever não lhe interessam e mais de metade dos campos ficam por preencher (referindo-se aos elementos Dublin Core que surgem na caixa de seleção de metadados). A linguagem das plataformas foi considerada como clara.

Para o **participante B** utilizar o Upbox é bastante simples e intuitivo, sendo uma ferramenta de utilização rápida e fácil. No entanto, referiu que teve dúvidas em perceber como dar início ao processo de anotação, mas não teria dificuldade em executar a tarefa novamente. Considerou que a linguagem é acessível em ambas as plataformas, sendo que a UPBox assemelha-se ao *Feup Load*, e o DataNotes lhe faz lembrar o *software* que o grupo de investigação utiliza. Por outro lado afirmou que o DataNotes peca pelo excesso de informação, referindo-se às várias opções que surgem no interface que não lhe pareceram relevantes. Por isso mesmo considerou que a pesquisa é confusa, com diferentes zonas de informação sem encadeamento natural.

Na opinião do mesmo a plataforma representa uma solução útil porque para além de manter os registos atualizados vai permitir que outros colaboradores possam ficar com um conhecimento que por norma fica apenas para a pessoa que produz os dados. Como sugestão indicou que o sistema deveria confirmar a adesão dos colaboradores, e que a pesquisa no DataNotes devia estar organizada com um padrão uniformizado e estruturado em árvore.

De acordo com o **participante C** a plataforma UPBox revelou-se bastante intuitiva e fácil de usar tendo em conta a simplicidade do interface. Tal como os resultados sugerem este participante não teve dificuldades em executar nenhuma das tarefas propostas, isto porque na opinião dele tanto a linguagem como os ícones são compreensíveis. Este considerou o UPBox similar às plataformas *DropBox* e *Redmine*⁴⁹, com as quais está habituado a trabalhar, afirmando que para efeitos de partilha o UPBox não acrescenta valor por si só. Relativamente ao DataNotes este participante afirmou tratar-se de um interface pouco intuitivo, o que lhe causou alguns problemas na compreensão dos objetivos a que esta plataforma se propõe. Observou ainda que talvez sentisse algumas dificuldades quando voltasse a utilizar a ferramenta. Para além disso salientou que a linguagem não é a mais apropriada, mais concretamente quanto à designação da caixa que dá acesso aos ficheiros “*UPBox Structure Files*”, sem contudo sugerir uma terminologia mais incisiva. Revelou interesse em voltar a utilizar o

⁴⁹ Aplicação web para uma gestão de projetos flexível. <http://www.redmine.org/>

DataNotes como ferramenta de trabalho, afirmando que a ferramenta que utiliza de momento para anotação de documentos é o *EndNote*.

Como sugestão referiu que a funcionalidade de adicionar colaboradores aos projetos podia ser mais vantajosa se apresentasse um histórico de partilha e estatísticas associadas, ou seja, o participante acharia pertinente saber o número de projetos partilhados com determinado colaborador. Por outro lado considera que seria confortável para o utilizador poder aceder diretamente do ambiente de trabalho às plataformas, como atalho, isto porque por norma trabalha com o *Google Desktop*. Outro aspeto que consideraria interessante seria a comunicação com outras plataformas, dando o exemplo do *ResearchId*⁵⁰, *EndNote*⁵¹ e redes sociais.

Segundo o **participante D** as duas ferramentas são no geral intuitivas, simples, de linguagem acessível e fáceis de usar, tendo todas as funcionalidades - “à mão de semear”. A perceção inicial deste participante foi que a única funcionalidade do UPBox seria guardar ficheiros e não lhe ocorria nada que fosse preciso modificar. Em termos de comparação desta plataforma com outras soluções, afirmou que esta é menos complicada que a *Dropbox*, que no seu entender tem elementos que confundem. Reforçou alguns dos aspetos que mencionou durante a execução das tarefas, ou seja, a necessidade da organização dos projetos em sentido decrescente, se possível com uma linha cronológica, por se tratar de um requisito vital para a gestão de documentos no grupo de investigação em causa. Sugeriu também a possibilidade do utilizador ter a hipótese de ordenar os projetos por ordem alfabética. Por fim e como o ecrã deste utilizador era de dimensões reduzidas (tamanho 3.3) alguns dos elementos do interface ficavam ocultos, por esse motivo recomendou uma adaptação a ecrãs de menor dimensão. No caso do DataNotes por exemplo não lhe permitia ver o campo de seleção de ficheiro, o que obrigou a fazer *scroll* para encontrar o botão mais relevante da página inicial.

No caso particular do DataNotes chamou a atenção para o facto de o sistema não lhe apresentar campos já editados. Considerou que a utilização desta ferramenta não era uma atividade aborrecida porque não exige o preenchimento obrigatório dos campos de edição. Outras das observações foi a recomendação de uma integração mais harmoniosa das duas plataformas, através de um desenho de interface consistente entre ambas para não confundir o utilizador, assim como a necessidade de efetuar um único registo.

Na opinião do **participante E** as ferramentas demonstraram-se de utilização muito simples e bem conseguidas em relação a outros sistemas, sendo rápidas no desempenho e intuitivas. Embora não considerando como limitações sugeriu que as funcionalidades viessem acompanhadas com informação

⁵⁰ Ferramenta que se propõe a solucionar o problema da ambiguidade entre investigadores. <http://www.researcherid.com>

⁵¹ Ferramenta para pesquisa, organização e partilha de documentos. <http://endnote.com/>

relacionada, nomeadamente no caso da UPBox. Apesar de achar a linguagem bem aplicada lembrou que podia estar prevista uma versão em língua portuguesa.

Este participante referiu que em termos de partilha de dados a investigação teria a ganhar com a utilização de uma ferramenta criada para servir as necessidades dos investigadores. Todavia afirmou conhecer algumas iniciativas similares no seu domínio que já estão numa fase mais completa. Alertou, ainda, que os processos podem não ser tão lineares para quem não está familiarizado com a necessidade de descrição dos conteúdos produzidos. Alertou que deve ser explícito que diretivas estão a ser seguidas para descrever os dados, como forma de contornar problemas de ambiguidade entre domínios.

O **participante F** afirmou estar satisfeito após experimentar ambas as ferramentas, realçando a rapidez com que foi capaz de executar as tarefas, classificando os interfaces como visualmente agradáveis. Apesar de não ter cometido erros, excetuado a hesitação antes de executar algumas tarefas, referiu que acreditava que a curva de aprendizagem para qualquer tipo de utilizador terá uma evolução no sentido positivo. Em termos de linguagem não teve dificuldades de compreensão. Não deixou recomendações finais porque foi expondo as suas opiniões ao longo do desenvolvimento das tarefas.

Por fim o **participante G** considerou as plataformas UPBox e DataNotes como soluções interessantes para a partilha de dados entre colaboradores, particularmente quando estes são pesados, afirmando que trabalhar na UPBox é mais simples que na *Dropbox*. Este participante revelou ter conhecimento de iniciativas na sua área de investigação que vão ao encontro das necessidades de gestão dos investigadores, mas que no entanto se encontram na fase de conceptualização, e por esse motivo a disponibilização de ferramentas operacionais é vista como uma iniciativa com valor. Relembrou que as funcionalidades de anotar, eliminar e de descarregar dados não foram nem totalmente perceptíveis nem imediatas. Contudo, afirmou que uma vez apreendidos os procedimentos são facilmente reproduzidos. Para este participante a linguagem aplicada é adequada, muito embora o interface do DataNotes deva ser mais limpo porque a tendência é a exploração de todos os botões e os que não têm utilidade só fazem perder tempo. Nesse sentido avançou que a melhor alternativa era disponibilizar as funcionalidades básicas numa posição central do interface do DataNotes.

4.2.3 Recomendações UPBox e DataNotes

Após a realização dos testes de utilização das plataformas UPBox e DataNotes, foi possível listar uma série de sugestões que podem conferir maior eficácia e eficiência e contribuir para uma melhor adaptação dos utilizadores de ambas as plataformas. Na tabela 29 são apresentadas todas as recomendações que foram apuradas para a melhoria da UPBox e do DataNotes.

Tabela 29: Recomendações de melhoria UPBox e DataNotes

UPBox	
1	Dispor os projetos criados cronologicamente, do mais recente para o mais antigo.
2	Possibilitar ordenação dos projetos por ordem alfabética.
3	Implementar um sistema de controlo de versões, onde seja possível ver o histórico de alterações aos ficheiros.
4	Permitir a remoção de projetos criados.
5	Tornar evidente que tipo de informação é necessária para a adição de colaboradores.
6	Permitir a adição de elementos externos à Universidade do Porto, mediante as necessidades de partilha dos grupos de investigação.
7	Mostrar mensagem de erro quando o utilizador não preencher corretamente o campo de adição de colaborador, ou sugestão de correção.
8	Estabelecer o preenchimento semiautomático na adição de colaboradores.
9	Permitir carregar ficheiros quando utilizado o <i>browser Mozilla FireFox</i> .
10	Permitir visualizar conteúdos sem a necessidade de ter de se fazer <i>download</i> .
11	Mostrar mensagem de confirmação de <i>download</i> antes de o executar.
12	Permitir o <i>download</i> completo de diretórios, mediante compressão num único ficheiro.
13	Editar o cabeçalho com informação sobre as funcionalidades a executar.
14	Aumentar o tamanho dos ícones e reforçar com cores mais vivas (vermelho para eliminar).
15	Sinalizar função de anotação com um elemento iconográfico mais apropriado (por exemplo: lápis).
16	Disponibilizar uma caixa de ajuda no espaço livre do interface.
17	Ajustar a plataforma a ecrãs de menor dimensão.
DataNotes	
21	Colocar uma caixa de seleção de ficheiros numa posição de destaque.
22	Manter o nome do projeto visível durante a anotação dos ficheiros.
23	Indicar quais os metadados que devem ser obrigatoriamente preenchidos, com recurso a um sinal tipográfico (por exemplo: asterisco).
24	Tornar evidente quais os esquemas de metadados que estão a ser utilizados no momento da seleção de metadados.
25	Organizar a secção de metadados de acordo com o domínio para evitar o número elevado de elementos a percorrer na caixa de seleção de metadados.

26	Aplicar um esquema de cores que torne perceptível ao utilizador o estado corrente da anotação de determinado ficheiro, tornando o processo gradual e mais fácil de acompanhar.
27	Ajustar dimensões dos campos a preencher.
28	Possibilitar adicionar campos repetidos através da colocação de um botão (+) à frente dos campos, sem ter de os ir buscar à caixa de seleção de metadados.
29	Disponibilizar um calendário para inserção da data, para resolver problemas de ambiguidade.
30	Apresentar mensagem de confirmação do abandono da página no caso de terem sido feitas anotações não gravadas.
31	Associar descrição dos ficheiros à descrição feita na pasta “mãe”, para evitar redundância na anotação. Permitindo a anotação particular dos ficheiros caso seja necessário.
32	Dar maior relevo ao botão “ <i>Save page</i> ”.
33	Flexibilizar a tabela de descrições previamente editadas para permitir a eliminação ou edição de elementos instantaneamente.
34	Enquadrar as funções de maior relevância no centro da página inicial, assim como os ficheiros.
35	Editar a página de ajuda ao utilizador.
36	Mostrar campos previamente editados quando se regressa à página de edição.
37	Limpar elementos do interface que não sejam pertinentes para o utilizador.
UPBox + DataNotes	
38	Sincronizar o registo de utilizadores, para que apenas tenha de ser feito uma vez.
39	Harmonizar o interface entre as plataformas.
40	Implementar a comunicação com outras ferramentas de gestão de dados e redes sociais.
41	Sincronizar plataformas com o computador pessoal dos investigadores (expansão a uma aplicação cliente).

Conclusão

A gestão dos dados de investigação é uma atividade que levanta alguns problemas. Esses problemas estão relacionados com o registo incorreto dos dados, devido à falta de prática dos investigadores ou do tempo que eles podem dedicar a esta tarefa. A falta de apoio institucional é também um dos aspetos que condiciona a gestão dos dados de investigação, sendo que em muitos casos os investigadores não dispõem de infraestruturas que garantam o correto armazenamento e preservação dos dados. Para além disso a noção dos benefícios da partilha de dados entre a comunidade científica é uma questão que deve ser reforçada na mente dos investigadores.

Estas circunstâncias podem ter consequências severas para o rumo da investigação, uma vez que colocam entraves ao acesso e interpretação dos dados o que dificulta a reutilização dos mesmos, provocando bloqueios à evolução da ciência. Nesse sentido a comunidade científica internacional tem feito esforços para fomentar abordagens de curadoria de dados e implementar serviços relacionados com repositórios de dados, que não só incentivam, como em alguns casos, obrigam à disponibilização dos dados de investigação. Existem também casos em que as entidades financiadoras e as publicações científicas recomendam planos de gestão de dados ou a ingestão destes num repositório.

O trabalho desenvolvido nesta dissertação reflete esta problemática, enquadrando-se no processo de implementação na U. Porto de um repositório de dados e de ferramentas desenvolvidas para facilitar a gestão dos dados de investigação.

Relembrando os objetivos expostos nas páginas de introdução, este trabalho foi segmentado em duas vertentes. Em primeiro lugar procedeu-se à descrição de dados de investigação de diferentes domínios, nomeadamente através da proposta de perfis de aplicação para cada um deles. Em segundo lugar realizaram-se testes de utilização às ferramentas de gestão de dados, a UPBox e o DataNotes, e ao protótipo do Repositório de dados da Universidade do Porto, com a finalidade de apresentar estas ferramentas aos investigadores e fazer um levantamento de eventuais pontos fracos das mesmas.

Mediante os resultados obtidos pode-se afirmar que os objetivos propostos foram alcançados. No que diz respeito à elaboração de perfis de aplicação a confirmação do sucesso do cumprimento deste objetivo chegou através da escrita de um artigo científico, desenvolvido durante a elaboração desta dissertação, submetido e aceite para publicação nas atas da *International Conference on Dublin Core and Metadata Applications*, a decorrer em setembro de 2013.

Esse artigo, com o título “*Designing an Application Profile using qualified Dublin Core: a case study with fracture mechanics datasets*” consistiu num relatório de projeto onde foi descrita a experiência do levantamento das circunstâncias de gestão de dados grupo de investigação em fratura de materiais, e o desenho de um perfil de aplicação para a anotação dos dados produzidos por este grupo (ver capítulo 3). Os comentários dos revisores salientam que a proposta de perfil de aplicação é facilmente transferível para outras disciplinas, providenciado a informação necessária para perceber os conceitos fundamentais da experiência realizada. Para além disso o artigo foi considerado como um importante contributo para a comunidade DCMI.

Tendo em linha de conta a observação dos revisores sobre a facilidade em transpor o modelo para outros domínios, depreende-se que para os restantes perfis de aplicação propostos os resultados sejam relevantes. Isto, salvaguardando a impossibilidade da reprodução da experiência nos grupos de investigação em estruturas dinâmicas, gravimetria e ecologia georreferenciada. Sendo que nos dois primeiros grupos os elementos para descrição de dados sugeridos foram validados por especialistas no domínio, e no caso do grupo de investigação em ecologia georreferenciada foram adotados elementos da norma INSPIRE, que os mesmos utilizam para anotar as experiências que desenvolvem.

Em relação aos testes e demonstração das funcionalidades das plataformas foi possível levantar um número considerável de recomendações por parte dos participantes. Foram recolhidas um total de 17 sugestões de melhoria para o protótipo do Repositório de dados e 41 para as ferramentas de gestão de dados integradas, UPBox e DataNotes. Apesar do número elevado de recomendações as plataformas podem ser entendidas, nesta fase de desenvolvimento, como operacionais, tendo os participantes demonstrado, grosso modo, satisfação e vontade de usufruir destas ferramentas, em virtude do sentido de utilidade que reconhecem nas mesmas e na escassez de serviços deste género orientados aos investigadores. Os problemas identificados na interação dos participantes com as plataformas, uma vez superados, servem para minorar a possível resistência de futuros utilizadores, que experimentem as plataformas sem a presença de um intermediário, à adoção destas como ferramenta de trabalho.

Todavia, convém salientar que a experiência não decorreu sem limitações. Uma dessas limitações está relacionada com questões de agenda dos investigadores, que na maioria dos casos não permitiu efetuar mais do que uma sessão com os mesmos. Outra limitação foi a indisponibilidade em recorrer a ferramentas apropriadas ao desenvolvimento dos teste de utilização, face à realização dos mesmos no local de trabalho e computador pessoal dos investigadores.

As perspetivas de trabalho futuro, do ponto de vista da gestão de informação, devem ponderar a investigação no desenvolvimento de vocabulários controlados, esquemas de codificação e ontologias a

validar por especialistas dos mais diversos domínios, que obriguem à adoção de termos exatos para retirar a ambiguidade no momento da anotação dos dados. O estabelecimento de um vocabulário controlado vai também permitir uma percentagem maior na recuperação de resultados de interesse no momento da pesquisa, para além de orientar na construção de índices de pesquisa e classificadores de navegação no Repositório de dados da Universidade do Porto. Outro aspeto que pode ser alvo de investigação é o desenvolvimento de estratégias de parametrização no momento da ingestão e indexação dos dados no repositório, antecipando casos onde a anotação com o mesmo conteúdo informacional seja feita com recurso a metadados diferentes, nomeadamente a partir de esquemas distintos.

O processo descrito ao longo da presente dissertação permitiu identificar paradigmas diferentes ao nível da consciencialização da importância de adoção das melhores práticas de gestão de dados. Por um lado lidou-se com investigadores que não tinham sensibilidade para a questão da descrição de dados, em que conceitos como metadados ou descritores não eram familiares. Neste caso a necessidade da anotação não era encarada como uma oportunidade de solução para o transtorno de gerir o volume crescentes de dados, não havendo sequer a evidência da gestão de dados como um problema identificado. Por outro lado existem grupos que se encontram a trabalhar a gestão de dados a um nível de maturidade elevado, em consequência da sua participação em projetos que obrigam a uma maior preocupação em termos de catalogação e organização dos dados, ou simplesmente porque operam em domínios que levam um avanço relativo quando se trata de gestão e preservação de dados.

Esta observação levanta diferentes desafios e oportunidades para um profissional envolvido na curadoria de dados, devendo este moldar a sua abordagem ao tipo de perfil dos investigador de cada um dos paradigmas. Enquanto a mediação com grupos de investigação pouco experientes deve passar por um processo de educação dos mesmos, na interação com grupos mais experientes a postura a adotar poderá passar pela disposição à aprendizagem, tirando partido das melhores práticas de gestão de dados que estes tenham para oferecer.

Esta experiência possibilitou promover a consciencialização do valor das práticas de boa descrição de dados entre os grupos de investigação e à medida que cada vez mais investigadores começarem a descrever os seus dados, mais facilmente estes serão transpostos para o Repositório de dados da Universidade do Porto. Com a continuidade deste trabalho, possivelmente alargado a grupos de investigação de diferentes domínios, pode-se obter um alargamento do conhecimento sobre as práticas de gestão de dados no contexto de investigação realizada na U. Porto, e contribuir para melhorar as ferramentas de gestão de dados.

Contudo, e qualquer que seja o grau de experiência dos investigadores na gestão de dados, ficou patente no contato efetuado com os investigadores que é do interesse destes em ver desenvolvidas ferramentas que lhes permitam armazenar, preservar, recuperar e partilhar (com as devidas reservas) os dados que produzem.

Para terminar, e na esperança que a prática se generalize, o crescimento do Repositório de dados da Universidade do Porto pode motivar outros investigadores a valorizar a curadoria de dados, tendo em vista o aumento da visibilidade através da citação de dados.

Referências Bibliográficas

- Akmon, D. (2011). Moving Beyond Sharing vs Withholding to Understand How Scientists Share Data Through Large-Scale, Open Access Databases. *iConference* (pp. 634–635). Seattle: ACM.
- Akmon, D., Zimmerman, A., Daniels, M., & Hedstrom, M. (2011). The Application of Archival Concepts to a Data-Intensive Environment: Working with Scientists to Understand Data Management and Preservation Needs. *Archival Science*, 11(3-4), 329–348.
- Albanese, A. (2009). Institutional Repositories: Thinking Beyond the Box. *Library Journal*, 134(4), 26–28. Acessível em: <http://www.libraryjournal.com/article/CA6639327.html>
- AUSTRALIAN NATIONAL DATA SERVICE (2011). Metadata Guide Working Level. Intersect. Disponível em: <http://www.ands.org.au/guides/metadata-working.html>
- Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., & Moorman, D. (2004). Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal*, 3(November), 135–152.
- Asunka, S., Chae, H. S., & Natriello, G. (2011). Towards an understanding of the use of an institutional repository with integrated social networking tools: A case study of PocketKnowledge. *Library & Information Science Research*, 33(1), 80–88. doi:10.1016/j.lisr.2010.04.006
- Ball, A. (2006). Briefing Paper: the OAIS Reference Model, (February), 1–18. Acessível em: www.ukoln.ac.uk/projects/grand-challenge/papers/oaisBriefing.pdf
- Bawden, D. (2006). Users, user studies and human information behaviour: A three-decade perspective on Tom Wilson's "On user studies and information needs". *Journal of Documentation*, 62(6), 671–679. doi:10.1108/00220410610714903
- Barbosa, José Pedro (2013). UPBox: Armazenamento na Nuvem para Dados de Investigação na Universidade do Porto. Tese de Mestrado. *Faculdade de Engenharia da Universidade do Porto, Portugal*
- Beyer, H., & Holtzblatt, K. (1999). Contextual Design. *Interactions*, 6(1), 32–42.
- Belkin, N. (1980). Anomolous States of Knowledge as a Basis for Information Retrieval. *The Canadian Journal of Information Science*, 5, 133–143.

- Borgman, C. L., Wallis, J. C., Mayernik, M. S., & Pepe, A. (2007). Drowning in Data: Digital Library Architecture to Support Scientific Use of Embedded Sensor Networks. *JCDL*, June(18-23), 269–277
- Buneman, P. (2004) The two cultures of digital curation. Proceedings. International Conference on Scientific and Statistical Database Management. 7-7
- CCSDS. (2012). Reference Model for an Open Archival Information System (OAIS): Magenta Book, (June).
- Chowdhury, G. (2010). From digital libraries to digital preservation research: the importance of users and context. *Journal of Documentation*, 66 (2).207–223. doi:10.1108/00220411011023625
- COMISSÃO EUROPEIA (2010). *Sharing knowledge: open access and preservation in Europe*. Doi:10.2777/63410
- COMISSÃO EUROPEIA (2011). *National open access and preservation policies in Europe - Analyses of a questionnaire to the European Research Area Committee*. Doi:10.2777/74207
- Conway, P. (1986). Facts and Frameworks: An Approach to Studying the Users of Archives. *American Archivist*, 49(October 1985), 393–407.
- Cronin, B. (2001), “T.D. Wilson: an appreciation”, *Journal of Documentation*, Vol. 57 No. 1, pp. 1-5
- Crow, R. (2002). The Case for Institutional Repositories: A SPARC Position Paper. *The Scholarly Publishing and Academic Resources Coalition*, (August), 37.
- Gouveia, Mariana (2013) DataNotes – um sistema colaborativo para a anotação de estruturas de directórios. Tese de Mestrado. *Faculdade de Engenharia da Universidade do Porto, Portugal*.
- Giglia, E. (2011). Open Access, Open Data: Paradigm Shifts in the Changing Scholarly Communication Scenario: Conference Report. *D-Lib Magazine*, 17(3-4). doi:10.1045/march2011-giglia
- Hariri, N., & Norouzi, Y. (2011). Determining evaluation criteria for digital libraries’ user interface: a review. *The Electronic Library*, 29(5), 698–722. doi:10.1108/02640471111177116
- Heery, Rachel & Patel, Manjula (2000). Application Profiles: Mixing and Matching Metadata Schemas. *Ariadne*, 25 (September). Acessível em: <http://www.ariadne.ac.uk/issue25/app-profiles>
- Hey, T., Transley, S., & Kirstin, T. (2009). *The Fourth Paradigm: Data Intensive-Scientific Discovery*. Redmond.Washington: Microsoft Research.

- Jeng, J. (2005). Digital Libraries: Effectiveness, Efficiency, Satisfaction, and Learnability. *Libri*, 55, 96–121.
- Joo, S., & Lee, J. Y. (2011). Measuring the usability of academic digital libraries: Instrument development and validation. *The Electronic Library*, 29(4), 523–537. doi:10.1108/02640471111156777
- Kani-Zabihi, E., Ghinea, G., & Chen, S. Y. (2006). Digital libraries: what do users want? *Online Information Review*, 30(4), 395–412. doi:10.1108/14684520610686292
- Kellie, S., Ballaux, B., Christensen-Dalsgaard, B., Hofman, H., Hansen, J. H., Innocenti, P., Nielsen, M. P., et al. (2008). Considering the User Perspective Research into Usage and Communication of Digital Information. *D-Lib Magazine*, 14(5/6).
- Kim, H. H., & Kim, Y. H. (2008). Usability study of digital institutional repositories. *The Electronic Library*, 26(6), 863–881. doi:10.1108/02640470810921637
- Li, Y., & Banach, M. (2011). Institutional Repositories and Digital Preservation: Assessing Current Practices at Research Libraries. *D-Lib Magazine*, 17(5/6). doi:10.1045/may2011
- Lyon, Liz (2010) Data Dimensions : Disciplinary Differences in Research Data Sharing , Reuse and Long term Viability A comparative review based on sixteen case studies. *A report commissioned by the DCC and SCARP Project, Key Perspectives Ltd*, (January)
- Michener, W. K. (2006). Meta-information concepts for ecological data management. *Ecological Informatics*, 1(1), 3-7.
- Mick, C. K., Lindsey, G. N., & Callahati, D. (1980). Toward Usable User Studies. *Journal of the American Society for Information Science*, 347–356.
- Murray-Rust, P. (2008). Open Data in Science. *Serials Review*, 34(1), 52–64. doi:10.1016/j.serrev.2008.01.001
- National Science Board (2005). Long-lived digital data collections: Enabling research and education in the 21st century. Disponível em <http://www.nsf.gov/pubs/2005/nsbo540/>. Consultado a 9 de abril de 2013
- Nielsen, J. (2003). “Usability 101: Introduction to usability”. *Useit.com: Usable Information Technology*, August. Disponível em: <http://www.nngroup.com/articles/usability-101-introduction-to-usability/>
- Norberg, L. R., Vassiliadis, K., Ferguson, J., & Smith, N. (2005). Sustainable design for multiple audiences: The usability study and iterative redesign of the Documenting the American South digital library. *OCLC Systems & Services*, 21(4), 285–299. doi:10.1108/10650750510629625

- Pfeiffenberger, H., & Carlson, D. (2011). Pfeiffenberger and Carlson (2011) Earth System Science Data (ESSD) - A Peer Reviewed Journal for Publication of Data. *D-Lib Magazine*, 17(1-2). doi:10.1045/january2011-pfeiffenberger
- Piwowar, H. a, Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PloS one*, 2(3), e308. doi:10.1371/journal.pone.0000308
- Ray, J. (2012). The rise of digital curation and cyberinfrastructure: From experimentation to implementation and maybe integration. *Library Hi Tech*, 30(4), 604–622. doi:10.1108/07378831211285086
- Ribeiro, Cristina & Fernandes, Maria Eugénia Matos (2011). Data Curation at U. Porto: Identifying current practices across disciplinary domains. *IASSIST Quarterly*, Winter 2011: 14-17
- Rice, Robin & Haywood, Jeff (2011). Research Data Management Initiatives at University of Edinburgh. *The International Journal of Digital Curation*, 6(2), 232–244.
- Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PloS one*, 4(9), e7078. doi:10.1371/journal.pone.0007078
- Shiri, A., Ruecker, S., Bouchard, M., Stafford, A., Mehta, P., Anvik, K., & Rossello, X. (2009). User evaluation of Searchling: a visual interface for bilingual digital libraries. *The Electronic Library*, 29(1), 71–89. doi:10.1108/02640471111111442
- Siatri, R. (1999). The Evolution of User Studies. *Libri*, (48)3, 132-141. doi:10.1515/libr.1999.49.3.132
- Smit, E., Van Der Hoeven, J., & Giaretta, D. (2011). Avoiding a Digital Dark Age for data: why publishers should care about digital preservation. *Learned Publishing*, 24(1), 35–49. doi:10.1087/20110107
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., et al. (2011). Data sharing by scientists: practices and perceptions. *PloS one*, 6(6), e21101. doi:10.1371/journal.pone.0021101
- Thomas, C., & McDonald, R. (2007). Measuring and Comparing Participation Patterns In Digital Repositories, Repositories by the Numbers Part I. *D-Lib Magazine*, 13(9/10).
- Treloar, A. (2009). Design and Implementation of the Australian National Data Service. *The International Journal of Digital Curation*, 4(1), 125–137.
- Wallis, J. C., Mayernik, M. S., Borgman, C. L., & Pepe, A. (2010). Digital Libraries for Scientific Data Discovery and Reuse : From Vision to Practical Reality. *JCDL*, June(21-25), 333–340.

- Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PloS one*, 6(11), e26828. doi:10.1371/journal.pone.0026828
- Wildemuth, B. M. (2003). Why Conduct User Studies? The Role of Empirical Evidence in Improving the Practice of Librarianship by. In Inforum (Ed.), “*Fine-Tuning information Strategies*” *INFORUM 2003; 9th Conference on Professional Information Resources*. Praga.
- Willis, Craig, Greenberg, Jane & White, Hollie (2012). Analyses and Synthesis of Metadata Goals for Scientific Data. *Journal of the American Society for Information Science*, 63(8): 1505-1520
- Wilson, J. A. J., Martinez-Urbe, L., Fraser, M. A., & Jeffreys, P. (2011). An Institutional Approach to Developing Research Data Management Infrastructure. *The International Journal of Digital Curation*, 6(2), 18–31.
- Wilson, T. D. (1981). on User Studies and Information Needs. *Journal of Documentation*, 37(1), 3–15. doi:10.1108/eb026702
- Wilson, T. D. (2000). Recent trends in user studies action research and qualitative methods. *Information Research*, 5(3).
- Wira-Alam, Andias, Dimitrov, Dimitar & Zenk-Möltgen, Wolfgang. (2012) Extending Basic Dublin Core for an Open Research Data Archive. *Proceedings of the Internation Congeference on Dublin Core and Metadata Applications*, 2012, 56-61
- Zhang, X., & Str, H. (2009). How Usable are Operational Digital Libraries – A Usability Evaluation of System Interactions. *EICS, July(15-17)*, 177–186.
- Zhang, Y. (2010). Developing a Holistic Model for Digital Library Evaluation. *Journal of the American Society for Information Science*, 61(2000), 88–110. doi:10.1002/asi

ANEXOS

Anexo I – Recomendações de esquemas para descrição de dados de investigação (DCMI)

RESEARCH DATA ALLIANCE (RDA) METADATA DIRECTORY

Access to Biological Collection Data (ABCD)

- Description: The Access to Biological Collections Data (ABCD) Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). ABCD version 2.06 has been recommended by the TDWG meeting in St. Petersburg as the actual standard and has since then been ratified by TDWG members.
 - Identifier of Latest Version: 2.06 <http://www.tdwg.org/standards/115/>
 - Term: <http://wiki.tdwg.org/twiki/bin/view/ABCD/AbcdPrimer>
 - Domain/Discipline/Subject: Biodiversity
 - Maintenance Authority: Access to Biological Collections Data Task Group <http://www.tdwg.org/activities/abcd/>
 - Year of Creation:[2005]
-

Astronomy Visualization Metadata Standard (AVM)

- Description: AVM focuses on print-ready and screen ready astronomical imagery, which has been rendered from telescopic observations. Such images can combine data acquired at different wavebands and from different observatories. The scheme is also appropriate for describing artwork and illustrations of astronomical subject matter.
 - Identifier of Latest Version: 1.2 http://www.virtualastronomy.org/AVM_DRAFTVersion12_rlh02.pdf
 - Terms: http://www.virtualastronomy.org/AVM_12_ref_table_rlh02_print2.pdf
 - Domain/Discipline/Subject: Astronomy, Geography
 - Maintenance authority: Virtual Astronomy Multimedia Project (VAMP) <http://www.virtualastronomy.org/>
 - Year of Creation: 2008
-

Content Standard for Digital Geospatial Metadata (CGDGM)

- Description: The standard provides a common set of terminology and definitions for the documentation of digital geospatial data.
 - Identifier of Latest Version: 2.0 <http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata>
 - Terms: http://www.fgdc.gov/metadata/documents/workbook_0501_bmk.pdf
 - Domain/Discipline/Subject: Geography, Biology, Oceanography
 - Maintenance authority: Federal Geographic Data Committee Secretariat <http://www.fgdc.gov/>
 - Year of Creation: 1994
-

Darwin Core (DwC)

- Description: The Darwin Core is body of standards. It includes a glossary of terms intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries. The Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, and samples, and related information.
 - Identifier of Latest Version: unknown, <http://rs.tdwg.org/dwc/>
 - Terms: <http://rs.tdwg.org/dwc/terms/>
 - Domain/Discipline/Subject: Biology, Species analysis, Organisms
 - Maintenance Authority: Darwin Core Task Group <http://www.tdwg.org/activities/darwincore/>
 - Year of Creation: 1998
-

Data Documentation Initiative (DDI)

- Description: DDI is an effort to create an international standard for describing data from the social, behavioral, and economic sciences. Expressed in XML, the DDI metadata specification now supports the entire research data life cycle. DDI metadata accompanies and enables data conceptualization, collection, processing, distribution, discovery, analysis, repurposing, and archiving.
 - Identifier of the Latest Version: 3.1 <http://www.ddialliance.org/Specification/DDI-Lifecycle/3.1/>
 - Terms: <http://www.ddialliance.org/Specification/DDI-Codebook/2.5/>
 - Domain/Discipline/Subject: Social, Behavioral, and Economic Sciences
 - Maintenance Authority: DDI Alliance <http://www.ddialliance.org/alliance>
 - Year of Creation: 2000
-

Directory Interchange Format (DIF) Writer's Guide

- Description: The Directory Interchange Format is used to create directory entries that describe a group of data. A DIF consists of a collection of fields which detail specific information about the data.
 - Identifier of Latest Version: unknown, <http://gcmd.gsfc.nasa.gov/User/difguide/difman.html>
 - Terms: <http://gcmd.nasa.gov/User/difguide/WRITADIF.pdf>
 - Domain/Discipline/Subject: General Science
 - Maintenance Authority: Global Change Master Directory, National Aeronautics and Space Administration
 - Year of Creation: [1988]
-

Dublin Core Metadata Element Set (DC)

- Description: The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description. The name “Dublin” is due to its origin at a 1995 invitational workshop in Dublin, Ohio; “core” because its elements are broad and generic, usable for describing a wide range of resources.
 - Identifier of Latest Version: 1.1 <http://dublincore.org/documents/dces/>
 - Terms: <http://dublincore.org/documents/dces/>
 - Domain/Discipline/Subject: General purpose metadata standard for all disciplines
 - Maintenance Authority: Dublin Core Metadata <http://dublincore.org/about-us/>
 - Year of Creation: 1995
-

Ecological Metadata Language (EML)

- Description: The Ecological Metadata Language is a metadata standard developed by the ecological discipline. It is implemented in an Extensible Markup Language (XML) as a series of document types used in a modular manner to document ecological data.
 - Identifier of Latest Version: 2.1.1<http://knb.ecoinformatics.org/software/eml/>
 - Terms:<http://knb.ecoinformatics.org/software/eml/eml-2.1.1/index.html>
 - Domain/Discipline/Subject: Ecological
 - Maintenance Authority: The Knowledge Network for Biocomplexity
 - Year of Creation: [2008]
-

NetCDF Climate and Forecast (CF) Metadata Convention

- Description: The conventions for climate and forecast (CF) metadata are designed to promote the processing and sharing of files created with the NetCDF API.
 - Identifier of Latest Version: 1.6,<http://cf-pcmdi.llnl.gov/>
 - Terms:<http://cf-pcmdi.llnl.gov/documents>
 - Domain/Discipline/Subject: Climate and Forecast, Earth Science
 - Maintenance Authority: CF Governance
 - Year of Creation: [2003]
-

NISO Metadata for Images in XML

- Description: NISO Metadata for Images in XML (MIX) is a metadata schema for a set of technical data elements required to manage digital image collections
 - Identifier of Latest Version: 2.0<http://www.loc.gov/standards/mix/>
 - Terms:http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=b897b0cf3e2ee526252d9f830207b3cc9f3b6c2c
 - Domain/Discipline/Subject: Digital Images
 - Maintenance Authority: Library Of Congress Network Development and MARC Standard Office, NISO Technical Metadata for Digital Still Images Standards Committee
 - Year of Creation: [2006]
-

Shoreline Metadata Profile of the Content Standards for Digital Geographic Data Committee (FGDC)

- Description: FGDC is used as an extension or profile to the existing Content Standard for Digital Geospatial Metadata (CSDGM). The metadata profile is to capture the critical processes and conditions that revolve around creating and collecting shoreline data.
- Identifier of Latest Version: unknown,http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/shoreline-metadata/index_html
- Terms:http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/shoreline-metadata/sp_endorsed.pdf
- Domain/Discipline/Subject: Geospatial, Shoreline, Coastal
- Maintenance Authority: NOAA Coastal Services Center

- Year of Creation: 2001

•

About this Project

The RDA prototype directory is being developed as part of the Research Data Alliance (RDA) (<http://rd-alliance.org/>) effort in collaboration with the DCMI SAM community, the DataONE DataNet, and other partners to-be-confirmed. Graduate students developing the registry are affiliated with the Metadata Research Center, School of Information and Library Science, University of North Carolina at Chapel Hill (<http://ils.unc.edu/mrc/>).

Anexo II– Guião de entrevista (Data Curation Tool Profile).

RELATÓRIO DA ENTREVISTA AO PROFESSOR [REDACTED]

Departamento de Mecânica Aplicada – Feup

08/03/2013

Descrição do contexto de investigação associado aos dados.

O contexto dos dados vem de ensaios de fratura de madeira em que nós (...) aplicamos uma solicitação na máquina e vai-se propagando uma fenda e nós medimos a energia associada a propagação dessa fenda. O que registamos em termos de dados é a evolução da força e deslocamento, que são os resultados que a máquina fornece, é um software que captura a curva força-deslocamento, que são os dados que vos forneci, que estão em bruto, e que depois são tratados para se chegar à energia de fratura, sendo esse o objetivo final.

1 – Ciclo de vida dos dados.

Pode fazer uma descrição do *data set*?

Os investigadores são alunos de doutoramento, que na fase experimental podem ser acompanhados pelo doutor Nuno Dourado, e alguns alunos de mestrado.

Os dados são registados na máquina (ver artigo), passados para excell e trabalhados com recurso a métodos analíticos próprios da área que permitem obter os dados pretendidos. O método passa por converter os dados (força e deslocamento) em valores de energia, através do registo direto no excell.

Duas fases: Captura dos dados com registo no excell e tratamento; cada ensaio produz um ficheiro; o tamanho dos ficheiros é em mb´s. O output é a máquina descrita no artigo e que fornece resultados em excell.

A segunda fase corresponde ao tratamento dos dados aplicando o método analítico.

2– Partilha e Acesso.

Que dado (em cada uma das etapas do ciclo de vida) está disponível a partilhar?

Partilhados em aberto dentro do grupo em cada uma das etapas, para o exterior já é mais complicado, não é fácil; partilhamos os dados convosco porque já foram publicados.

Partilha, atualmente, esses dados com alguém?

São sempre partilhados dentro do grupo mas nunca fora. Temos cuidados porque dá muito trabalho

recolher os dados e não queremos que chegue alguém e se aproprie deles, não publicamos prematuramente antes da publicação, para não correremos o risco de ser divulgados por outros antes de nós. Somos um grupo de investigação e entre nós não a segredo. O nosso grupo é quase todo exterior à Feup, a maioria da UTAD e Aveiro e não propriamente da FEUP, poucos elementos do grupo são da Universidade do Porto, e é com esses elementos com os quais a partilha é feita.

Não existe método para a partilha, é feita através da Dropbox por exemplo, ou mesmo copiado do disco com a presença dos alunos, não existem restrições de confidencialidade na partilha entre nós porque publicamos em conjunto, e a partilha é feita no momento da produção dos dados.

Disponibilidade para partilha

Etapa	Não partilho	Colaboradores mais próximos	Pessoas da mesma instituição	Investigadores da área	Outros investigadores	Com qualquer um
1		x				
2		x				

Em que etapas do ciclo de vida, os dados podem ser divulgados num repositório? Quais as principais preocupações na submissão de dados num repositório de dados aberto?

Não podemos divulgar os dados em aberto, são dados experimentais.

É necessário um periodo de tempo aos quais os dados têm de estar fechados, nomeadamente enquanto os alunos terminam o doutoramento, enquanto se aguarda pela publicação dos dados.

O objetivo é estar disponíveis para circular entre o grupo.

Produzir dados dá muito trabalho, é um investimento grande, a compra de materiais envolve dinheiro, e não podemos arriscar tornar os dados públicos sem ter tudo devidamente publicado.

Após a submissão de dados, pretende requerer um periodo de embargo desses dados?

Sem embargo, acesso imediato -

1-3meses- 4-6 meses- 7meses a 1 ano- 1 a 2 anos-
2 a 5 anos- mais de 5 anos- x Outro

Prioridade do serviço

	Nenhuma	Baixa	Média	Alta	Não sei
Aptidão para citar data sets nas minhas publicações	x				
Data set deve ser citado por outros quando usados			x		
Aptidão para restringir o uso de data sets a outros utilizadores				X – restrito a elementos do grupo	

Outro serviço – um serviço que me agradaria era ter a capacidade de partilhar com facilidade os dados com os outros elementos, e parece que o vosso serviço podia facilitar isso, interessa-nos ter os dados depositados e facilmente consultados e outras coisas que nos interessam. Todavia existe o problema de ter de pedir os acessos de elementos externos, que correspondem a maioria das pessoas do nosso grupo, este é o maior problema.

Não me ocorre a necessidade de ter aptidão para citar as minhas próprias publicações. Não existe prioridade nesse aspeto. Não temos necessidade de citar dados que dão origem aos dados. Mais importante que citar o dataset é importante a citação do artigo quando esses dados são usados.

O acesso tem de estar restrito aos elementos do grupo. A partir do momento do embargo e publicação não posso dar resposta por mim mesmo, eventualmente, mas teria de ser discutido com colegas da UTAD, uma vez que os dados são gerados lá, aqui tratamos questões de ordem teórica e numérica.

3– Transferência de dados / Ingestão dos dados num Repositório.

Que ações devem estar previstas para que os dados possam ser depositados num repositório ou transferidos do seu controlo para curadoria?

Isso seria feito por nós, não teríamos dificuldade em fazer isso caso o serviço seja user-friendly. O entrevistado mostrou vontade de manter a sua autonomia.

4 – Organização e Descrição dos dados.

Como são os datasets organizados e descritos (anotações detalhadas; tabelas, etc..)?

Utilizada algum esquema standard para a descrição de dados? Qual?

É necessária alguma descrição adicional para que alguém com conhecimentos similares possa interpretar os dados?

Existe informação adicional (contextual ou documental) necessária para que outros interpretem e usem os dados?

A informação em princípio não é detalhada, diria que é mesmo escassa. Basicamente todos os elementos estão dentro do assunto, e conseguem interpretar os dados sem dificuldade. A informação

circula bem entre pessoas que conhecem o domínio. Nos dados que cedi tem uns desenhos que para elementos do grupo não seriam precisos. Quem chega vem preparado para compreender os dados.

Não utilizamos esquemas de metadados standard.

Pessoas da mesma área não precisam de informação adicional para compreender os dados em brutos, qualquer aluno saído da licenciatura vai compreender esses dados. Na segunda fase é que já é preciso informação adicional para interpretar os dados trabalhados, é necessário conhecer o método, que está publicado nos artigos, e não acompanham os dados.

Pessoas fora da área não deveriam ter interesse nos dados, a não ser ligadas a engenharia mecânica, e em concreto aqueles dentro da caracterização da fratura dos materiais.

5- Propriedade Intelectual

Quem é/são o(s) proprietário(s) dos dados?

Quais são as fontes de financiamento?

Existem outros stakeholders associados aos dados?

Os financiadores exigem:

Um plano de gestão de dados- Não

Partilhar, publicar ou depositar os dados num repositório- Não

Preservação dos dados para além da “vida” do financiamento- Não

Proteção dos dados através de uma política de privacidade/confidencialidade- Não

Eu diria que são os membros do grupo, em particular da UTAD porque é lá que os dados são produzidos. Diria que essa propriedade é partilhada porque é cá que os dados são tratados.

Quase todos os dados são produzidos sob financiamento da FCT, mas não tenho conhecimento de exigências da parte deles, a não ser uma menção nas publicações, isso é o que estamos obrigados a fazer, em relação aos dados não tenho esse conhecimento.

Não existem outros stakeholders.

6- Interoperabilidade.

Nos periódicos em que publica, com maior regularidade, os dados, ou documentação associada, são aceites para publicação?

Não nos pedem essa informação, os gráficos resultantes dos dados vão no artigo. Não existe nenhum interesse em relação aos dados em bruto ou exigência de disponibilização dos dados num repositório.

7- Factor de impacto

Prioridade do serviço

	Nenhuma	Baixa	Média	Alta	Não Sei
Ter conhecimento de estatísticas sobre o uso dos dados				x	
Ter informação sobre quem usou ou acedeu aos dados				x	

Que informação gostaria de obter sobre os utilizadores que acedem ou usam os seus dados?

Para além de estatísticas de uso, há outras medidas ou métodos de análise que gostaria de ver associados aos seus dados?

Gostaria de saber a atividade profissional deles, o que andam a estudar, isso seria interessante. Pode dar impulso a novas colaborações e portanto o entrevistado vê a hipótese de obter informação sobre os utilizadores, um factor importante.

Não demonstrou interesse nouro tipo de recolha de informação

8- Gestão e Preservação dos dados.

Que atividades desenvolve para gerir os seus dados? (Incluir software de armazenamento, ou outras ferramentas usadas para fazer essa gestão.)

Faz *back-ups* desses dados?

Atualmente toma medidas de proteção desses dados? Quais?

Basicamente faço back-ups, mais do que um, através de um disco rígido, uma vez por mês.

Quais são as partes (etapas do ciclo de vida) dos dados mais importantes a preservar? (a gerir a longo prazo).

Os dados tratados são mais importantes, apesar de os dados tratados incluírem os dados em bruto. No entanto, se tivesse de escolher entre quais perder, não me importava de perder os dados em bruto.

Os dados que pretende partilhar são os que precisam de ser preservados? Existem componentes adicionais que precisam de ser preservadas para além dos dados a partilhar?

Certo. Uma coisa que o meu colega se queixa com pertinência é da falta de organização das figuras. Temos dificuldade em encontrar figuras, que necessitamos de usar com frequência. Temos figuras que se repetem. Não sabemos qual é a figura original e perdemos muito tempo a procurar essa figura para fazermos alterações às mesmas. Seria útil um serviço de gestão e partilha das imagens produzidas internamente para contornar esse problema de localização das imagens.

Durante quanto tempo os dados mantêm valor tanto para si, como para potenciais utilizadores?

Sim, estes dados são importantes. Às vezes pegamos em dados que já não víamos há muito tempo. Os dados não ficam datados. São dados de observação e apesar de poderem ser reproduzidos, os resultados obtidos não são necessariamente os mesmos, mesmo quando por exemplo se estuda a mesma árvore no exterior. Os dados estão relacionados com a amostra e não são obrigatoriamente reproduzidos mesmo conhecendo os métodos e materiais. No caso de material artificial talvez seja mais fácil reproduzir estes dados.

Não precisam de ser preservados -

Menos de 3 anos-

3 a 5 anos-

5 a 10 anos-

Mais de 10 anos-

Indefinido- **x**

FIM!

Anexo III – Guião teste de utilização Repositório de dados da Universidade do Porto

TESTE DE UTILIZAÇÃO DO REPOSITÓRIO DE DADOS DE INVESTIGAÇÃO DA UNIVERSIDADE DO PORTO

Investigador:

Data:

Questionário pré-tarefas:

- 1- Quais as plataformas que utiliza para pesquisar informação científica?
- 2- Utiliza algum repositório digital para armazenar, recuperar dados de investigação?
- 3- Considera-se um utilizador experiente neste tipo de plataforma?
- 3- No geral, o que acha desta plataforma?

Tarefas:

1. Visualizar coleção “Aereal Gravimetry Runs”
2. Visualizar ficheiros da autoria de “Richard Deurloo”
3. Visualizar ficheiros do assunto “azores”
4. Ordenar ficheiros da coleção “Aereal Gravimetry Runs”, de acordo com a data de emissão, por ordem descendente.
5. Visualizar os ficheiros relativos a “Final Results of the gravimetry runs”; mostrar registo de metadados e abrir um dos ficheiros
6. Explore dados relativos a “Final Results of the gravimetry runs” e visualize os metadados associados e termine com o download dos dados
7. Filtre os dados com altitude igual ou superior a “110.846” e inferior a “110.911”
8. Subscreeva coleção “Aereal Gravimetry Runs”

Questionário pós-tarefas:

- 1- Principais impressões – Pontos positivos e negativos?
- 2- Quais as principais dificuldades sentidas na realização das tarefas?
- 3- Considera a linguagem utilizada nesta plataforma apropriada?
- 4- Se voltasse a realizar as tarefas acha que as faria sem qualquer dificuldade?
- 5- Considera esta plataforma útil? Em que sentido? Em comparação com outras plataformas?
- 6- Sugestões de funcionalidade - O que mudaria nesta plataforma?
- 7- Numa escala de 0 a 10, como avalia o desempenho desta plataforma?
- 8- Numa escala de 0 a 10, qual o seu nível de satisfação após este experimentar esta plataforma?

Tarefa	Regressa página inicial	Pede ajuda	Erro	Abandona tarefa	Tempo despendido
1					
2					
3					
4					
5					
6					
7					
8					

FIM!

Anexo IV – Guião de teste de utilização das ferramentas UPBox e DataNotes

TESTE DE UTILIZAÇÃO DA UPBOX E DATANOTES

Investigador:

Data:

Tarefas:

- 1.Criação de um projeto
- 2.Partilhar ficheiro com um colaborador
- 3.Criar uma pasta e adicionar ficheiro(s)
- 4.Efetuar download de ficheiro
- 5.Remover ficheiro do projeto
- 6.Iniciar processo de anotação
- 7.Iniciar anotação (Selecionar ficheiro a anotar)
- 8.Editar anotação:
 - dc.title*
 - dc.subject*
 - dc.creator*
 - dc.date*
 - dc.description*
 - Guardar página
- 9.Eliminar uma das anotações
- 10.Regressa a página inicial e pesquisar pelo nome do criador/autor
- 11.Voltar a página de anotação e adicionar campo
 - dc.type*

Questionário pós-tarefas:

- 1- Principais impressões – Pontos positivos e negativos?
- 2- Quais as principais dificuldades sentidas na realização das tarefas?
- 3- Considera a linguagem utilizada nesta plataforma apropriada?
- 4- Se voltasse a realizar as tarefas acha que as faria sem qualquer dificuldade?
- 5- Considera esta plataforma útil? Em que sentido? Em comparação com outras plataformas?
- 6- Sugestões de funcionalidade - O que mudaria nesta plataforma?

Tarefa	Regressa página inicial	Pede ajuda	Erro	Abandona tarefa	Tempo dispendido
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					

FIM!

Anexo V – Exemplo de relatório de teste de utilização

Relatório do teste de usabilidade da UPBox e do DataNotes

Investigador: ██████████ | Departamento de Engenharia Mecânica

Data: 14 de março de 2013

Tarefas:

1. Criação de um projeto
2. Partilhar ficheiro com um colaborador
3. Criar pasta e adicionar ficheiros
4. Efetuar download de ficheiro
5. Remover um ficheiro do projeto
6. Iniciar processo de anotação
7. Selecionar ficheiro a anotar
8. Editar anotação:
 - dc:title
 - dc:subject
 - dc:creator
 - dc:date
 - dc:description
 - guardar página
9. Eliminar uma das anotações
10. Regressa a página inicial e pesquisar pelo nome do criador/autor
11. Voltar a página de anotação e adicionar campo
 - dc:type

Tarefa	Regressa à página inicial	Pede ajuda	Comete erro	Abandona tarefa	Tempo dispendido
1					38 segundos
2		2			54 segundos
3					42 segundos
4					6 segundos
5					9 segundos
6		4	2		1 minuto e 06 segundos*
7					37 segundos
8					3 minutos e 56 segundos
9		2	1		37 segundos
10					16 segundos
11	3	2	4	x	Não terminou tarefa

Comentários:

Tarefa 1- Nada a apontar

Tarefa 2 - O utilizador revelou alguma dificuldade em executar esta tarefa. Durante a realização da mesma fez as seguintes perguntas:

- *Onde está o partilhar aqui?*

- *Clico aqui?*

E afirmou – *Não sei como fazer.*

- *Talvez necessite de afinação.*

Tarefa 3 - O utilizador comenta a facilidade em adicionar folder, dizendo que está claro, assim como o upload e escolha de ficheiro foram consideradas como atividades óbvias do ponto de vista do utilizador.

Tarefa 4 - A tarefa foi executada de forma intuitiva em simultâneo com a proposta para proceder ao download do ficheiro.

Tarefa 5 - Nada a apontar

Tarefa 6- Esta tarefa revelou-se complexa. O utilizador ficou confundido com a forma pela qual se faz a transição para o DataNotes, ou seja, como começar o processo de anotação. Os comentários feitos durante a tarefa foram os seguintes:

- *Não estou a ver*
- *É isto aqui, UPBox*
- *Daqui não consigo sair não é?*
- *Ainda não percebi bem, existem duas plataformas, UPBox e esta?*
- *Há algum link para lá?*

O problema foi a falta de percepção em relação ao botão que encaminha o utilizador do UPBox para o DataNotes – em vez do ícone que sugere anotação deverá aparecer por extenso algo relacionado com a anotação.

Tarefa 7 - O participante não cometeu erros mas levou algum tempo (37 segundos) à procura do botão para seleccionar ficheiros

Tarefa 8 - O utilizador não teve dificuldade em adicionar os descritores pedidos, o tempo de execução da tarefa tem a ver com o conteúdo da descrição.

Tarefa 9 - O utilizador passou por algumas dificuldades para voltar a entrar na edição e eliminar um dos descritores, alguns comentários feitos durante esta tarefa foram os seguintes:

- *Não sei o que quer dizer “editform”*
- *Sei lá!*
- *Qual a diferença entre “edit” e “editform”?*

O erro cometido está ligado a exploração do botão “edit”.

Após a tarefa seguiu-se uma explicação da diferença entre ambos os botões!

O utilizador afirmou que a funcionalidade do botão “edit” não presta.

Tarefa 10 - Nada a apontar

Tarefa 11 - O utilizador revelou dificuldades em adicionar novo descritor. Os comentários feitos foram os seguintes:

- *Não está evidente.*
- *Newpage?*
- *Não consigo lá chegar.*
- *Não sei.*
- *Não sei lá chegar*
- *Será isto?*

Com esta tarefa foi possível detetar um erro no serviço. Esse erro foi a falta de apresentação dos descritores anteriormente editados. O utilizador fica assim sem saber quais os campos já preenchidos, o que se agrava se for um projeto partilhado. Só quando seleccionados de novo é que se percebe quais os campos que já estão preenchidos. De qualquer forma o utilizador não conseguiu adicionar o descritor proposto.

Impressões Gerais:

O utilizar considerou o interface da UPBox como user friendly. A linguagem foi definida como clara. O utilizador diz que a plataforma UPBox é semelhante a que utiliza habitualmente, a *Dropbox*.

A maior dificuldade sentida foi em relação ao processo de adicionar colaborador, uma vez que essa função não foi assimilada de forma intuitiva, pelo que o utilizador sugere uma afinação, no entanto considera que após saber como fazer não terá dificuldade em repetir o procedimento. Tirando um ou dois passos a experiência foi tida como positiva. O utilizador considera-se satisfeito com a experiência, mas acrescenta que há coisas difíceis de perceber, principalmente quando se trata de voltar a editar um ficheiro ou pasta.

A linguagem foi classificada como clara. O utilizador nunca tinha trabalho com nada deste género, acreditando que este é um serviço útil porque a informação fica organizada, registada e permite autonomia, facilitando a gestão dos dados. Não lhe ocorreu nada a sugerir mas mostrou desagrado em relação ao excesso de informação, porque segundo a maioria dos descritores não lhe interessam e mais de metade fica por preencher.

Notas:

Durante a conversa após o final da experiência ficou claro o cenário ideal no qual os ficheiros que

pertencem a uma pasta-mãe descrita, assumem automaticamente os valores da última para evitar trabalho redundante. Contudo esses ficheiros podem também ser editados mediante a sua especificidade.

Outra das alterações a sugerir que resultam desta experiência é a necessidade de alargamento das plataformas a utilizadores fora da U. Porto, isto porque os colaboradores deste utilizador são investigadores externos a esta instituição.

FIM!