

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

DendroShare: Partilha e citação de conjuntos de dados de investigação

Lázaro Gabriel Barros da Costa



FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia Informática e Computação

Orientador: João Miguel Rocha da Silva, PhD

30 de Julho de 2019

DendroShare: Partilha e citação de conjuntos de dados de investigação

Lázaro Gabriel Barros da Costa

Mestrado Integrado em Engenharia Informática e Computação

Resumo

O crescente poder computacional ao dispor da ciência tem vindo a tornar a gestão dos dados de investigação um dos temas da atualidade. A ligação dos artigos científicos aos seus dados base torna possível a validação e clarificação do seu conteúdo, facilita a reutilização dos dados por terceiros e atribui o devido crédito aos criadores dos dados. Apesar destas preocupações, nem todos os conjuntos de dados de investigação são identificados e armazenados, o que todos os anos coloca em risco muitos conjuntos de dados de investigação.

O conceito de Open Science é a prática da ciência de tal forma que possibilita a colaboração e contribuição da sociedade na reprodução de dados de investigação, relatórios entre outros processos de pesquisa, assim como o seu livre acesso, redistribuição e utilização para posteriores pesquisas. Enquanto que Open Data compreende a disponibilização dos dados em acesso aberto, o que permite impulsionar o desenvolvimento de novas ideias e produtos.

Com a disponibilização de dados surgiram os princípios FAIR, com o objetivo de facilitar o processo de pesquisa, acessibilidade, interoperabilidade e reutilização dos diferentes dados, quer por humanos, que por máquinas. Facilitou o avanço na ciência, reduziu os custos na reprodução de resultados duplicados e melhorou o aproveitamento do tempo. Ao mesmo tempo, a associação de dados FAIR às publicações fomentam a reprodutibilidade dos resultados obtidos assim como a verificação dos mesmos.

Os vários repositórios de armazenamento de dados de investigação abordam o tema do preenchimento dos metadados de maneira diferente: a plataforma CKAN tem um conjunto de valores a preencher, estando representado na forma de chave-valor. Por outro lado, DSpace tem um número arbitrário de esquemas de metadados para classificação dos dados. Por último, o Dendro permite o preenchimento de um número arbitrário de metadados, como a flexibilidade quanto ao seu tipo de dados.

A plataforma "Dendro"(github.com/feup-infolab/dendro), desenvolvida no InfoLab da FEUP, tem como principal objetivo suportar a gestão de documentos e dados de investigação de forma colaborativa em preparação para o seu depósito num repositório de dados.

O DendroShare, que é o Dendro com a implementação de uma extensão, tem como objetivo tornar o Dendro mais FAIR. Torna possível o depósito e catalogação de conjuntos de dados de investigação na própria plataforma, deixando de ser obrigatória a existência de um repositório para os conjuntos de dados acabados.

A eficiência da plataforma Dendro vai ser testada em comparação com a plataforma CKAN e Dataverse. Um conjunto de utilizadores vão realizar um conjunto de tarefas, que abrangem os tópicos de descrição, armazenamento, pesquisa de conjuntos de dados e a produção da citação dos conjuntos de dados.

Com estas tarefas, vai ser possível verificar a taxa de sucesso das tarefas realizadas por cada utilizadores, o tempo médio gasto por cada utilizador para a realização de cada tarefa, a taxa de

erros cometidos por cada utilizador ao realizar as tarefas e por último, uma medida do quanto um utilizador ficou satisfeito com a utilização de cada plataforma. Prevendo-se assim melhores resultados com a utilização da extensão do DendroShare.

Com este módulo de gestão de dados acabados será mais fácil disponibilizar dados base dos artigos científicos a longo prazo e de forma interoperável, contribuindo deste modo para uma ciência mais transparente e eficiente.

Palavras chave: Ciência aberta, Citação de dados, Partilha de dados, Repositórios de dados, Dados abertos, Gestão de dados de investigação, Metadados

Abstract

The increasing computational power at the disposal of science has made the management of research data one of the current issues. Linking scientific articles to their base data makes it possible to validate and clarify their content, facilitate the re-use of data by third parties and assign due credit to data creators. Despite these concerns, not all research data sets are identified and stored, which puts many research data sets at risk each year.

The concept of Open Science is the practice of science in such a way that it enables the collaboration and contribution of society in the reproduction of research data, reports among other research processes, as well as its free access to redistribution and use for further research, while that Open Data comprises the availability of data in open access, which allows to boost the development of new ideas and products.

With the availability of data, FAIR principles emerged with the objective of facilitating the research process, accessibility, interoperability and re-use of different data, both by humans and by machines. It facilitated advancement in science, reduced costs in reproducing duplicate results, and improved time utilization. At the same time, the association of FAIR data with publications encourages the reproducibility of the results obtained as well as the verification of the results.

The various research data storage repositories address the subject of filling in the metadata in a different way: the CKAN platform has a set of values to fill, being represented in the key-value form, on the other hand, DSpace has an arbitrary number of metadata schemes for data classification. Finally, Dendro allows you to fill in an arbitrary number of metadata, such as flexibility in your data type.

The "Dendro" platform (github.com/feup-infolab/dendro), developed in FEUP's InfoLab, has as main objective to support the management of documents and research data in a collaborative way, in preparation for its deposit in a data repository .

DendroShare, which is the Dendro with the implementation of an extension, aims to make the Dendro more FAIR. Make it possible to store and catalog research data on the platform itself. The existence of a repository for the finished data sets is no longer mandatory.

The efficiency of the Dendro platform will be tested in comparison to the CKAN platform and Dataverse. A set of users will perform a set of tasks, covering the topics of description, storage, search for datasets, and the production of citation of datasets.

In these tasks, it will be possible to verify the success rate of the tasks performed by each user, the average time spent by each user to perform each task, the rate of errors committed by each user when performing the tasks and, finally, a measure of how much the user was satisfied with the use of each platform. Predicting better results with the use of DendroShare extension.

With this finished data management module, it will be easier to provide base data for scientific articles in the long term and in an interoperable way. Thus contributing to a more transparent and

efficient science.

Keywords: Open Science, Data Citation, Data Sharing, Data Repositories, Open Data, Research Data Management, Metadata

Agradecimentos

A realização da presente dissertação de mestrado, sem a ajuda e contribuição de algumas pessoas não teria sido possível de tornar realidade, aos quais expresso o meu sincero obrigado.

Ao meu orientador, o professor João Silva, pela disponibilidade, colaboração e paciência demonstrada na resolução dos problemas que iam surgindo ao longo do trabalho.

Aos meus colegas de laboratório e amigos, pelos momentos de divertimento, ajuda e motivação que me ofereceram durante a realização do projeto.

À minha namorada Mariana Amorim, pela disponibilidade para a tradução do resumo da dissertação para inglês e por todo o apoio que me tem transmitido ao longo deste tempo.

Por último, um obrigado especial aos meus pais e irmão por todo o apoio, incentivo e paciência demonstrada para me ajudarem a ultrapassar todos os obstáculos que foram surgindo ao longo destes meses de trabalho. Um muito obrigado!

Lázaro Costa

Financiamento

Este trabalho é financiado por Fundos FEDER através do Programa Operacional Competitividade e Internacionalização - COMPETE 2020 e por Fundos Nacionais através da FCT - Fundação para a Ciência e a Tecnologia no âmbito do projeto POCI-01-0145-FEDER-016736.

Conteúdo

Acrónimos	xiii
1 Introdução	1
1.1 Contexto	1
1.2 Motivação	1
1.3 Objetivos	2
1.4 Estrutura do Documento	3
2 Ciência Aberta	5
2.1 Introdução	5
2.2 Objetivos	5
2.2.1 Atribuição de crédito aos autores dos dados	6
2.2.2 Reprodutibilidade de resultados	7
2.2.3 Redução de esforços duplicados	7
2.3 Recomendações	8
2.3.1 Publicações Open Access	9
2.3.2 Princípios dos dados FAIR	10
2.3.3 Curation Lifecycle Model do DCC	11
2.4 Iniciativas e projetos	11
2.4.1 FOSTER - Open Science	12
2.4.2 GO FAIR	12
2.4.3 EUDAT	13
2.5 Conclusões	13
3 Partilha e citação	15
3.1 Citação de publicações	15
3.1.1 Normas e estilos de citação de documentos	15
3.1.2 Identificador persistente PID e as suas funcionalidades	17
3.2 Citação de conjuntos de dados	18
3.2.1 A diversidade de práticas	18
3.2.2 Recomendações	19
3.3 Partilha de conjuntos de dados	20
3.3.1 Stakeholders	20
3.3.2 Fatores motivadores e desmotivadores	21
3.4 Plataformas de partilha e citação de dados	22
3.4.1 Comparação de funcionalidades	22
3.5 Conclusões	23

4	Partilha e citação de dados no DendroShare	25
4.1	Levantamento de requisitos	25
4.1.1	Casos de uso	25
4.1.2	Mockups de interfaces	25
4.2	Descrição técnica	29
4.2.1	Modelo de dados	29
4.2.2	Detalhes de implementação	33
4.3	Demonstração da utilização do sistema	35
5	Experiência de avaliação	41
5.1	Características dos utilizadores avaliados	41
5.2	Questionário pré-experiência	41
5.3	Tarefas de avaliação	43
5.4	Métricas de avaliação	47
5.5	Análise de resultados	47
5.6	Questionário pós-experiência	50
5.7	Conclusões do desempenho das 3 plataformas e o grau de satisfação dos utilizadores	65
6	Conclusões e trabalho futuro	67
6.1	Publicações resultantes deste trabalho	67
6.2	Balanço do trabalho realizado	67
6.2.1	Conclusões do trabalho realizado	68
6.3	Trabalho futuro	70
	Referências	71
A	Guião da experiência	77
A.1	Descrição da experiência	77
A.2	Questionário pré-experiência	77
A.3	Tarefas de Avaliação	77
A.4	Questionário pós-experiência	83
B	Questionários	85
B.1	Questionário de pré-experiência	85
B.2	Questionário de pós-experiência	88
C	Formulário de Avaliação	95

Lista de Figuras

1.1	Árvore de objetivos	3
2.1	Taxonomia da ciência aberta [38]	9
2.2	Ciclo de vida do tratamento de dados de investigação [19]	11
4.1	Casos de uso	26
4.2	Pesquisa preliminar do Dendro	26
4.3	Interface de pesquisa de depósitos no DendroShare	27
4.4	Resultado da pesquisa em lista no DendroShare	28
4.5	Visualização da informação de um conjunto de dados no DendroShare	29
4.6	Fluxo de navegação entre as páginas da plataforma DendroShare	30
4.7	Modelo de dados do DendroShare	31
4.8	Parametrização do ecrã de pesquisa de conjuntos de dados do DendroShare	34
4.9	Fluxo de navegação entre as funcionalidades da plataforma DendroShare	35
4.10	Página de um projeto, com os ficheiros à esquerda e metadados no centro	36
4.11	Interface da criação de um depósito no DendroShare	37
4.12	Listagem de todos os depósitos da plataforma	38
4.13	Pedido de acesso ao depósito	39
4.14	Notificação recebida pelo criador para pedido de visualização de um depósito	39
4.15	Gestão dos acessos dos utilizadores ao depósito	40
4.16	Página do depósito	40
5.1	Resultados da pergunta QI1	43
5.2	Resultados da pergunta QI2	44
5.3	Resultados da pergunta QI3	44
5.4	Resultados da pergunta QI4	45
5.5	Resultados da pergunta QI5	45
5.6	Resultados da pergunta QI6	46
5.7	Resultados da pergunta QI7	46
5.8	Mediana dos tempos de realização das tarefas de avaliação	48
5.9	Média dos tempos de realização das tarefas de avaliação	49
5.10	Desvio padrão dos tempos de realização das tarefas de avaliação	49
5.11	Número de ajudas prestadas durante as tarefas de avaliação	51
5.12	Número de erros cometidos pelos utilizadores durante as tarefas de avaliação	51
5.13	Resultados da pergunta QF1 na plataforma DendroShare	53
5.14	Resultados da pergunta QF2 na plataforma DendroShare	53
5.15	Resultados da pergunta QF3 na plataforma DendroShare	54
5.16	Resultados da pergunta QF4 na plataforma DendroShare	54

5.17	Resultados da pergunta QF5 na plataforma DendroShare	55
5.18	Resultados da pergunta QF6 na plataforma DendroShare	55
5.19	Resultados da pergunta QF7 na plataforma DendroShare	56
5.20	Resultados da pergunta QF1 na plataforma Dataverse	57
5.21	Resultados da pergunta QF2 na plataforma Dataverse	57
5.22	Resultados da pergunta QF3 na plataforma Dataverse	58
5.23	Resultados da pergunta QF4 na plataforma Dataverse	59
5.24	Resultados da pergunta QF5 na plataforma Dataverse	59
5.25	Resultados da pergunta QF6 na plataforma Dataverse	60
5.26	Resultados da pergunta QF7 na plataforma Dataverse	60
5.27	Resultados da pergunta QF1 na plataforma CKAN	61
5.28	Resultados da pergunta QF2 na plataforma CKAN	62
5.29	Resultados da pergunta QF3 na plataforma CKAN	62
5.30	Resultados da pergunta QF4 na plataforma CKAN	63
5.31	Resultados da pergunta QF5 na plataforma CKAN	63
5.32	Resultados da pergunta QF6 na plataforma CKAN	64
5.33	Resultados da pergunta QF7 na plataforma CKAN	64
6.1	Diagrama de Gantt	68
6.2	Balanço dos objetivos alcançados	69

Lista de Tabelas

5.1	Caracterização dos utilizadores avaliados	42
5.2	Perguntas do questionário pré-experiência	42
5.3	Tarefas realizadas pelos utilizadores	47
5.4	Resultados da conclusão da TAREFA6 do CKAN pelos utilizadores	50
5.5	Perguntas do questionário de pós-experiência	52
5.6	Respostas à pergunta QF8 na plataforma DendroShare	56
5.7	Respostas à pergunta QF8 na plataforma Dataverse	59
5.8	Respostas à pergunta QF8 na plataforma CKAN	62
5.9	Desempenho das plataformas	65

Acrónimos

APC Article Processing Charges. 9

CDI Collaborative Data Infrastructure. 13

CNRI Corporation for National Research Initiatives. 17

DCC Digital Curation Centre. 11

DOI Digital Object Identifiers. 17

FAIR Findable, Accessible, Interoperable, and Reusable. 2, 10–13

IDF International DOI Foundation. 17

ORR Open Reproducible Research. 8

OS Open Science. 5, 8

PID Persistent Identifiers. 2, 17, 28

PLOS Public Library of Science. 7, 22

RDA Research Data Alliance. 19, 20

Capítulo 1

Introdução

1.1 Contexto

A gestão de dados de investigação e a evolução do processo científico é um dos temas da atualidade. Como tal, ao anexar aos artigos científicos os seus dados base, para além de ser possível a validação e clarificação do seu conteúdo, também facilita o processo de reutilização em posteriores investigações.

No entanto, nem todos os conjuntos de dados de investigação são identificados e armazenados. Desta forma, podem perder-se muitos e valiosos conjuntos de dados de investigação todos os anos.

Daí surge da plataforma Dendro¹, desenvolvida na FEUP, que tem como principal objetivo a gestão de documentos e dados de investigação, entre outros materiais digitais. Permite aos seus utilizadores depositar ficheiros de qualquer tipo, funcionando como uma “Dropbox privada”.

Após os dados serem depositados no Dendro e devidamente descritos com metadados adequados, é necessário exportar o conjunto de dados e metadados para um ambiente de partilha de dados acabados. O Dendro suporta a integração com várias plataformas de armazenamento de dados existentes (CKAN, ePrints, DSpace, entre outros) e com vista ao alargamento destes horizontes, tem em conta a preservação a longo prazo de dados disciplinares.

Como objetivo, pretende-se implementar uma extensão para o depósito e catalogação de conjuntos de dados de investigação na plataforma Dendro. Deste modo, deixa de ser necessário empacotar os dados e partilhá-los noutras plataformas.

Com esta nova funcionalidade de armazenamento e gestão de dados de investigação, passará a ser possível partilhar conjuntos de dados de forma citável no Dendro.

1.2 Motivação

O crescente poder computacional ao dispor da ciência tem vindo a tornar a gestão dos dados de investigação um dos temas da atualidade.

¹<https://github.com/feup-infolab/dendro>

A ligação dos artigos científicos aos seus dados base torna possível a validação e clarificação do seu conteúdo, facilita a reutilização dos dados por terceiros e permite a atribuição de crédito aos criadores dos dados.

Apesar destas preocupações, nem todos os conjuntos de dados de investigação são identificados e armazenados, o que todos os anos coloca em risco muitos conjuntos de dados de investigação [26]. Atualmente, a UPorto juntamente com o INESC TEC estão em processo de realização de um estudo de adoção de uma solução para a gestão de dados de investigação, que inclui o Dendro.

Após uma análise das limitações deste workflow, sentiu-se a necessidade da implementação de uma extensão para o depósito e partilha de dados acabados.

1.3 Objetivos

O desenvolvimento deste projeto tem um conjunto de objetivos representados estes sob a forma de árvore Fig 1.1 e cada objetivo identificado por um identificador. A árvore de objetivos apresenta objetivos alcançados em trabalhos anteriores, representado pelo objetivo 1.1, especifica o principal objetivo deste trabalho, representado pelo objetivo 1.2 e representado pelo identificador 1.3 o objetivo que será realizado por outro aluno do mestrado que colabora no mesmo projeto.

A plataforma Dendro tem como principal objetivo melhorar a reutilização de dados de investigação. Atualmente, o Dendro permite armazenar, descrever e depositar conjuntos de dados de investigação (objetivo 1.1).

Todo o desenvolvimento deste projeto está focado num único objetivo, permitir a partilha e citação de dados de investigação na plataforma Dendro (objetivo 1.2).

Deste modo, é necessário atribuir Persistent Identifiers (PID) a conjuntos de dados para permitir a citação dos mesmos (objetivo 1.2.1), produzir registo de metadados Findable, Accessible, Interoperable, and Reusable (FAIR) (objetivo 1.2.2), gerar *snippets* de citação para conjuntos de dados de investigação (objetivo 1.2.3) e recuperar conjuntos de dados depositados na plataforma (objetivo 1.2.4).

A plataforma Dendro pretende ser integrada com uma plataforma de geração de PID (objetivo 1.2.1.1), sendo esta integração essencial para a obtenção de forma automática do PID do conjunto de dados depositados no Dendro e concluir o objetivo 1.2.1.

A fim de concluir o objetivo 1.2.2, pretende-se selecionar os descritores dos metadados essenciais para representar os conjuntos de dados (objetivo 1.2.2.1) e assegurar a qualidade do preenchimento desses mesmos descritores (objetivo 1.2.2.2).

Para a geração de *snippets* de citação dos conjunto de dados de investigação (objetivo 1.2.3), é necessário selecionar um padrão para a citação dos dados (objetivo 1.2.3.1).

Por último, para tornar possível o processo de recuperação de dados depositados na plataforma (objetivo 1.2.4), é necessário a geração de um índice e uma interface de pesquisa sobre os metadados associados ao conjunto de dados depositados (objetivo 1.2.4.1).

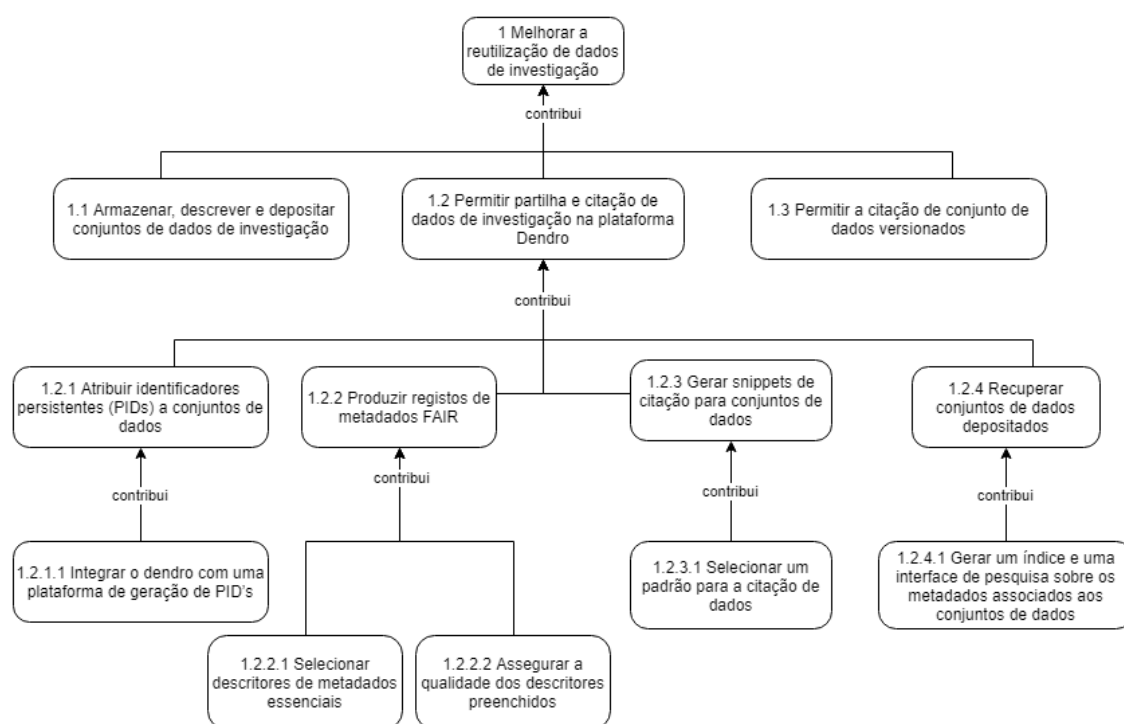


Figura 1.1: Árvore de objetivos

Para trabalho a realizar por outro aluno do mestrado que colabora no mesmo projeto, espera-se possibilitar a citação de conjunto de dados de investigação versionados, como referido no objetivo 1.3.

1.4 Estrutura do Documento

Este documento está estruturado em seis capítulos. O presente e primeiro capítulo introduz o tema trabalhado na dissertação, definindo o contexto, a motivação e os objetivos deste trabalho. O Capítulo 2 apresenta uma parte da pesquisa realizada do estado da arte, que inclui a introdução à Ciência Aberta, os objetivos, as recomendações e as iniciativas e projetos existentes nesta área. No Capítulo 3 a outra parte do estado da arte onde é apresentado as normas de citação de publicações, a diversidade de pratica de citação de conjuntos de dados, os fatores motivadores e desmotivadores de partilha de conjuntos de dados assim como os stakeholders desta partilha dos conjuntos de dados. No Capítulo 4 é apresentada a plataforma Dendro, realizado o levantamento dos requisitos para a solução dos problema, realizada uma descrição técnica do modelo de dados e de detalhes de implementação e finalmente realizada uma demonstração da utilização do sistema. No Capítulo 5 é apresentado as características dos utilizadores avaliados, os questionários de pré e pós experiência, as tarefas realizadas pelos utilizadores, as métricas de avaliação da extensão realizada, os resultados das experiências realizadas e por último uma comparação do desempenho das plataformas utilizadas na experiência de avaliação. Finalmente, no Capítulo 6 são apresentadas as contribuições resultantes deste trabalho, as conclusões e o trabalho futuro.

Capítulo 2

Ciência Aberta

Neste capítulo será introduzido o conceito de Ciência aberta e a importância desta no avanço da ciência.

2.1 Introdução

O conceito de Ciência Aberta, ou Open Science (OS), é a prática da ciência de forma a tornar possível a colaboração e contribuição da sociedade na reprodução de dados de investigação, na elaboração de artigos de laboratório, e na elaboração de relatórios de pesquisa, entre outras contribuições e colaborações em dados de investigação. Permitindo o livre acesso, a redistribuição e reutilização para posteriores pesquisas [38, 42].

Um conceito próximo é o de Dados Abertos ou *Open Data*, que compreende a disponibilização dos dados em acesso aberto. Permite abordar e obter respostas aos desafios humanos, sociais, agrícolas, na nutrição e mudanças climáticas [22], através do desenvolvimento de novas ideias e produtos [26].

A disponibilização dos dados e das publicações científicas em acesso aberto à comunidade, permite que estes conteúdos sejam redistribuídos e reutilizados, desde que seja realizada a devida citação de atribuição de mérito ao criador dos mesmos.

2.2 Objetivos

Atualmente, ainda existem vários obstáculos nesta área, impedindo a plena utilização dos seus benefícios por toda a comunidade.

O movimento da ciência aberta tem sido um facilitador para a participação pública na ciência. Contribui para uma ciência melhor e mais eficiente, possibilitando o crescimento económico das pequenas e médias empresas que pretendam capitalizar os resultados das investigações e, por último, aumentando a transparência dos resultados obtidos nas investigações científicas [26].

Inúmeros relatórios, artigos científicos e declarações políticas foram realizadas com o intuito de apoiar a necessidade de disponibilização e abertura dos dados (bem como outros resultados digitais das investigações) para reutilização como uma prática comum da sociedade [22].

Apesar de todo o trabalho de investigação realizado pelos investigadores, nem todas as partes do trabalho ficam disponíveis para a comunidade, apesar de em muitos casos todas as partes serem de valor para a comunidade. Alguns investigadores desejam partilhar sempre os seus dados, outros desejam não partilhar dados. Alguns investigadores preenchem os metadados necessários para permitir a citação dos dados, outros simplesmente não preenchem [9, 13]. A existência destas várias perspetivas na sociedade contribui para a criação de obstáculos ao eficaz envolvimento dos investigadores no mundo da investigação e na atribuição de crédito ao criador desses mesmos dados [13].

Pretende-se, portanto, uma mudança na cultura dos investigadores, que ao preencherem os metadados dos dados de investigação, passam a permitir a identificação dos mesmos. Ao mesmo tempo, pretende-se aumentar o envolvimento dos investigadores no acesso aberto aos dados através do surgimento de incentivos sistemáticos [26, 2]. Uma forma de aumentar esta disponibilização de dados em acessos aberto é a redução de barreiras ao depósito de dados, como por exemplo, redução do esforço por parte do investigador no momento da publicação de dados. Desta forma, pretende-se que todos os metadados dos dados de investigação não sejam preenchidos no momento da publicação, mas sim, sejam preenchido no momento da recolha desses mesmos dados [15, 2].

Por último, pretende-se um aumento da contribuição social a nível global na adoção da tradição ética da ciência aberta através de atividades, apoiadas e promovidas pelas entidades financiadoras e pela sociedade de investigação, com o intuito de melhorar o processo de investigação e aumentar a participação pública nas investigações [26].

2.2.1 Atribuição de crédito aos autores dos dados

Os reduzidos incentivos para o envolvimento dos investigadores no acesso aberto dos dados, e a reduzida atribuição de reconhecimento do esforço realizado na reprodução dos dados, é um dos principais obstáculos nesta área [26].

Muitos investigadores não adotam a prática da ciência aberta nos dados de investigação, estando este facto relacionado com a falta de atribuição de reconhecimento ao trabalho realizado, ao verem os seus dados utilizados sem citação. Para tal, é necessário os investigadores preencherem os metadados dos dados, o que passa a facilitar a identificação dos mesmos, coisa que não acontece na maioria dos dados depositados nos repositórios [18].

Para permitir a atribuição de reconhecimento aos autores dos dados, é necessária a utilização da citação daquele conjunto de dados por parte de quem os reutiliza. No entanto, nem todos os repositórios o têm presente. Apesar da existência de padrões de citação para todo o tipo de documento e publicação, ainda não existe um padrão para a citação de dados de investigação, e com isto os próprios investigadores têm dificuldade na construção da citação do seu conjunto de dados, necessária para possibilitar o reconhecimento do trabalho [37, 2].

Os investigadores ao tornarem as descrições dos metadados ricos, facilitam a citação dos dados e aumentam a taxa de citação desses mesmo dados, tendo-se chegado a estas conclusões em estudos anteriores [37].

As novas tendências das entidades financiadores focam-se na criação de iniciativas de encorajamento para os investigadores disponibilizarem os seus dados, inclusive a Public Library of Science (PLOS), exige o envio dos dados biomédicos de forma detalhada como condição para publicação do artigo [36, 41]. Da mesma forma, as próprias entidades financiadoras exigem publicações como condição para financiamento dos projetos e iniciativas, deste modo, ao ser incentivada a investigação, é promovido o desenvolvimento e o reconhecimento das instituições [26, 44].

2.2.2 Reprodutibilidade de resultados

O processo científico defende a disseminação e reprodutibilidade da investigação realizada. Durante séculos, artigos e livros científicos têm sido os principais mecanismos de disseminação dos resultados científicos. No entanto, a validação dos resultados obtidos baseia-se no acesso a que o leitor tem aos materiais necessários para validação destes mesmos resultados [10].

O aumento do uso de software na ciência dificulta a reprodutibilidade dos resultados, isto leva à não publicação do código fonte nem dos dados necessários para a reprodução dos resultados por parte dos investigadores [10].

Daí, os resultados e as conclusões obtidas nas publicação não poderem ser validadas pelos leitores, pois é difícil para estes recriar o ambiente experimental dos autores [10].

Com isto, a reprodutibilidade está a tornar-se uma preocupação primária em muitas áreas da ciência. Em particular, na ciência da computação, com o objetivo de tornar as contribuições científicas mais acessíveis e reproduzíveis pela comunidade científica e do público em geral e com isto, aumentar a fiabilidade dos resultados obtidos nas publicações [17, 30].

Ao associar os dados de investigação às publicações, permite-se a reprodutibilidade da experiência, reduzindo os erros e a manipulação dos resultados obtidos. Estas praticas ajudam a tornar a ciência mais rigorosa e com mais qualidade de investigação [37].

2.2.3 Redução de esforços duplicados

Os custo de produção de dados de investigação dependem das experiências ou da observação a realizar. Que envolve um grande valor monetário e na maioria das vezes, não se encontra ao alcance dos centros de investigação [37].

Como a disponibilização dos dados reproduzidos nas investigações em acesso aberto, torna-se possível a reutilização dos dados por parte dos laboratórios de investigação para futuras investigações [9].

Por outro lado, ao reutilizar os resultados ou conjuntos de dados anteriormente realizados por outros investigadores, proporciona uma base de comparação, ou “baseline” com o qual os investigadores podem comparar os seus resultados [29, 44].

As entidades financiadoras focam-se na criação de iniciativas de encorajamento para os investigadores disponibilizarem os seus dados das publicações, algumas delas, só fornecem financiamento com a disponibilização dos dados em acesso aberto [26].

Portanto, a disponibilização e partilha de dados de investigação permite reduzir os custos da produção de dados duplicados. Permite o impulsionamento a novas investigações a utilizadores dos dados base das investigação já realizadas nessa área, e aumenta o aproveitamento do tempo, bem como dos fundos de investimento, facilitando o avanço da ciência [36].

2.3 Recomendações

Como a ciência aberta é uma área relativamente nova, foram criados conceitos recorrendo a uma taxonomia.

A taxonomia Fig 2.1 foi criada no projeto FOSTER¹ baseando-se em dois objetivos. Primeiro, oferecer uma representação detalhada dos conceitos e dos componentes da ciência aberta para aqueles que não estavam familiarizados com o conceito. Segundo, fornecer uma terminologia estruturada e consistente, reduzindo os conflitos descritivos atuais na área da ciência aberta e promover o desenvolvimento através dessas práticas de consistência.

Deste modo, o objetivo do projeto FOSTER não era apenas criar uma taxonomia para classificar os campos da sua área, mas aproveitar a capacidade organizacional das taxonomias para sistematizar os recursos educacionais hospedados no portal, o que também facilitava o processo de revisão dos materiais no portal [38].

Portanto, o primeiro nível da taxonomia da ciência aberta divide-se em nove termos: *Open Access*, *Open Data*, *Open Reproducible Research (ORR)*, *Open Science Definition*, *Open Science Evaluation*, *Open Science Guidelines*, *Open Science Policies*, *Open Science Projects* e *Open Science Tools*.

Embora a taxonomia esteja dividida em nove termos, vai ser dado especial destaque aos termos *Open Access*, *Open Data* e ORR, visto serem as áreas em que o tema deste trabalho está focado.

O conceito de Dados Abertos ou Open Data, está inserido no tópico da ciência aberta OS e refere-se à disponibilização dos dados em acesso aberto. Quer dos resultados de investigação, como que da revisão dos pares recolhidos durante um projeto de investigação. Permite a redistribuição e reutilização após serem realizados os corretos processos de citação [38].

O conceito de acesso aberto ou *Open Access*, está inserido no tópico da ciência aberta OS e refere-se à disponibilização para leitura dos resultados de investigação como da revisão dos pares, podendo existir algumas restrições de direitos de copyright e licenciamento ou, simplesmente, a inexistência destas restrições [38].

O conceito da pesquisa aberta e reproduzível ou ORR, está inserido no tópico da ciência aberta OS e refere-se ao ato de praticar a investigação numa dada área, possibilitando a reprodutibilidade dos resultado apenas com os resultados da pesquisa previamente realizada por parte da entidade responsável pela publicação dos resultados [38].

¹<https://www.fosteropenscience.eu/>

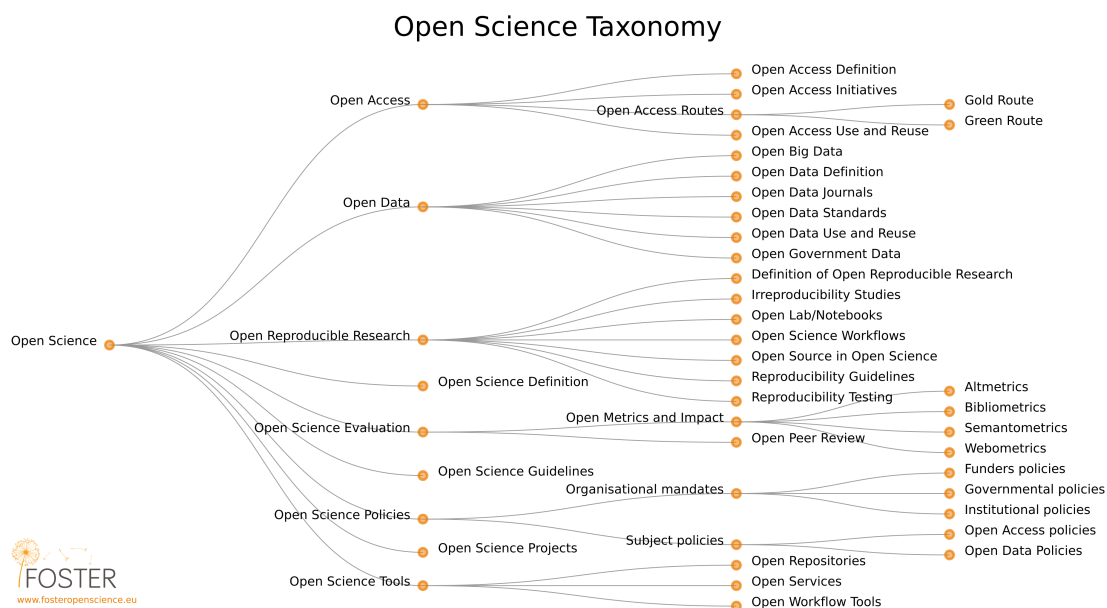


Figura 2.1: Taxonomia da ciência aberta [38]

O facto de a taxonomia incluir uma variedade de termos, demonstra uma tendência universal para a mudança do comportamento dos investigadores em relação à adoção das práticas de disponibilização de conteúdo em acesso aberto, que se relacionam com todo o ciclo de vida da pesquisa [38].

2.3.1 Publicações Open Access

O movimento *Open Access* ajudou a comunidade de investigação na divulgação de publicações a um público mais abrangente e, ao mesmo tempo, no levantamento de discussões sobre os lucros e custos envolvidos nas publicações científicas [23].

O papel das editoras que armazenam publicações das investigações, provavelmente mudará à medida que o acesso aberto dos dados se tornar um tema mais recorrente. Em vez de estas adquirirem material produzido noutros locais para clientes locais os adquirirem, passarão a adquirir todo o tipo de publicações produzidas localmente, procedendo à disseminação desse mesmo material a nível mundial, por meio de canais de publicações de acesso aberto [23].

Gold Open Access

O *Gold Open Access* tem como principal vantagem tornar o acesso livre às publicações, desde o momento da sua primeira publicação.

As licenças de acesso *Gold* concedem abundantes direitos exploratórios, alcançando um maior nível de visibilidade e com isto, um impacto positivo no número de vezes que é divulgado e na frequência com que é citado [12]. Normalmente, este tipo de publicação incorre em taxas para o autor chamadas de Encargos do Processamento do Artigo ou Article Processing Charges (APC).

Green Open Access

O *Green Open Access* não oferece a mesma estrutura de restrições de conteúdo que o *Gold Open Access*. Como resultado, a exploração do conteúdo passa a ser apenas permitida dentro dos limites das restrições das leis de direitos de copyright. Deste modo, o autor deve ter cuidado no contrato que realiza com a editora, a fim de poder reutilizar o artigo de forma a cumprir todas as leis impostas. Não existe uma regra dos padrões de acessibilidade das publicações, assim, diferentes editoras podem impor diferentes períodos de embargo antes da disponibilização dos artigos gratuitamente [12].

No modelo *Green Open Access*, as organizações de investigação criam repositórios institucionais que permitem aos investigadores arquivarem as suas publicações, mesmo que estas tenham sido publicadas em jornais com o sistema *paywall*, não resultando em quaisquer custos para o autor [23].

Tradicional

A transferência de informação de forma livre e eficiente entre a sociedade foi uma previsão do matemático Andrew Odlyzko (1994), que mencionava um novo mundo das publicações eletrônicas, onde estas passariam a ser drasticamente mais baratas que as publicações tradicionais em jornal de papel [32]. Depois de mais de vinte anos, pode-se confirmar que o futuro não foi bem previsto pelo matemático, pois, não houve queda acentuada nos preços de compra das revistas eletrônicas e atualmente são consideradas uma das práticas mais lucrativas e bem sucedidas das editoras [23].

2.3.2 Princípios dos dados FAIR

Nos últimos anos assistimos a uma revolução na tecnologia digital, com a exponencial quantidade de volume de dados criados, armazenados, processados e analisados. Deste modo a facilitar e melhorar a qualidade de pesquisa, acessibilidade, interoperabilidade e reutilização dos dados, surgiu os princípios FAIR [22].

Reconhecendo-se que a adoção dos princípios e práticas para tornar os dados FAIR era do total interesse de algumas áreas científicas, nestes domínios, a partilha e reutilização de dados é promovida e aplaudida, desde que seja realizada a citação por parte de quem os utiliza, para atribuição de crédito ao criador dos dados [22, 24].

Os investigadores reconheceram que não é suficiente publicar na web tanto dados como outros materiais relacionados com a investigação sem permitir a sua reutilização por parte de outros. Entre os requisitos extra incluem-se a presença de informações contextuais e de suporte (metadados) para permitir a descoberta, acessibilidade, interoperabilidade e reutilização dos dados—para sistematizar estas necessidades surgiram os princípios FAIR [24].

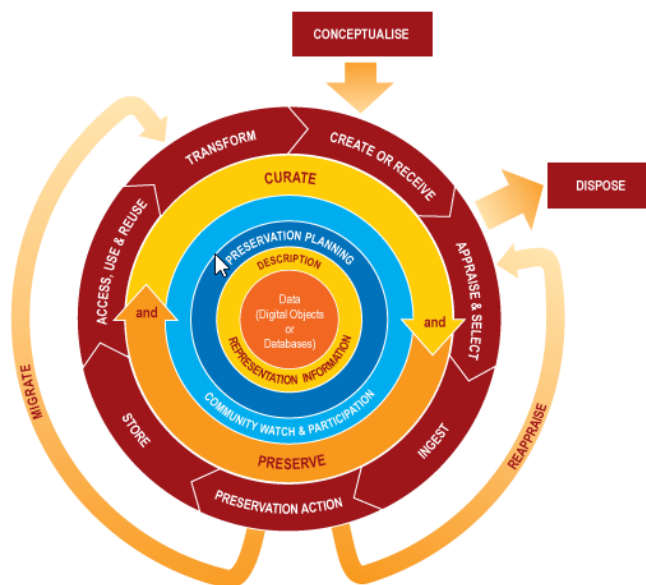


Figura 2.2: Ciclo de vida do tratamento de dados de investigação [19]

2.3.3 Curation Lifecycle Model do DCC

A gestão do ciclo de vida dos materiais digitais é necessária para garantir a organização, documentação e organização dos dados, a fim de permitir a sua partilha e reutilização pela comunidade científica.

O Modelo de Ciclo de Vida de Curadoria do Digital Curation Centre (DCC), representado na Fig 2.2, é uma ferramenta específica para o tratamento de dados, podendo ser usada em conjunto com outros padrões para planeamento de atividade, tratamento e preservação de dados com diferentes níveis de granularidade [20].

A curadoria digital ou “Digital Curation” em inglês, procura preservar e acrescentar valor aos dados de investigação ao longo do seu ciclo de vida, através de uma gestão ativa, que mantém a interoperabilidade dos dados ao longo do tempo [28].

Deste modo, após a conclusão dos financiamentos e/ou conclusão dos projetos, os investigadores devem permitir a exploração e a atualização dos dados através do armazenamento dos mesmos em repositórios digitais confiáveis. Deste modo, os dados são disponibilizados entre toda a comunidade de investigação, reduzindo os esforços da criação de dados de investigação duplicados e aumentando a fiabilidade dos dados ao permitir a reprodutibilidade dos resultados. Por último, facilita a reutilização e citação dos mesmos em futuras investigações [20].

2.4 Iniciativas e projetos

Os incentivos para tornar os dados FAIR são vistos de uma forma recorrente ao nível de investigadores individuais. No entanto, esta necessidade de mudança da cultura da sociedade em

tornar os dados FAIR, tem que acontecer de uma forma mais ampla, incluindo todos os projetos e investigações realizadas [22].

As instituições de investigação de grande ou de pequena escala têm um papel importante na definição e no incentivo de tornar os dados FAIR. Existe portanto a necessidade de um planeamento estratégico nestas infra-estruturas para promover e incentivar os seus investigadores a tornar os dados FAIR [22].

É produzida uma grande quantidade de dados nas instituições de investigação. Por vezes, muitos deles valiosos e com necessidade de os documentar, computar e armazenar para facilitar a pesquisa, localização e reutilização destes mesmos dados. Ao serem realizadas estas ações, são evitadas situações de dados criados para um estudo específico e que não puderam mais ser reutilizados em investigações devido à falta de documentação, ou que simplesmente não foram armazenados e perderam-se [22].

É por todas estas razões que existe a necessidade de iniciativas e missões para incentivar a pesquisa interdisciplinar a nível internacional, fornecendo conhecimento para compreender e adotar as práticas de promoção de mudanças a nível global em tornar os dados FAIR.

2.4.1 FOSTER - Open Science

O portal FOSTER², é uma plataforma *e-learning* que reúne diversos recursos para desenvolver estratégias e habilidades ao utilizar as práticas da Open Science nas investigações realizadas. Com a criação desta plataforma que reúne uma vasta variedade de utilizadores, desde investigadores até administradores da plataforma, surgiu um projeto que tem como principal objetivo tornar a *Open Science* uma norma através da mudança do comportamento dos investigadores europeus [38].

Os investigadores, as instituições de investigação e as entidades financiadores reconheceram que o aumento das aptidões da *Open Science* está relacionado com a realização de investigações e inovações responsáveis. No entanto, há uma concordância crescente entre todas as partes interessadas em que deveria ser aumentada a orientação aos investigadores da UE na aprendizagem de como inicializar uma pesquisa e na utilização das práticas corretas na citação da informação existente [38].

Destas necessidades surgiu o projeto FOSTER PLUS, que tem como objetivo melhorar o material existente, e co-produzir novos conteúdos de aprendizagem através de atividades de formação dirigidas a todas as partes interessadas no Espaço Europeu da investigação, principalmente destinada a jovens cientistas e estudantes. Procura atingir um público amplo e torná-los capazes de incorporar as boas práticas da *Open Science* no seu processo de trabalho [38].

2.4.2 GO FAIR

Um crescente número de investigadores, entidades financiadoras e os responsáveis pela criação das políticas defendem a transição para a *Open Science* e, ao longo do tempo, foram sendo

²<https://www.fosteropenscience.eu/>

lançados projetos pela comunidade destinados a tornar os dados de investigação localizáveis, acessíveis, interoperáveis e reutilizáveis—no entanto estes esforços carecem de sinergia a nível social [1].

Com a necessidade da construção de um ambiente comum, que permitiria a pesquisa e inovação orientada por dados, surgiu o GO FAIR³. O GO FAIR tem como objetivo evitar a concorrência e a fragmentação indevida, ao promover o esforço conjunto dos participantes. Estes princípios assentam em três pilares: GO CHANGE, GO TRAIN e GO BUILD [1]. O GO FAIR disponibiliza um ecossistema aberto à sociedade e instituições de todo o mundo, comprometida com a definição de criação de material e ferramentas com as características dos dados FAIR [1]. É uma iniciativa liderada e auto-governada pela comunidade, que atua em vários países e em diferentes áreas.

2.4.3 EUDAT

A EUDAT⁴ tem como principal objetivo a partilha e a preservação dos dados entre fronteiras e disciplinas, permitindo a administração de dados entre as comunidades de investigação europeias através de uma Collaborative Data Infrastructure (CDI). Sendo esta uma infraestrutura capaz de gerir todos os centros de dados de investigação e todos os repositórios de dados existentes [14, 39].

Os investigadores e profissionais europeus de qualquer área de investigação podem preservar, pesquisar, aceder e processar dados na CDI da EUDAT. A rede utilizada pelos colaboradores para a gestão dos repositórios é confiável, e para além disso, os repositórios têm permanência e persistência nos maiores centros de dados da Europa [14, 39].

A EUDAT oferece um vasto conjunto de serviços de gestão de dados de investigação e de recurso de armazenamento em várias comunidades de investigação. Utilizam uma rede distribuída por 15 países europeus e o seu conteúdo armazenada no conjunto dos melhores computadores da Europa [14, 39].

O principal objetivo da EUDAT é fazer a ligação entre as infraestruturas de investigação e as tecnológicas, através de uma participação ativa da comunidade [14, 39].

2.5 Conclusões

Das iniciativas e projetos existentes, a plataforma Dendro vai servir-se da projeto EUDAT para implementar a partilha e a preservação dos dados entre fronteiras mas irá melhorar a descrição e pesquisa de dados na plataforma.

Com isto pretende-se flexibilizar os descritores metadados a preencher e recolher todos os metadados que de alguma forma contribuam para uma melhor descrição dados. O investigador poderá escolher os metadados que o seu conjunto de dados deve ter e não limitar o investigador ao preenchimento dos metadados impostos. Para além disto, a plataforma também permitirá realizar uma pesquisa facetada sobre todos os metadados associados aos dados da plataforma. Das iniciativas e projetos existentes, nenhum deles permite inserir metadados específicos do domínio aos

³<https://www.go-fair.org/>

⁴<https://eudat.eu/>

dados que serão guardados na plataforma e numa posterior realizar uma pesquisa sobre todos os metadados associados aos dados da plataforma.

Capítulo 3

Partilha e citação

Neste capítulo será introduzido o conceito de normas de partilha, citação de publicações e dados de investigação. Partindo da sua ligação às práticas já existentes de citação de publicações.

3.1 Citação de publicações

A citação e as referências bibliográficas é um tema muito debatido e abordado. Existe normalmente para cada área científica um estilo próprio para a representação das referências das fontes. O principal objetivo da citação é identificar as publicações que foram consultadas e das quais foram recolhidos conteúdos que serviram à elaboração da publicação. Expor a distinção entre o seu trabalho e o trabalho realizado por outros ao especificar de forma clara a localização de tal fonte, para permitir os leitores consultar tal informação se assim o pretenderem [13].

3.1.1 Normas e estilos de citação de documentos

Atualmente existem normas de citação de dados para todos os tipos de documentos. Desde documentos impressos, cartazes e filmes, documentos não publicados até aos documentos eletrónicos. Atualmente destacam-se os seguintes estilos de citação:

- APA
- Harvard Referencing
- Chicago
- MLA
- IEEE

O estilo de citação APA (6ª edição) é representado pelo sobrenome do autor e a data da publicação entre parênteses, sempre que seja referenciada outra fonte na narrativa. Embora seja principalmente utilizado pelas ciências sociais e comportamentais, o formato APA é usado noutras publicações científicas devido à sua eficiência de edição e representação [27]. A título de exemplo, na citação do texto é apresentado (*Apelido, Ano de publicação*) e na citação da lista de referências por *Apelido, Inicial(s). (Ano de publicação). Título. Editora. Localização do recurso.*

O estilo de citação *Harvard Referencing* é usado principalmente por estudantes universitários, para citarem as fontes de informação utilizadas nos seus trabalhos. Este estilo inclui portanto dois tipos de citações: citações usadas diretamente no texto com o intuito de parafrasear uma frase, localizadas estas no corpo do trabalho e contendo um fragmento da citação completa. Por outro lado, existem as listas de referências localizadas no final do trabalho, que mostram a citação completa dessa referência [35].

As listas de referências são criadas para permitir aos leitores localizarem as fontes originais utilizadas no trabalho. Cada citação completa pode conter os seguintes elementos:

- Nome do(s) autor(es)
- Ano da publicação
- Título
- Cidade em que foi publicado
- Editora
- Páginas utilizadas

Por exemplo numa citação em Harvard Referencing, cada entrada na lista de referências tem a estrutura *Apelido, Primeiro Nome. (Ano de publicação). Título. Editora, Página(s)*

O estilo de citação Chicago é o mais utilizado a nível internacional, para representação das fontes de referência. Este estilo pode ser utilizado de duas diferentes formas, uma utilizando o sistema de nota de rodapé e outro usando o sistema Autor-data [27].

As citações em notas de rodapé são frequentemente utilizadas na área das humanidades, enquanto que, o sistema de citação Autor-data é mais utilizado nas área das ciências [27].

Numa citação em Chicago, as entradas na lista de referências têm a estrutura *Primeiro nome último nome. Título. (Lugar de publicação: Editora, Ano de publicação)*

O estilo de citação MLA foi desenvolvido pela Modern Language Association como uma forma consistente de documentar as fontes usadas na escrita académica. É principalmente utilizado nas áreas de humanidades [27].

Por exemplo numa citação em MLA na lista de referências é apresentado *Último Nome, Primeiro Nome. Título, Editora, Data de publicação, Localização do recurso.*

O estilo de citação IEEE¹, é usado principalmente nas publicações técnicas da ciência da computação e engenharia. Uma das características que o distingue dos outros estilos, é o uso de parênteses retos em vez dos parênteses responsáveis pela enumeração das citações no texto e da lista de referências. Além disso, a lista de referências é representada pela ordem numérica que aparece no texto, em vez da ordem alfabética do sobrenome do autor.

Quando é adotado o estilo IEEE, as listas de referências têm entradas com a estrutura *Primeiro nome Último Nome. Título. Lcal de publicação: Editora, Ano de publicação, Localização do recurso.*

A utilização de um dos diferentes estilos de citação depende da área onde está inserido, dos requisitos da universidade ou da publicação para a qual está a ser realizado o trabalho. O crucial

¹<https://iee-dataport.org/>

é apresentar o trabalho de forma consistente e com as atribuições dos créditos aos respectivos autores dos materiais utilizados nos trabalhos. Com isto, demonstra-se a atenção ao detalhe e aumenta a credibilidade do trabalho escrito. Cada estilo de citação contém elementos essenciais, recomendáveis e facultativos para a sua correta representação [27, 35].

3.1.2 Identificador persistente PID e as suas funcionalidades

Para ser possível referenciar com maior grau de segurança e fiabilidade um documento, um ficheiro, uma página web ou outros objetos, surgiram os identificadores de longa duração PID. Naturalmente o conceito de identificador é anterior à Internet—ao longo dos séculos, os escritores procuraram desenvolver um padrão para citação de documentos, de forma a possibilitar aos leitores encontrarem de forma eficiente a fonte que o escritor mencionou.

Com o surgimento da Internet, os identificadores passaram a ter maior importância no mundo da Internet. A persistência de um identificador está relacionada com o comprometimento de um sistema institucional a realizar o serviço que mantenha o recurso identificado pelo identificador persistente. Deste modo, o identificador é persistente enquanto esse serviço se comprometer a guardar esse recurso, podendo ao fim de algum tempo desaparecer a ligação aos dados se o serviço deixar de estar disponível [21].

3.1.2.1 Digital Object Identifiers (DOI)

Um DOI é uma sequência de números, letras e símbolos usados para identificar permanentemente e de forma interoperável um recurso de informação em redes digitais. No entanto, não é o local de armazenamento dos dados [34].

Na maioria das publicações realizadas recentemente, o DOI é impresso no próprio artigo, podendo ser encontrado na página inicial, no cabeçalho ou até mesmo no rodapé [34]. Um DOI é um identificador persistente, que contém as informações atuais desse objeto, estando todo o sistema de gestão de atribuições de DOI's a encargo da International DOI Foundation (IDF). O DOI pode ser usado para gestão de dados comerciais e não comerciais. Este combina o sistema de resolução *Handle* e o Dicionário de Dados *indecs* no seu sistema de desenvolvimento.

Em primeiro lugar, será falado do INDECS, acrónimo de "Interoperability of data in electronic commerce systems". Uma parte do projeto foi financiado pela iniciativa Info 2000 da comissão Europeia e por várias organizações que representam os autores, as editoras de texto e outros setores entre o ano de 1998 e o ano 2000. Desde então, tem sido utilizado em atividades que envolvem metadados [34].

Em segundo lugar, será falado do Sistema *Handle*, em que a entidade responsável é a Corporation for National Research Initiatives (CNRI). Este sistema atribui identificadores persistentes ou identificadores a recursos de informação necessários para localizar, aceder e utilizar os recursos. O Sistema *Handle* não codifica a informação sobre os recursos subjacentes, daí ser apenas veiculado aos metadados do recurso. Deste modo, o identificador não é considerado inválido apesar da mudança dos metadados.

3.2 Citação de conjuntos de dados

A citação de dados de investigação é tão valorizado como documentos e monografias. Devido ao grande crescimento da produção de dados nos dias de hoje, a reutilização de dados de investigação, respeito pelas novas políticas de pesquisa e utilização das ferramentas de gestão de dados está a ganhar maior importância [9]. A citação de dados é uma prática importante no suporte de acesso aos mesmo, partilha, reutilização, pesquisa e reprodutibilidade dos resultados obtidos. O que resulta numa das técnicas largamente reconhecida em muitas áreas de investigação [3].

A crescente produção de dados de investigação e a crescente procura de acesso a grandes dados de coleções faz-se sentir na educação, ciência, política e no comércio. Ao reduzir as barreiras à descoberta e ao uso desses mesmos dados, possibilita a ligação dos dados às publicações científicas, promovendo novas publicações interdisciplinares nas investigações, o que leva à redução dos custos de replicação dos dados e promoção da continuidade de investigações após reutilização dos dados e resultados obtidos [3, 13].

Deste modo, toda a comunidade científica, nomeadamente editores e instituições financiadoras, têm lançado alguns esforços para melhorar a confiabilidade, reprodutibilidade e disponibilidade dos dados e resultados das investigações nas várias áreas da ciência, ao promover a adoção padronizada de uma norma de citação e partilha de dados de investigação [3].

3.2.1 A diversidade de práticas

Devido à ausência de padrões de citação de dados de investigação, existem dificuldades em referenciar e citar dados na comunidade científica. Muitos autores nas suas publicações recorrem a uma maneira tradicional de referenciar os dados diretamente subjacente à publicação. Pode ainda existir uma declaração de acesso aos dados caso não sejam dados abertos. No caso dos dados abertos, estes passam a estar disponíveis a partir do fornecimento do identificador URL do acesso aos dados [7].

A plataforma Virtual Datacenter, desenvolvida em regime *open source*, foi projetada para suportar o acesso e citação dos conjuntos de dados de investigação que armazena. A cada conjunto de dados é atribuído um identificador persistente e uma citação, que é um URL da localização dos dados [3]. No entanto, este método de localização dos conjuntos de dados de investigação não é eficaz quando pretendido localizar subconjuntos dos dados localizados no URL.

Um conjunto de dados de investigação pode ser simplesmente uma tabela, um ficheiro ou até mesmo um conjunto de ficheiros. Sendo cada conjunto de dados de investigação identificado pelo seu URL. Um autor quando tenta citar um subconjunto de dados de investigação encontra dificuldades, devido ao identificador dos dados identificar um conjunto de dados e não um subconjunto específico dos dados.

Para os autores, a solução encontrada foi usar o identificador do conjunto de dados, e no caso de pretender apenas citar um subconjunto dos dados, deveria fornecer informações específicas para o leitor localizar o subconjunto dos dados [7]. No entanto, esta metodologia de localização de conjuntos de dados e subconjuntos de dados de investigação através de identificadores ou URL

não é a mais apropriada para citar dados de investigação, e com isto acarreta alguns riscos à sua utilização:

- No caso de um erro tipográfico no identificador ou no URL não há informações adicionais para localização dos dados de investigação [7];
- Os autores podem fornecer o URL do repositório dos dados, em vez do URL do conjunto específico dos dados [7];
- Não há atribuição de crédito aos criadores dos dados de investigação [7].

Por outro lado, a organização internacional DataCite², que tem como objetivo melhorar a citação de dados ao estabelecer acesso fácil aos dados das investigações, aumentar a contribuição e produção de dados de forma fiável e por último, fornecer suporte ao armazenar os dados permitindo que os resultados sejam verificados e reutilizados em futuras investigações [18].

O DataCite assume-se assim como uma plataforma que armazena e partilha os metadados dos dados de investigação de forma eficiente por toda a sociedade. O esquema de metadados criado pelo grupo de trabalho do DataCite *Metadata Working Group*, apresenta metadados de preenchimento obrigatório. Metadados recomendados que são opcionais, mas altamente recomendados, e metadados opcionais que fornecem informação mais detalhada dos dados [18].

3.2.2 Recomendações

Os princípios de citação de dados de investigação devem-se estender a todas as áreas e a todos os tipos de dados. Deste modo, foram criadas recomendações de citação de dados e não especificações detalhadas de implementação ou de tecnologias e repositórios de dados a usar.

Ao fim de algumas reuniões entre o grupo Research Data Alliance (RDA) e alguns comentários do público, em Fevereiro de 2014, foi finalizada a declaração conjunta dos princípios de citação de dados, que reflete uma ampla convergência de alguns princípios fundamentais [3, 2]:

1. Importância - Os dados devem ser considerados produtos legítimos e citáveis na investigação e desta forma citáveis como as publicações.
2. Crédito e atribuição - A citação de dados deve facilitar a concessão e a atribuição de crédito a todos os colaboradores dos dados.
3. Evidencia - Na literatura, sempre que uma evidencia se baseie em dados, os correspondentes dados devem ser citados.
4. Identificação única - A citação de dados deve incluir um identificador persistente globalmente reconhecido e amplamente usado pela comunidade.
5. Acesso - As citações de dados devem facilitar o acesso aos dados e a todos os seus metadados, tanto às máquinas como a humanos.

²<https://datacite.org/>

6. Persistência - Os identificadores únicos dos dados e os seus metadados devem persistir para além do tempo de vida dos dados que descrevem.
7. Especificidade e verificabilidade - As citações de dados devem facilitar o acesso e a verificação dos dados que apoiam a obtenção dos resultados, fornecendo informações para montar o mesmo ambiente experimental.
8. Interoperabilidade e flexibilidade - Os métodos de citação de dados devem ser flexíveis para acomodar as diferentes práticas entre as comunidade, mas por outro lado, não devem comprometer a interoperabilidade das práticas das citação de dados entre as comunidades.

Como resultado do trabalho do grupo RDA, surgiu um exemplo genérico de citação de dados de investigação:

Autor(es), Ano, Conjunto de dados, Identificador persistente global, Repositório dos dados ou arquivo, Versão ou subconjunto

Pouco tempo depois da oficialização dos princípios, já trinta organizações utilizavam estas recomendações, no entanto, o grupo comprometeu-se com um plano para a disseminação destes princípios por várias organizações e por diferentes áreas da investigação para levar os princípios à escala global.

A criação destes princípios de citação de dados unificados e amplamente difundidos modificará os fluxos das publicações, criará novas tecnologias de citação de dados, definirá novas métricas de impacto e de conhecimento de investigação e fornecerá acesso aos dados de forma persistente e interoperável entre a comunidade de investigação [3].

3.3 Partilha de conjuntos de dados

A gestão de dados de investigação é uma prática comum nos dias de hoje, ao registar e preservar os resultados das investigação ao longo do tempo.

A partilha pública de dados detalhados da investigação, facilita à ciência através da confirmação dos resultados obtidos nas investigação, assim como possibilita a reutilização dos dados em bruto para explorar novas hipóteses e realizar investigações relacionadas [36].

Atualmente, existem várias plataformas geridas por equipas de investigadores que garantem o depósito e partilha dos dados de investigação pela comunidade de investigação [5].

3.3.1 Stakeholders

Com o surgimento da gestão de dados de investigação, surgiram também alguns stakeholders, nomeadamente as entidades financiadoras, as instituições de investigação, as editoras, os repositórios, e claro, os investigadores.

Todas estas entidades estão interessadas no ciclo de vida dos dados de investigação, exigindo que os dados sejam validados, armazenados, descritos e preservados a longo prazo, a fim de possibilitar a descoberta, recuperação e reutilização dos mesmos por outros investigadores [11].

3.3.2 Fatores motivadores e desmotivadores

Dado que a citação de dados procura satisfazer os vários requisitos de cada área de investigação, foram identificadas seis motivações comuns para citação de dados de investigação nas diferentes áreas:

- i. Atribuição de dados: Recolher e gerir dados de investigação por vezes é um trabalho moroso e árduo, sendo muitas vezes realizado por cientistas e investigadores qualificados, tornando-se essencial atribuir crédito pela granularidade de trabalho realizado e não no conjunto de dados como um todo, quer para atribuição de responsabilidades, quer para alcançarem reconhecimento científico [43, 44].
- ii. Conexão de dados: A citação de dados é essencial para ligar as publicações aos dados em que são baseados, para permitir a consulta e a recuperação dos dados durante a leitura do artigo em formato digital [11, 43].
- iii. Descoberta de dados: Citar dados significa ser capaz de identificar, pesquisar, aceder e recuperar conjunto de dados ou um subconjunto de dados, sendo estas operações fundamentais para tornar os dados detetáveis. No entanto, as citações de dados podem atuar como pontos de entrada para fontes de dados ocultas, não sendo por isso indexadas por mecanismos de pesquisa e tornando-se praticamente inacessíveis aos investigadores que não tenham conhecimento para aceder a essas fontes de dados [43].
- iv. Partilha de dados: Os investigadores ao citarem dados de investigação, estão a atribuir crédito aos criadores dos dados, quer investigadores como instituições. Est é um dos fatores decisivos para o aumento da disposição dos investigadores a partilharem os dados de investigação [43].
- v. Impacto dos dados: A citação de dados permite determinar o impacto dos dados através de métricas, como o número de utilizações dos dados. Apesar de o número de utilizações não ser sinónimo de qualidade, é particularmente importante para as entidades financiadoras que investem na criação dos dados. Realçando a necessidade de aumentar o número da citação dos dados [43].
- vi. Reprodutibilidade: A citação de dados tem um grande impacto na reprodutividade da ciência [6], ao ser muito utilizado nas várias áreas científicas, como a astronomia [25], biologia [8], física [17] entre outras. Nos últimos anos, várias revistas autorizadas solicitaram a partilha de dados e o fornecimento de metodologias para a validação das experiências realizadas através da citação de dados, tornando a publicação mais confiável e reutilizável entre a comunidade de investigação [43].

As editoras e os repositórios com disponibilização dos dados em acesso aberto permitem reduzir os erros dos dados e reduzir a manipulação dos dados, assegurando a qualidade e a integridade dos dados ao longo do tempo [5].

As instituições também são motivadas para que os dados dos seus investigadores sejam reconhecidos, partilhados e armazenados. Pois, algumas vezes, são requisitos das instituições de financiamento. Ao valorizar a conformidade dos metadados com os padrões existentes, possibilita a integração dos dados nos repositórios de armazenamento dos dados e consequentemente o aumento da visibilidade da instituição [5, 11].

Para além disso, ao agregar os dados de investigação às publicações científicas, abrem-se portas a outros investigadores para continuarem as investigações anteriormente realizado, e os iniciantes do projeto, receberem crédito através das citações feitas por outros ao seu conjunto de dados criados [15, 33].

Os investigadores concordam com a partilha de dados de investigação, mas têm medo de não receberem crédito da reprodução dos dados e assim perderem financiamento ou oportunidades de publicação, como vêm muitas vezes os seus trabalhos publicados em nome de outros. Isto deve-se à inexistência de sistemas de reconhecimento do trabalho atualmente realizado [29].

Para além disso, para manter os dados documentados e formatados, acarretam-se custos de manutenção, sendo muitas das vezes suportados pelos investigadores [37].

Espera-se assim que os benefícios superem os custos de partilha de dados de investigação, pois estão constantemente a serem criadas iniciativas de encorajamento para os investigadores disponibilizarem os seus dados, inclusive a PLOS, exige o envio dos dados biomédicos de forma detalhada como condição para publicação do artigo [36, 41].

3.4 Plataformas de partilha e citação de dados

Dada a reconhecida preocupação da gestão de dados de investigação para as instituições e investigadores, surgiram várias plataformas capazes de suportar o depósito de dados, destacando-se **Zenodo**, **CKAN**, **DSpace**, **Figshare**, **EUDAT B2SHARE**, **Dataverse** e o **Dendro**. Em seguida vão ser comparadas cada uma delas no que diz respeito às funcionalidades de descrição, partilha e recuperação de dados.

3.4.1 Comparação de funcionalidades

Os utilizadores do Zenodo, após o depósito, podem especificar um período de embargo para o novo recurso. Ao mesmo tempo, a plataforma reserva um link exclusivo e gera um DOI para o depósito, mesmo quando o recurso permanece sob embargo. O Zenodo suporta metadados Dublin Core e MARC e uma pesquisa facetada [4, 22]. Infelizmente, o Zenodo suporta apenas um conjunto fixo de descritores de metadados e não permite especificar metadados específicos do domínio.

O CKAN inclui recursos direcionados à gestão de dados, incluindo versionamento de metadados. Ao contrário do Zenodo, uma instalação padrão do CKAN sem *plugins* pode incluir mais descritores nos registos de metadados, mas estes são essencialmente um conjunto de pares de valores-chave. Isso significa que a validação não existe e, como tal, é difícil usar descritores formalmente definidos a partir de esquemas específicos de domínio. Na experiência particular com o

CKAN, também foram observadas duas questões principais. Em primeiro lugar, as permissões do conjunto de dados são geridas no nível “ Organizacional ”, tornando necessário criar novas organizações para cada grupo de pesquisa e deste modo, evitar que conjuntos de dados privados sejam visíveis para outros grupos de investigadores que trabalham na mesma instituição. Em segundo lugar, muitos recursos essenciais (como embargo, gestão de acesso a utilizadores ou validação de metadados) estão disponíveis apenas através de *plugins*. Como estes são mantidos por voluntários, às vezes ficam desatualizados e incompatíveis com as versões mais recentes do CKAN.

Tal como o Zenodo, o DSpace permite que o criador do depósito especifique um período de embargo ao criar um novo depósito. Por padrão, o DSpace integra-se aos muitos provedores de identificadores (incluindo o DOI) desde a versão 3.0³. Suporta muitos esquemas, permite metadados genéricos e específicos do domínio com uma estrutura formal (ao contrário do CKAN), e inclui recursos de pesquisa facetada [4, 41].

O Dataverse atribui automaticamente um DOI com um prefixo padrão a cada depósito, com uma integração configurável com o DataCite para que os repositórios mantenham automaticamente os seus próprios DOIs. O Dataverse também permite versionamento de versões possui uma pesquisa facetada [24].

O Dendro foi inicialmente projetado sem recursos de depósito de dados, mas sim como uma plataforma colaborativa de dados para auxiliar os investigadores na organização e descrição dos ficheiros de dados com o recurso a descritores recomendados para preenchimento [16]. O modelo de dados é baseado em grafo [40], onde os metadados de recursos são propriedades que podem ser extraídas de múltiplas ontologias, permitindo não apenas metadados genéricos, mas também específicos de domínio.

3.5 Conclusões

A plataforma com a implementação da nova extensão terá de possuir as principais características de cada uma das plataformas analisadas, desde o embargo de depósitos no momento de publicação dos dados, a atribuição de um DOI a um depósito e a geração do snippet de citação de forma automática. Por outro lado, possuirá uma característica não presente em cada uma das plataformas analisadas, pois permitirá inserir metadados específicos do domínio aos dados inseridos na plataforma e realizar uma pesquisa facetada nos depósitos da plataforma pesquisando nos metadados associados aos dados.

Visto a plataforma Dataverse e CKAN serem umas das plataformas mais reconhecidas e utilizadas pelos investigadores para a gestão dos dados de investigação, estas serão utilizadas na experiência de avaliação para avaliar o desempenho da extensão implementada na plataforma Dendro.

³<https://wiki.duraspace.org/display/DSDOC4x/DOI+Digital+Object+Identifier>

Capítulo 4

Partilha e citação de dados no DendroShare

Este capítulo numa primeira fase abordará o levantamento de requisitos da plataforma, incluindo os casos de uso e as mockups de interfaces. Numa segunda fase, abordar-se-á a descrição técnica da plataforma, incluindo o modelo de dados e detalhes de implementação e por último, um pequena demonstração da utilização do sistema pelo utilizador.

4.1 Levantamento de requisitos

Nesta secção foi realizado o levantamento dos novos casos de uso à implementar na plataforma, assim como demonstrados os mockups de interface para as funcionalidades implementadas.

4.1.1 Casos de uso

A nova extensão à plataforma Dendro, denominada *DendroShare*, vai possibilitar ao utilizador realizar novas ações na plataforma, nomeadamente: depositar e partilhar conjuntos de dados a partir dos projetos em que participa no Dendro, pesquisar conjuntos de dados, listar conjunto de dados e por último, citar os conjunto de dados da plataforma. Estes casos de uso encontram-se representados no diagrama da Fig 4.1.

4.1.2 Mockups de interfaces

A plataforma Dendro, na sua página inicial, já apresentava uma interface de pesquisa de projetos (ver Fig 4.2), mas ainda se encontrava numa fase muito preliminar.

De modo a melhorar a interface de pesquisa dos conjuntos de dados acabados no Dendro, serão adicionados novos descritores e melhorados os existentes, com o intuito de tornar a interface de pesquisa mais semelhante ao que se pode ver na Fig. 4.3.

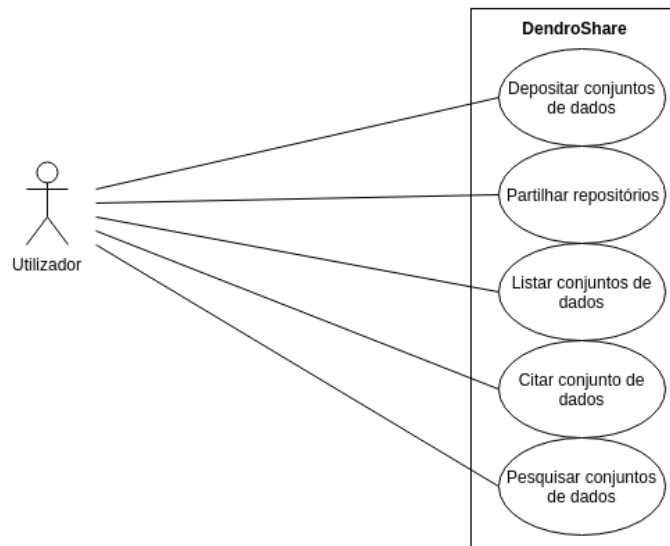


Figura 4.1: Casos de uso

O formulário de pesquisa preliminar do Dendro contém os seguintes campos e opções:

- Filters** (título)
- Username: campo de texto
- Project: campo de texto
- Date Start: campo de texto com máscara mm / dd / yyyy
- Date End: campo de texto com máscara mm / dd / yyyy
- Private Deposits only:
- Platform Used: lista de plataformas com caixas de seleção marcadas:
 - CKAN
 - DSpace
 - EPrints
 - Figshare
 - Zenodo
 - EUDAT B2Share
- Order By: menu suspenso com a opção **DATE** selecionada
- Deposits per page: menu suspenso com a opção **10** selecionada

Figura 4.2: Pesquisa preliminar do Dendro

Filters







Username	<input type="text" value="Username"/>
Project	<input type="text" value="Project"/>
Start date	<input type="text"/> 
End date	<input type="text"/> 
Platform used	<input type="radio"/> Plataform 1 <input type="radio"/> Plataform 2 <input type="radio"/>
Order By	<input type="text"/> 
Deposit type	<input type="radio"/> Private deposits <input type="radio"/> Open deposits <input type="radio"/> Embargoed deposits
Keyword	<input type="text" value="Any of/All of"/>  <input type="text" value="Keyword"/>
Popular Tag	<input type="text" value="Any of/All of"/>  <input type="text" value="Tag"/>
Description text	<input type="text" value="description"/>
persistent PID	<input type="text" value="PID"/>
Deposit per Page	<input type="text"/> 

Figura 4.3: Interface de pesquisa de depósitos no DendroShare

List dataset		
Title of dataset 1	Creator	Description
Title of dataset 2	Creator	Description
Title of dataset 3	Creator	Description
Title of dataset 4	Creator	Description

Figura 4.4: Resultado da pesquisa em lista no DendroShare

Os possíveis descritores utilizados para preenchimento e filtragem dos resultados de pesquisa serão: o username do proprietário do projeto, o nome do projeto, a data a partir da qual se pretende realizar a pesquisa, a data final do intervalo da pesquisa, qual a plataforma usada para o depósito e como se pretende que os resultados dos projetos sejam ordenados, isto é, ordenados por ordem crescente ou decrescente do nome do username ou da data de criação.

Para além destes descritores, também é possível especificar o tipo de depósito a procurar (depósitos privados, com acesso aberto ou com acesso embargado). Outro descritor usado é o de palavras chave, representadas por uma lista de palavras, permitindo encontrar o dataset que contém todas as palavras ou apenas algumas. Da mesma forma é possível pesquisar por *tags* do conjunto de dados, em que uma *tag* representam a área científicas em que o conjunto de dados está inserido. Como não poderia deixar de ser referido, também é possível inserir uma lista de palavras a encontrar na descrição do conjunto de dados. No caso de possuir conhecimento prévio do PID do depósito, o conjunto de dados pode ser recuperado diretamente através desse descritor. Por último, o utilizador pode especificar o número de resultados que pretende obter por página.

O resultado desta pesquisa é uma lista com os nomes dos conjunto de dados, os seus proprietários e uma breve descrição de cada conjunto de dados, como pode ser visto na Fig 4.4.

Por último, quando o utilizador clicar no título de um elemento da lista de resultados da pesquisa, este será redirecionado para as páginas desse conjunto de dados, em que é possível visualizar os vários ficheiros e as informações de todos os descritores desse conjunto de dados, como pode ser visto na Fig 4.5.

Pode ser visualizado o fluxo de navegação entre os vários mockups da plataforma (Fig 4.6) onde estão representadas as ações realizadas para interagir entre as diferentes páginas.

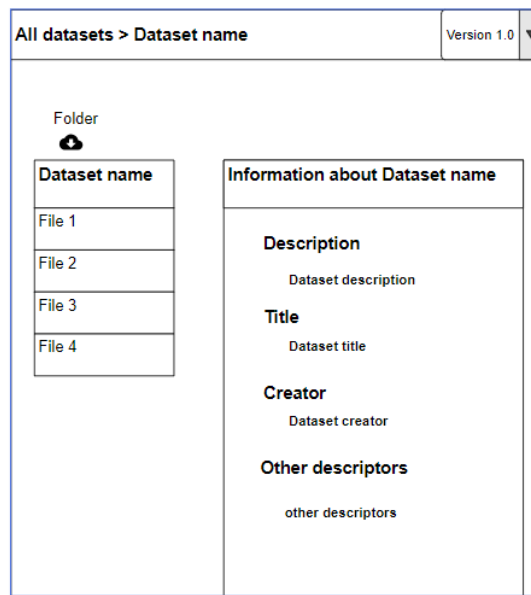


Figura 4.5: Visualização da informação de um conjunto de dados no DendroShare

4.2 Descrição técnica

Nesta secção serão abordados os detalhes técnicos da extensão realizada, desde o modelo de dados até a alguns detalhes de implementação.

4.2.1 Modelo de dados

O modelo de dados desenhado para demonstrar a estrutura lógica da extensão implementada, está assente em 7 classes: *Deposit*, *Project*, *User*, *ConditionsAcceptance*, *Notification*, *Folder* e *File*, representado no diagrama da Fig 4.7. Para facilitar a visibilidade do diagrama, não foi incluído a classe *Resource*. Todas as classes da plataforma herdam da classe *Resource*, em que esta contém entre outros atributos, um atributo Uri. O Uri é uma string que identifica cada recurso de forma única dentro da plataforma.

As classes *Folder* e *File* já faziam parte da solução do Dendro, por este motivo não foram representados os atributos destas classes, nem dado especial detalhe organizacional. Estas classes só foram representadas no diagrama para demonstrar que um Depósito e um Projeto têm como `rootFolder` um elemento da classe *Folder* e que este *Folder*, pode conter vários elementos da classe *File* e vários elementos da classe *Folder*, isto é, o Projeto e o Depósito têm uma estrutura de ficheiros em árvore.

Demonstrar-se-á como está organizada cada classe e o que cada atributo representa, assim como se explicará as relações entre as várias classes.

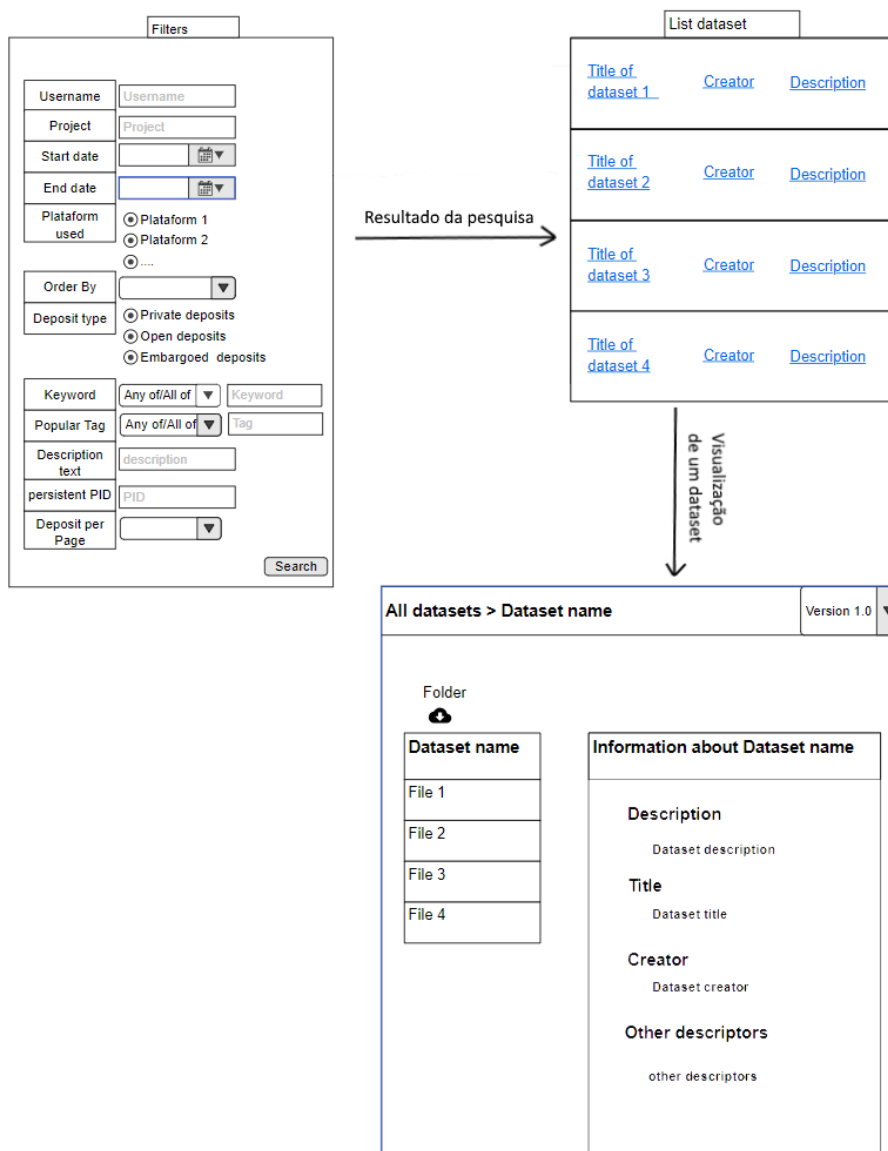


Figura 4.6: Fluxo de navegação entre as páginas da plataforma DendroShare

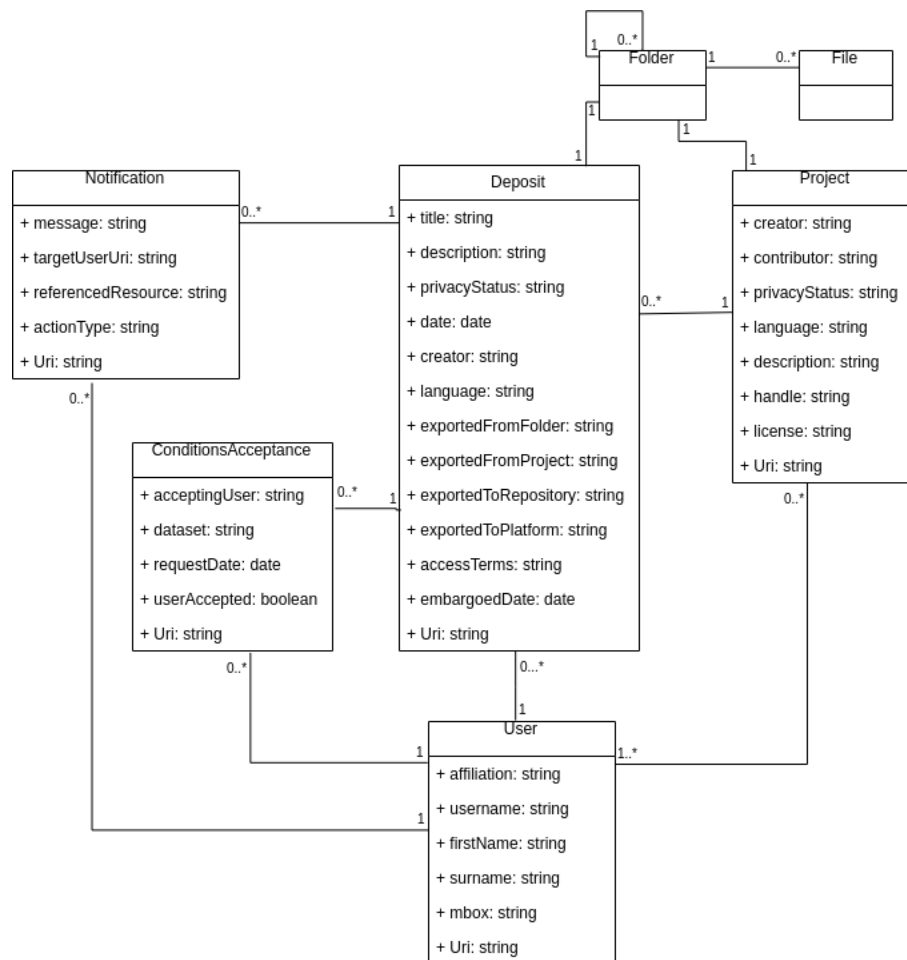


Figura 4.7: Modelo de dados do DendroShare

Classe Deposit

A classe *Deposit*, que é a classe principal desta extensão implementada, guarda o título do depósito, a descrição, a data de criação do depósito, a visibilidade do depósito representado por `privacyStatus`, o Uri do criador do depósito representado pelo atributo `creator`, a língua em que está os dados do depósito representada por `language`, o Uri do projeto de onde o depósito foi exportado representado por `exportedFromProject`, o Uri do `rootFolder` do projeto de onde foi exportado representado por `exportedFromFolder`, o nome da plataforma para onde foi exportado o depósito, visto que pode ser exportado para além da plataforma DendroShare, representado por `exportedToPlatform`, o Uri do repositório para onde vai ser exportado o depósito, no caso de ser para fora do DendroShare, representado por `exportedToRepository`, os termos de acesso inseridos pelo criador do depósito e por último a data de fim do embargo no caso da visibilidade do depósito ser embargado.

Classe Project

A classe *Project* armazena todos os dados e descrições realizadas pelo criador e contribuidores do depósito. No momento da exportação, todos os ficheiros e descrições são exportadas para o depósito, daí alguns atributos coincidirem entre as duas classes. Nesta classe apenas se dará especial detalhe aos atributos `privacyStatus`, `handle` e `license`. O atributo `privacyStatus`, tal como na classe *Deposit*, representa a visibilidade. Este pode ter um valor no projeto, mas no momento da exportação, o utilizador tem a possibilidade de escolher a nova visibilidade daquele conjunto de dados a exportar, a `license` guarda os termos de acesso e utilização dos dados escolhidos pelo criador, e por último, o `handle`, que guarda um identificador único do projeto no DendroShare.

Classe User

A classe *User*, representa o utilizador da plataforma, em que para além do atributo `Uri`, guarda a afiliação que representa a instituição para a qual realiza o trabalho de investigação, o `username`, o `firstName` e `surname`. Por fim, o atributo `mbox` guarda o email do utilizador. Todos os atributos da Classe *User* são do tipo `string`.

Classe ConditionsAcceptance

A classe *ConditionsAcceptance* é responsável por representar uma condição de pedido de visualização de um depósito pelo utilizador. Assim que o utilizador realiza um pedido de visualização, é criada uma nova instância da classe. O atributo `acceptingUser` irá armazenar o Uri do utilizador que deverá aceitar o pedido, neste caso, de um elemento da classe *User*. O atributo `dataset` irá armazenar o depósito a que a condição está associado, neste caso irá guardar o Uri de um elemento da classe *Deposit*, o atributo `requestDate` irá guardar a data do pedido para visualização

do depósito e finalmente, o atributo `userAccepted` representado por um *boolean*, indicará se a condição foi aceite ou não pelo criador do depósito.

Classe Notification

A classe *Notification* representa a notificação que é gerada assim que o utilizador pede acesso de visualização a um depósito, ou assim que o criador aceita um pedido de visualização a um depósito.

Esta classe contém os atributos `referenceResource`, que armazena o Uri do depósito que desencadeou a notificação. O atributo `actionType` representa o tipo de ação a desencadear assim que a notificação é criada, o atributo `message` armazena a mensagem a ser mostrada ao utilizador, e por último o `targetUserUri` armazena o Uri do Utilizador a quem vai ser mostrado a notificação. Deste modo, no caso de o utilizador pedir acesso de visualização a um depósito, o `targetUserUri` é o criador do depósito. No caso de o criador do depósito aceitar o pedido de acesso o `targetUserUri`, é o utilizador que tinha realizado o pedido de acesso ao depósito.

4.2.2 Detalhes de implementação

A extensão foi implementada, pensada e desenvolvida com o intuito de ser configurável através da alteração de certos valores ou expressões nos ficheiros de configuração.

Serão discutidas duas das funcionalidades implementadas, em que serão mencionadas as suas finalidades, como foram implementadas e como podem ser facilmente configuráveis nos ficheiros de configuração.

4.2.2.1 Job para retirar o embargo de um depósito

No momento de exportação dos dados e descrições para um depósito no DendroShare, o utilizador tem a possibilidade de escolher a visibilidade que pretende atribuir ao depósito. No caso de o utilizador escolher o tipo de visibilidade `embargoed`, terá de inserir uma data de fim de embargo do depósito. Depois da data de embargo passar, o depósito terá que alterar a sua visibilidade de `embargoed` para `public` e apagar o descritor `embargoedDate` do depósito. Deste modo, implementou-se um *job*, que é executado periodicamente de acordo com uma expressão de sintaxe *cron*, que é parametrizável num ficheiro de configuração da aplicação. Neste caso, para um *job* que seja executado diariamente às 00:00h, recorre-se à expressão `00 * * *`, que se encontra associada ao parâmetro **disembargo_cron_expression** localizado no ficheiro **dev.yml**, com o directório **conf/deployment_configs/dev.yml**. Para alterar quando o *job* é executado, basta alterar a expressão do descritor **disembargo_cron_expression** e possibilitar que o *job* seja executado a outras horas ou até mesmo várias vezes ao dia.

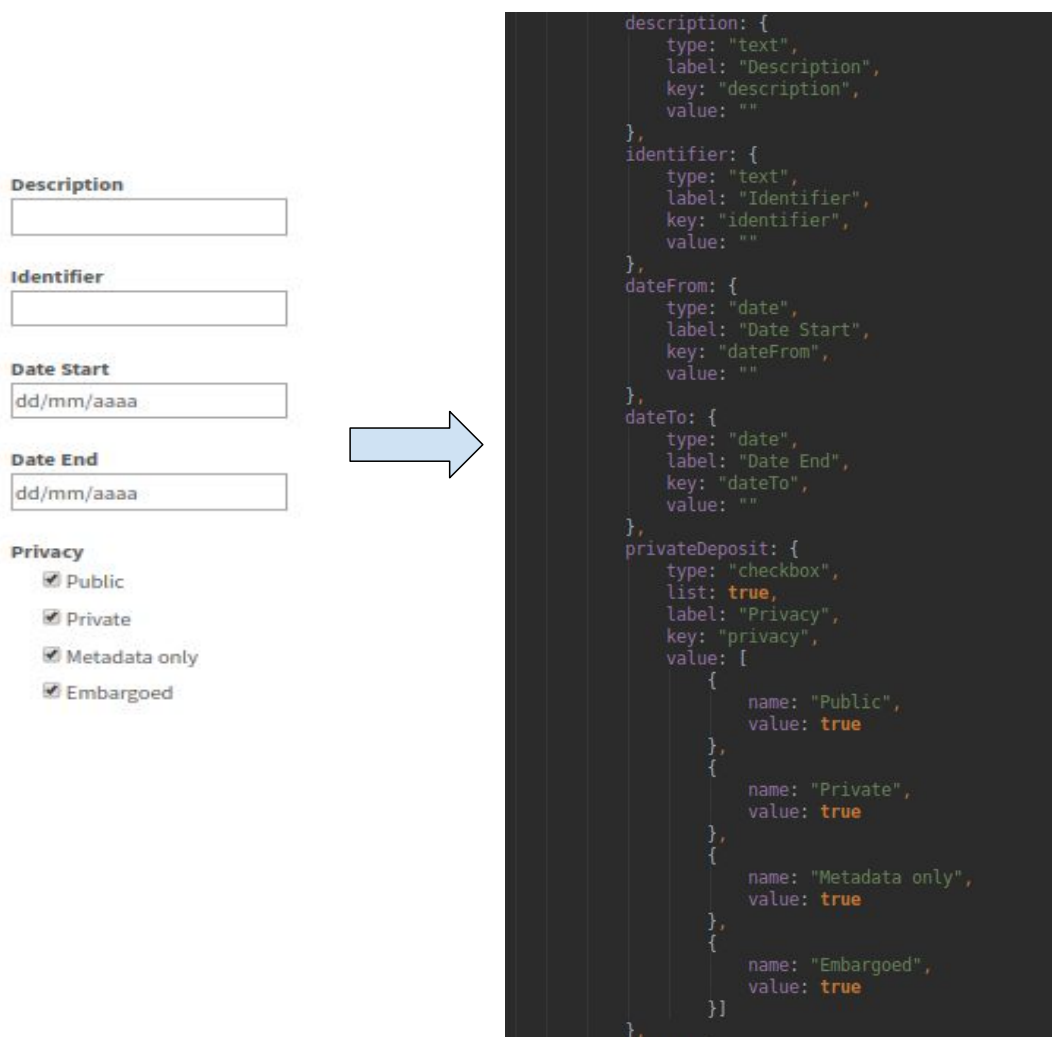


Figura 4.8: Parametrização do ecrã de pesquisa de conjuntos de dados do DendroShare

4.2.2.2 Parametrização do ecrã de pesquisa

Os critérios de pesquisa de conjuntos de dados na página inicial do DendroShare podem ser facilmente configuráveis no controlador do ecrã de pesquisa, disponível em **public/app/controllers/deposit_controller.js**, com o diretório **confdeployment_configs/dev.yml**.

Os descritores encontram-se escritos em formato JSON, para facilitar a configuração dos descritores a utilizar. Na Fig 4.8, é possível no lado esquerdo visualizar a interface de pesquisa gerada a partir do bocado de código escrito no lado direito. Neste ficheiro é possível adicionar vários tipos de restrições, cujos controlos podem ser caixas de texto, caixas de combinação ou caixas de verificação, e respetivos valores possíveis.

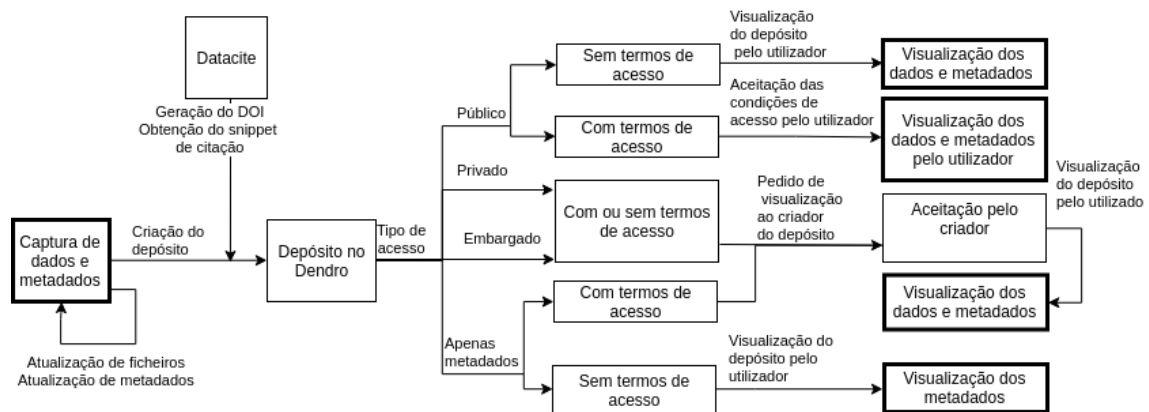


Figura 4.9: Fluxo de navegação entre as funcionalidades da plataforma DendroShare

4.3 Demonstração da utilização do sistema

Os utilizadores da plataforma podem usar a plataforma DendroShare para armazenar, descrever e organizar os seus conjuntos de dados de investigação. Demonstrar-se-á como empacotar e publicar um conjunto de dados no DendroShare com diferentes níveis de acesso, recuperá-lo através da pesquisa facetada e obter rapidamente um *snippet* de citação do depósito gerado automaticamente pela plataforma.

Vamos ilustrar o uso do DendroShare em torno de um fluxo de trabalho completo, que abrange os casos de uso comuns de gestão de dados. O utilizador começa por criar um projeto (que é semelhante a uma pasta partilhada na solução de armazenamento em nuvem *Dropbox*). Depois disso, pode carregar alguns ficheiros que de seguida podem ser descritos, usando vários descritores genéricos e específicos de domínio. Os descritores podem ser adicionados a quaisquer arquivos e pastas, bem como ao próprio projeto, pois é semelhante a uma estrutura de diretórios Fig 4.10.

Posteriormente, o utilizador poderá adicionar um colaborador ao projeto, para permitir mais *uploads* de ficheiros ou descrições. Assim que uma pasta esteja pronta para depósito, o utilizador poderá realizar a exportação no DendroShare. Durante o depósito, é possível escolher a visibilidade do mesmo, que varia entre privado, público, embargado e apenas visíveis os metadados Fig 4.11.

Após a exportação do depósito, este aparecerá no interface de pesquisa que é visível na Fig 4.12). Através da interface de pesquisa facetada, é possível encontrar depósitos ao combinar filtros, como o identificador do criador do depósito, o título do depósito, a data de publicação, o status de visibilidade ou pesquisa de texto livre sobre um descritor exato. Além disso, é possível adicionar filtros com base em qualquer descritor de metadados presente nos depósitos (seleccionando pares de descritor e respetivo valor). Um conjunto de depósitos será retornado consoante as restrições especificadas na forma de uma lista paginada.

Nos resultados da pesquisa, é possível ver o título do depósito, a descrição, a data de exportação e um ícone que representa a visibilidade do depósito (canto inferior esquerdo de cada linha de resultado da pesquisa).

The screenshot shows the DendroShare interface. At the top, there is a navigation bar with the Dendro logo, 'My' and 'Find' dropdown menus, a search bar labeled 'Search Dendro' with a 'SUBMIT' button, and a user profile for 'Dendro 1 Demo User 1' with a notification bell icon.

On the left side, there is a sidebar with an 'UPLOAD' button and a list of items: 'My Projects', a 'Folder', and a PDF file named 'inqueritoInicial.pdf'. Below this is a list of project files, including 'Urban Sense 2901'.

The main content area is titled 'Information about Urban Sense 2901' and contains a list of metadata fields:

- Description:** Estes dados são resultados do projeto UrbanSense, pertencem ao domínio ambiental.
- Title:** Urban Sense 2901
- Creator:** /r/user/9391ab68-a99e-4228-9763-1df6cf5f7e52
- Language:** en
- Project Handle:** urbansense2901
- Privacy:** private
- license:** Creative Commons Attribution 4.0

Figura 4.10: Página de um projeto, com os ficheiros à esquerda e metadados no centro

Select a repository ×

My repository bookmarks

SELECT A DESTINATION REPOSITORY ▾ ↻ ✕ CLEAR ALL

Connect to a new repository

CHOOSE REPOSITORY TYPE ▾

Share current folder to Dendro

Label

Dendro

Title of deposit:

geosapi: GeoServer REST API R Interface

Description of deposit:

This repository contains all publicly available numerical relativity surrogate data for waveforms produced by the Spectral Einstein Code.

User affiliation:

Inesc Tec 👤

Privacy:

Make this deposit public

Make this registry embargoed until

Make this deposit private

Make public metadata only

Access conditions:

Complete your conditions of access to the deposit

✕ DELETE THIS BOOKMARK 📤 SEND

Figura 4.11: Interface da criação de um depósito no DendroShare

Dendro^β My Find Search Dendro SUBMIT Dendro 1 Demo User 1

Latest deposits

Advanced search

Username

Deposit title

Description

Identifier

Date Start

Date End

Privacy

- Public
- Private
- Metadata only
- Embargoed

Repository Used

- CKAN (0)
- Dendro (4)
- DSpace (0)
- EPrints (0)
- Figshare (0)
- Zenodo (0)
- EUDAT B2Share (0)

Order By

DATE DESCENDING

Descriptor Value

ANY

Depth of search in the deposit

ONLY ON CHILDREN'S NODES

Search for descriptor...

Deposits per page

5

GS2 v8.0.2 <small>GS2 is a physics application, developed to study low-frequency turbulence in magnetized plasma.</small>	26.06.2019 10:53:47
Gene Ontology Data Archive <small>Archival bundle of GO data release.</small>	26.06.2019 10:53:21
D2.2 PID Metadata Provenance <small>We describe the implementations by Crossref and ORCID done previously, the...</small>	26.06.2019 10:53:00
Real-time optical and electronic sensing with a β-amino enone linked, triazine-containing 2D covalent organic framework <small>*Real-time optical and electronic sensing with a β-amino enone linked, triazine-containing 2D covalent organic framework*...</small>	26.06.2019 10:52:25

Figura 4.12: Listagem de todos os depósitos da plataforma

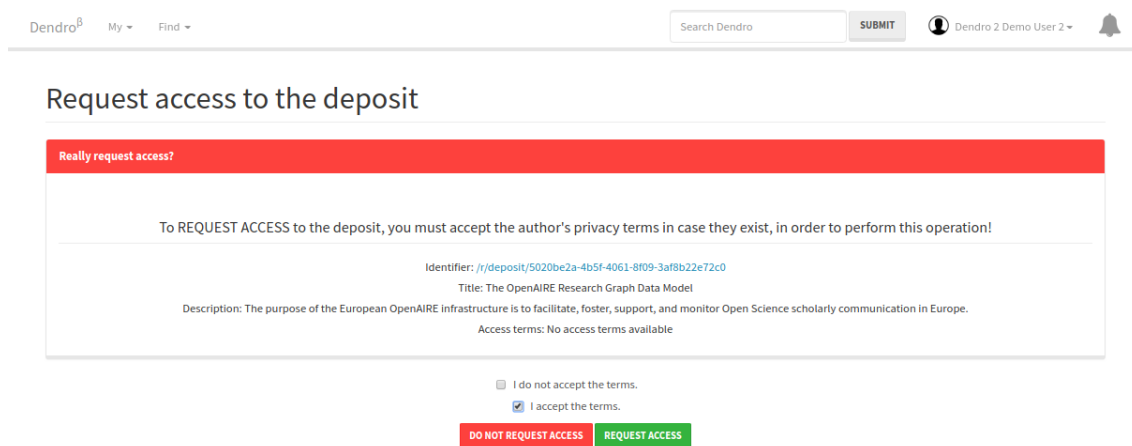


Figura 4.13: Pedido de acesso ao depósito

Os utilizadores apenas podem aceder a depósitos da plataforma no caso da visibilidade ser pública, ou aceder apenas aos metadados no caso da visibilidade ser “apenas metadados”. No caso do depósito ser privado ou embargado, o utilizador tem a opção de solicitar acesso de visualização ao criador do depósito, como se pode ver na Fig 4.13.

O criador do depósito será notificado sobre a solicitação na próxima vez que fizer login na plataforma (ver Fig 4.14). Depois de abrir a notificação, o criador pode aceitar ou recusar o pedido de acesso ao depósito. Mesmo depois de aprovado o acesso, o criador retém o controlo total, podendo revogar quaisquer acessos previamente concedidos (Fig 4.15).

Após um depósito de dados na plataforma, o utilizador pode ver não apenas os dados, mas também os metadados adicionados ao depósito, assim como os seus ficheiros e pastas individuais. O DendroShare facilita a citação ao fornecer, um trecho de citação do BibTeX (gerado pelo DataCite). Os metadados incluem um DOI, que também está presente no trecho de citação (Fig 4.16).

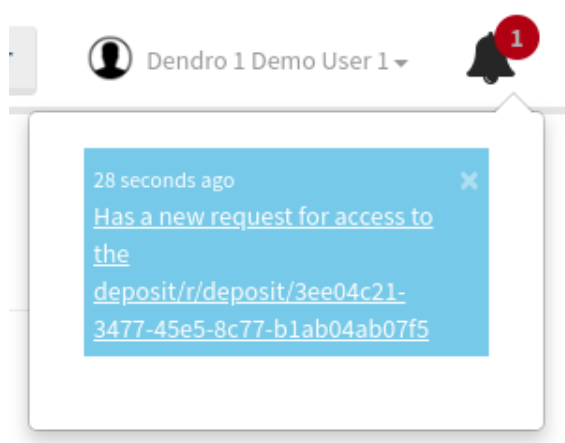


Figura 4.14: Notificação recebida pelo criador para pedido de visualização de um depósito

Information Management User access

Users with the accepted conditions

/r/user/88059641-f1a8-4308-8238-0fe332ae2a67 NOT ALLOW ACCESS

Conditions to accept

/r/user/2c0c4ac-67bd-40f2-9085-dcbc22d8aa90 DECLINE ACCEPT

Figura 4.15: Gestão dos acessos dos utilizadores ao depósito

Dendro^β My Find Search Dendro SUBMIT Dendro 1 Demo User 1

Folder

- Urban Sense
- paper_66.pdf

Information Management User access

Information about Urban Sense

Proposed Citation:

@misc[<https://doi.org/10.23673/bb0e4ae0-96d4-11e9-883c-61694c3189a9>, doi = {10.23673/BB0E4AE0-96D4-11E9-883C-61694C3189A9}, url = {<http://127.0.0.1:3001/r/deposit/f02e3dd3-1bbb-4cfe-b692-da04b3d90adf>}, author = {Demo User 1, Dendro 1}, title = {Urban Sense}, year = {2019}]

Description

description

Title

Urban Sense

creator

/r/user/9391ab68-a99e-4228-9763-1df6cf5f7e52

Date

"2019-06-24T23:06:00.000Z"

identifier

10.23673/bb0e4ae0-96d4-11e9-883c-61694c3189a9

Language

en

Privacy

public

Figura 4.16: Página do depósito

Capítulo 5

Experiência de avaliação

Para avaliar a eficiência da extensão realizada e compará-la com as plataformas existentes (CKAN e Dataverse), foi realizado um conjunto de tarefas nas 3 plataformas por um conjunto de utilizadores, seguindo o workflow típico de cada das plataformas.

As experiências consistem na realização de pequenas tarefas específicas do ciclo da gestão de dados de investigação. Ao longo deste capítulo, serão especificadas as experiências realizadas, as métricas de avaliação da eficiência da extensão da plataforma, os questionários de pré e pós experiência e o guião para as tarefas de avaliação. Os questionários de pré e pós experiência, assim como o guião para as tarefas de avaliação encontram-se em anexo.

Ao longo deste capítulo, e sempre que for utilizado o termo DendroShare, pretende-se referir a plataforma Dendro com a adição do módulo de partilha de dados DendroShare.

5.1 Características dos utilizadores avaliados

Nesta experiência participaram 16 utilizadores, representados pelos seus respetivos ID's e caracterizações, representados na Tab 5.1.

5.2 Questionário pré-experiência

No sentido de obter informação da caracterização dos utilizadores avaliados, realizou-se um pequeno questionário, apresentado no Anexo B.1, que tinha como objetivo avaliar a experiência do utilizador na utilização de plataformas de armazenamento de dados e na frequência com que cita e partilha dados de investigação.

Na elaboração do questionário recorreu-se à escala de Likert, para permitir a distribuição das respostas por cinco níveis e obter o nível de concordância ou não concordância dos utilizadores às afirmações impostas.

Cada pergunta do questionário tinha como opções de resposta: Concordo totalmente, Concordo parcialmente, Nem concordo nem discordo, Discordo parcialmente, Discordo totalmente.

As perguntas do questionário de pré-experiência encontram-se representadas na Tab 5.2.

Tabela 5.1: Caracterização dos utilizadores avaliados

ID	Idade	Sexo	Descrição
USER1	23	Feminino	Estudante do MIEIC
USER2	23	Masculino	Estudante do MIEIC
USER3	20	Feminino	Estudante da UFP
USER4	22	Masculino	Estudante do MIEIC
USER5	23	Masculino	Estudante do MIEM
USER6	24	Masculino	Estudante do MIEIC
USER7	34	Masculino	Estudante do MAP-I
USER8	22	Feminino	Estudante do MCI
USER9	37	Feminino	Estudante do PDMD
USER10	25	Feminino	Estudante do PDMD
USER11	36	Feminino	Professora do ensino secundário
USER12	19	Masculino	Estudante do ISEP
USER13	20	Feminino	Estudante de ciências da comunicação
USER14	23	Feminino	Estudante de Enfermagem ESSSM
USER15	18	Masculino	Estudante de MIEEC
USER16	32	Masculino	Professor assistente da FEUP

Tabela 5.2: Perguntas do questionário pré-experiência

ID	Descrição
QI1	Estou familiarizado com o conceito de partilha e citação de dados de investigação
QI2	Estou a par dos benefícios da reutilização de dados de investigação
QI3	Regularmente descrevo e armazeno dados de investigação
QI4	Regularmente cito dados de investigação
QI5	Regularmente partilho dados de investigação produzidos por mim ou pelo meu grupo de investigação
QI6	Tenho dificuldades na criação da citação do conjunto de dados
QI7	Nos projetos nos quais participo existe a obrigatoriedade de partilha de dados de investigação em regime de acesso aberto

Estou familiarizado com o conceito de partilha e citação de dados de investigação

16 respostas

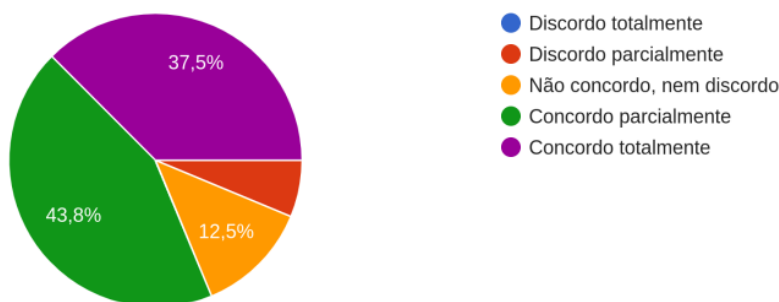


Figura 5.1: Resultados da pergunta Q11

Os resultados obtidos na pergunta **Q11**, representados na figura 5.1, demonstram que a maioria da amostra escolhida (81.3%), apresenta alguma familiarização com o conceito de partilha e citação de dados de investigação.

Para os resultados obtidos da pergunta **Q12**, representados na figura 5.2, indicam que grande parte da amostra escolhida conhece os benefícios da reutilização dos dados de investigação.

Os resultados à pergunta **Q13**, representados na figura 5.3, demonstram que nem toda a amostra costuma descrever e armazenar dados de investigação, mas mesmo assim 43.8% das pessoas da amostra concordaram com a afirmação.

Com os resultados à pergunta **Q14**, representados na figura 5.4, evidenciam que a maioria dos utilizadores costumam citar dados de investigação.

Nos resultados obtidos na pergunta **Q15**, representados na figura 5.5, indicam que nem todos os utilizadores da amostra costumam partilhar dados de investigação por eles ou pelo seu grupo de trabalho.

Nos resultados da pergunta **Q16**, representados na Fig 5.6, é possível verificar que a maioria da amostra escolhida (56.3%) tem dificuldades na criação da citação do conjunto de dados.

Com os resultados da última pergunta do questionário de pré-experiência, representados na Fig 5.7, indicam que os projetos onde os utilizadores participam (75.1%), existe a obrigatoriedade de partilha dos dados de investigação em regime de acesso aberto.

5.3 Tarefas de avaliação

A extensão DendroShare na plataforma Dendro foi testada em comparação com as plataforma CKAN e Dataverse por 16 utilizadores.

Estou a par dos benefícios da reutilização de dados de investigação

16 respostas

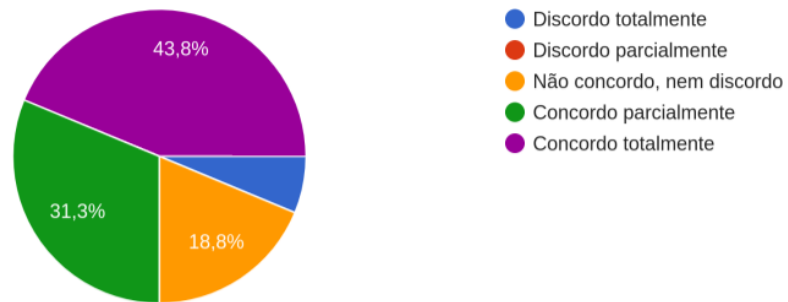


Figura 5.2: Resultados da pergunta Q12

Regularmente descrevo e armazeno dados de investigação

16 respostas

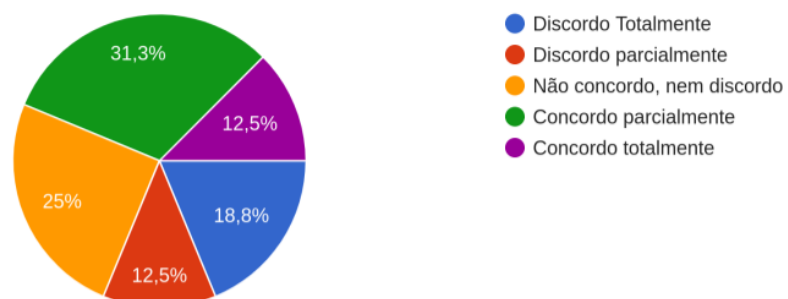


Figura 5.3: Resultados da pergunta Q13

Regularmente cito dados de investigação

16 respostas

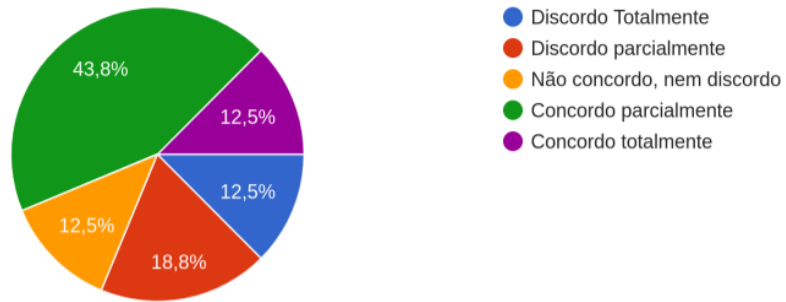


Figura 5.4: Resultados da pergunta QI4

Regularmente partilho dados de investigação produzidos por mim ou pelo meu grupo de investigação

16 respostas

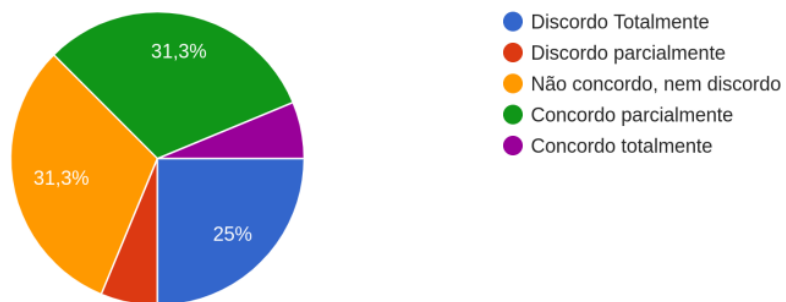


Figura 5.5: Resultados da pergunta QI5

Tenho dificuldades na criação da citação do conjunto de dados

16 respostas

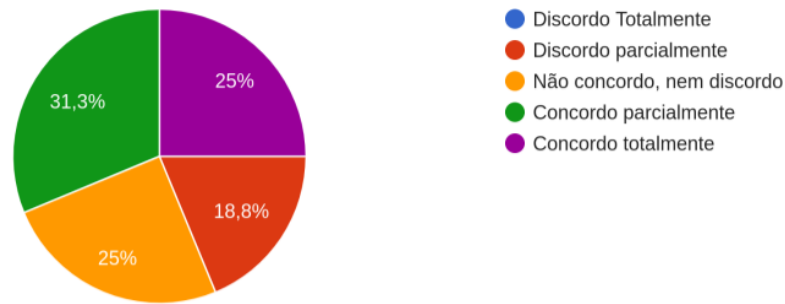


Figura 5.6: Resultados da pergunta Q16

Nos projetos nos quais participo existe a obrigatoriedade de partilha de dados de investigação em regime de acesso aberto

16 respostas

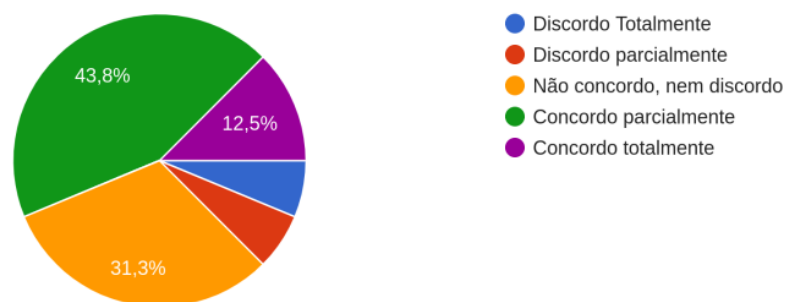


Figura 5.7: Resultados da pergunta Q17

Tabela 5.3: Tarefas realizadas pelos utilizadores

ID	Descrição
TAREFA1	Criar um projeto ou um Dataset na plataforma
TAREFA2	Fazer upload de ficheiros na plataforma
TAREFA3	Adicionar alguns metadados ao Projeto ou Dataset criado
TAREFA4	Exportar um depósito de dados acabados no DendroShare
TAREFA5	Pesquisar um depósito na plataforma
TAREFA6	Obter o <i>snippet</i> de citação de um depósito da plataforma

Os utilizadores realizaram o mesmo conjunto de tarefas em cada plataforma; no entanto, foi seguido o workflow típico de cada plataforma, logo certos passos não foram realizados pela mesma ordem.

Os utilizadores avaliados realizaram as tarefas, representadas na Tab 5.3, nas 3 diferentes plataformas (DendroShare, CKAN e Dataverse). No entanto, a **TAREFA4** só foi realizada no DendroShare, visto que esta tarefa não se assemelha a nenhuma tarefa do workflow das outras plataformas.

Nestas tarefas de avaliação, pretende-se avaliar a funcionalidade de armazenamento de ficheiros em cada plataforma, a descrição de ficheiros e datasets nas plataformas, a pesquisa de datasets específicos em cada uma das plataforma e por último, a obtenção do *snippet* de citação de um conjunto de dados em cada uma das 3 plataformas.

As tarefas de avaliação encontram-se representadas na Tab 5.3.

5.4 Métricas de avaliação

Para avaliar a eficiência da extensão realizada na plataforma DendroShare, pretende-se verificar a taxa de sucesso das tarefas realizadas por cada utilizadores nas 3 plataformas, o tempo médio da realização de cada tarefa pelo utilizador em cada uma das 3 plataformas, a taxa de erros cometidos pelo utilizador em cada uma das tarefas das 3 plataforma e por último, uma avaliação subjetiva de cada utilizador para cada uma das plataformas [31].

Os tempos da realização de cada tarefa foram cronometrados e anotado o número de erros cometidos por cada utilizador na realização das tarefas, assim como o número de ajudas fornecidas aos utilizadores para a realização das tarefas. Estes registos foram todos inseridos de forma correta num questionário de avaliação disponibilizado no Anexo C.

5.5 Análise de resultados

Nesta secção, são comparados os desempenhos dos utilizadores na execução das tarefas nas 3 plataformas (DendroShare, Dataverse e CKAN).

Mediana dos tempos das tarefas realizadas pelos utilizadores

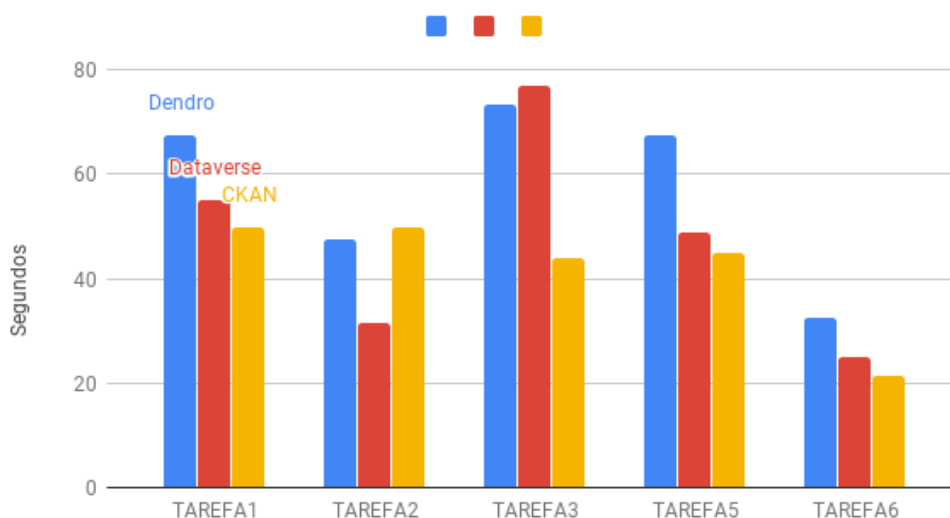


Figura 5.8: Mediana dos tempos de realização das tarefas de avaliação

Resultados da experiência

Apenas foram comparadas as Tarefas 1, 2, 3, 5 e 6 entre as diferentes plataformas, visto a Tarefa 4 ter sido apenas realizada no DendroShare, pois não existe nenhuma tarefa nas outras plataformas que se assemelhe a esta tarefa.

Para cada plataforma foi contabilizado o tempo de realização de cada tarefa pelos utilizadores e realizados gráficos para a média, mediana e o desvio padrão dos resultados obtidos. A plataforma CKAN foi a que obteve melhor mediana nos tempos de realização das tarefas em praticamente todas as tarefas. Tais resultados podem ser consultados na Fig 5.8.

A média dos tempos de realização das tarefas nas diferentes plataformas apresenta pouca variação de tempos, não sendo maior do que 20 segundos a diferença de tempos entre a plataforma que obteve melhor desempenho e a que obteve pior desempenho em todas as tarefas. O gráfico pode ser consultado na Fig 5.9.

O desvio padrão apresenta pouca variação entre as várias plataformas em cada uma das tarefas, no entanto, na TAREFA6 destaca-se um elevado valor do desvio padrão na plataforma CKAN. Tal valor deve-se à diferença de tempos dos utilizadores que não concluem a tarefa e os utilizadores que concluem a tarefa. Os resultados podem ser consultados na Fig 5.10.

Entre as várias tarefas pedidas aos utilizadores para realizarem nas diferentes plataformas, todas as tarefas foram sempre concluídas ao contrário da TAREFA6, na plataforma CKAN. Deste modo, são apresentados através da tabela Tab 5.4 os ID's dos utilizadores avaliados e o resultado da conclusão da tarefa.

Para além dos tempos de realização das tarefas, foi contabilizado o número de ajudas prestadas e o número de erros cometidos pelos utilizadores em cada uma das tarefas. Com isto, pretende-se

Média dos tempos das tarefas realizadas pelos utilizadores

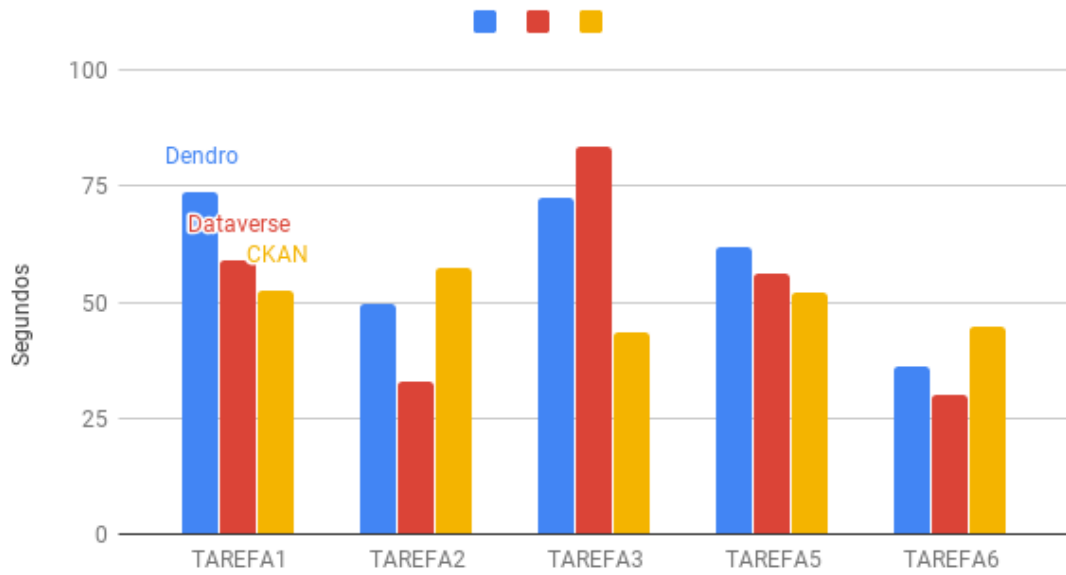


Figura 5.9: Média dos tempos de realização das tarefas de avaliação

Desvio padrão dos tempos das tarefas realizadas pelos utilizadores

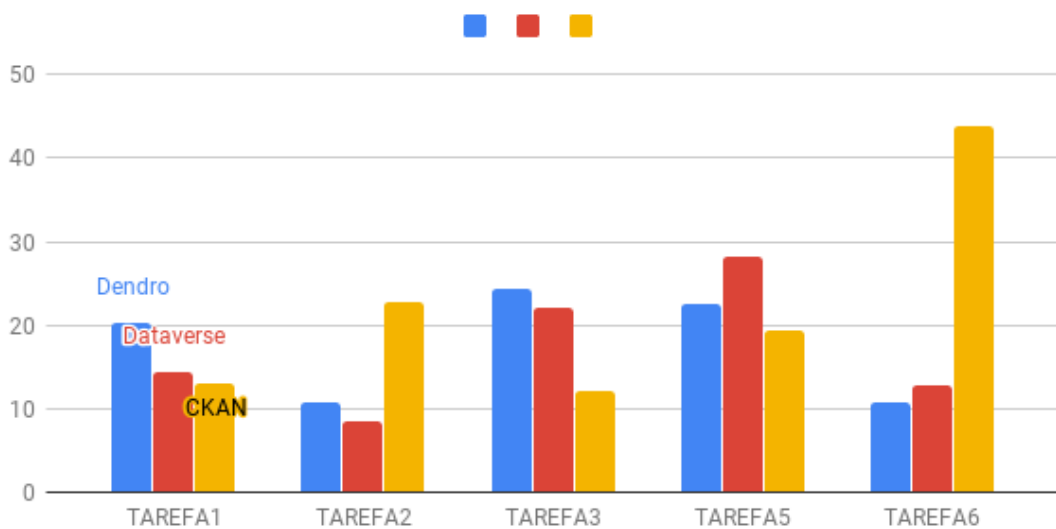


Figura 5.10: Desvio padrão dos tempos de realização das tarefas de avaliação

Tabela 5.4: Resultados da conclusão da TAREFA6 do CKAN pelos utilizadores

ID	Conclui a tarefa6
USER1	Não
USER2	Não
USER3	Não
USER4	Não
USER5	Não
USER6	Não
USER7	Sim
USER8	Não
USER9	Não
USER10	Sim
USER11	Sim
USER12	Não
USER13	Não
USER14	Não
USER15	Não
USER16	Sim

retirar o quanto a plataforma é intuitiva e de fácil utilização para o utilizador. Os resultados pode ser consultados na Fig 5.11 e na Fig 5.12. Entre os 16 utilizadores utilizados para a realização das tarefas nas 3 plataformas, foram contabilizados poucas ajudas e erros cometidos pelos utilizadores em cada uma das plataformas, no entanto, no CKAN os utilizadores cometeram mais erros e necessitaram de mais ajudas na TAREFA2.

5.6 Questionário pós-experiência

No sentido de tentar recolher informações relativas às experiências realizadas em cada plataforma, recorreu-se a um questionário de pós-experiência, Anexo B.2, com o objetivo de avaliar a dificuldade na realização e conclusão das tarefas pelos utilizadores em cada uma das plataformas, bem como avaliar a satisfação dos utilizadores na utilização de cada uma das plataforma.

Na elaboração do questionário de pós-experiência, recorreu-se à escala de Likert para permitir a distribuição das respostas por cinco níveis e possibilitar a obtenção do grau de concordância ou não concordância às afirmações impostas.

Cada pergunta do questionário tem como opções de resposta: Concordo totalmente, Concordo parcialmente, Nem concordo nem discordo, Discordo parcialmente e Discordo totalmente.

As perguntas do questionário de pós-experiência encontram-se representadas na Tab 5.5.

Número de ajudas prestadas durante as tarefas de avaliação

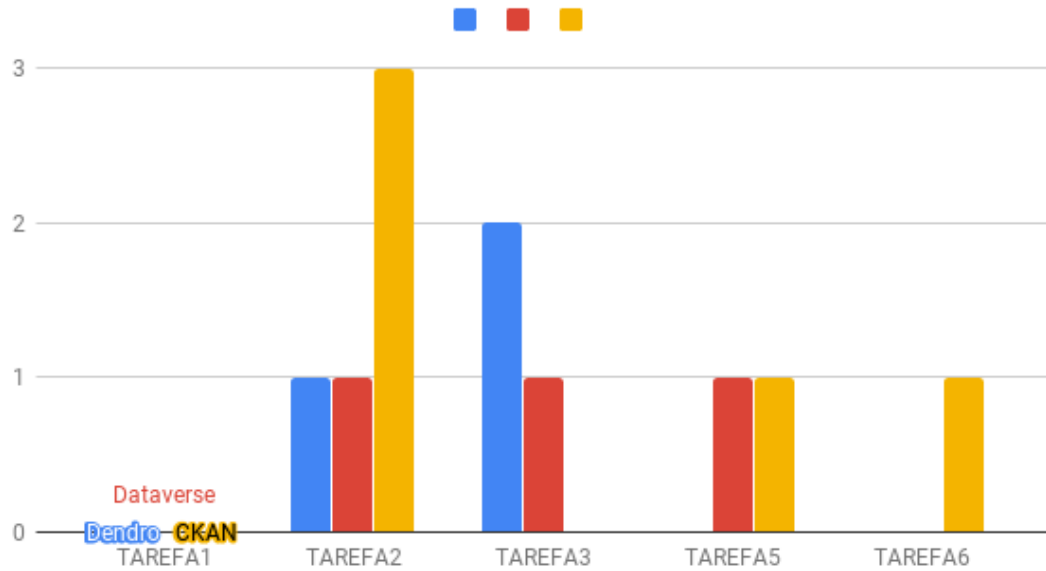


Figura 5.11: Número de ajudas prestadas durante as tarefas de avaliação

Número de erros cometidos pelos utilizadores durante as tarefas de avaliação

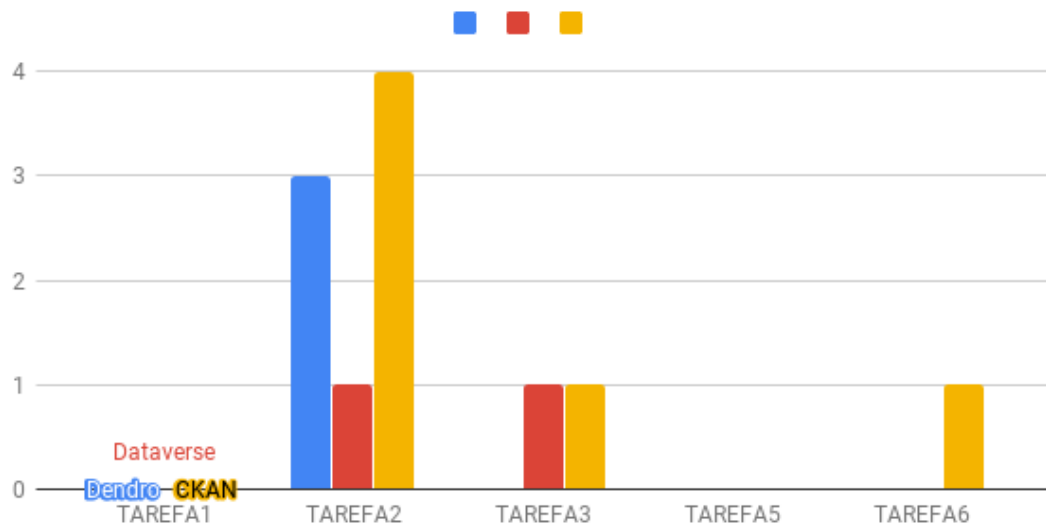


Figura 5.12: Número de erros cometidos pelos utilizadores durante as tarefas de avaliação

Tabela 5.5: Perguntas do questionário de pós-experiência

ID	Descrição
QF1	Tive dificuldade na realização das tarefas
QF2	Consegui concluir as tarefas propostas
QF3	Tive dificuldade na criação da citação do conjunto de dados
QF4	A plataforma usada tem uma interface <i>user-friendly</i>
QF5	Utilizaria esta plataforma no meu grupo de investigação
QF6	A plataforma é capaz de satisfazer as necessidades de partilha de dados no meu grupo de investigação
QF7	A plataforma é capaz de satisfazer as necessidades de citação de dados no meu grupo de investigação
QF8	Caso queira adicionar mais alguma informação relativamente à plataforma, indique-a neste espaço abaixo.

Resultados do Questionário de pós-experiência para a plataforma DendroShare

Perante os resultados da questão **QF1** podemos verificar que a maioria dos utilizadores não tiveram qualquer dificuldade na realização das tarefas no DendroShare, o que mostra o quanto a plataforma é intuitiva. Os resultados estão representados na Fig 5.13.

Os resultados obtidos na questão **QF2**, permitem concluir que a maioria dos utilizadores conseguiu concluir as tarefas propostas e obter os resultados esperados. A apuração a esta pergunta encontram-se representados na Fig 5.14.

A plataforma DendroShare tem na página inicial de cada depósito o *snippet* de citação do depósito, deste modo, o utilizador não tem qualquer dificuldade na criação do *snippet* de citação do conjunto de dados. Os resultados relativos à questão **QF3** estão disponíveis na Fig 5.15.

Os resultados à da questão **QF4**, permitem verificar que a plataforma DendroShare apresenta uma interface *user-friendly* e intuitiva para os utilizadores. Os resultados encontram-se na Fig 5.16.

Na questão **QF5**, os utilizadores avaliados na sua maioria, responderam que utilizariam a plataforma DendroShare no seu grupo de investigação, o que demonstra mas uma vez o quanto a plataforma é intuitiva e de fácil utilização para o utilizador. Os resultados encontram-se representados na Fig 5.17.

Na questão **QF6** todos os utilizadores concordaram que a plataforma DendroShare satisfaz as necessidades de partilha de dados de investigação, o que reforça os resultados da pergunta **QF5**. A apuração à questão **QF6** encontra-se disponível na Fig 5.18.

A questão **QF7**, está de certa forma relacionada com a questão **QF3**, pois se um utilizador tivesse dificuldade na obtenção do *snippet* de citação para um depósito da plataforma, esta plataforma não satisfazia as necessidades de citação de dados de investigação do seu grupo de investigação. Os resultados da **QF7**, encontram-se na Fig 5.19.

Tive dificuldade na realização das tarefas

16 respostas

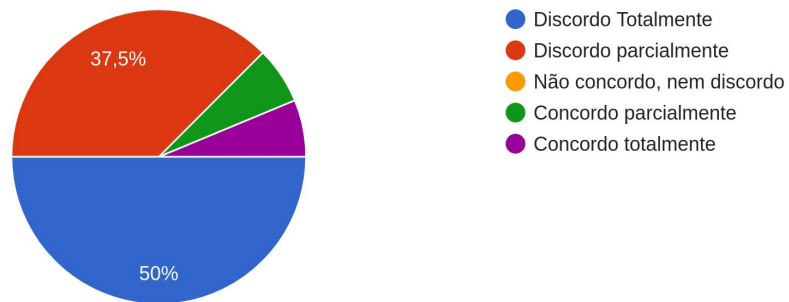


Figura 5.13: Resultados da pergunta QF1 na plataforma DendroShare

Consegui concluir as tarefas propostas

16 respostas

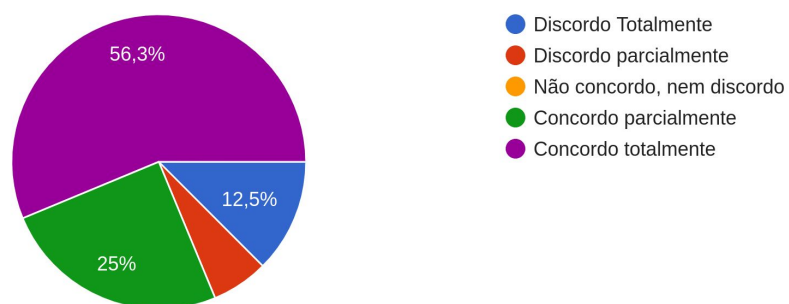


Figura 5.14: Resultados da pergunta QF2 na plataforma DendroShare

Tive dificuldade na criação da citação do conjunto de dados

16 respostas

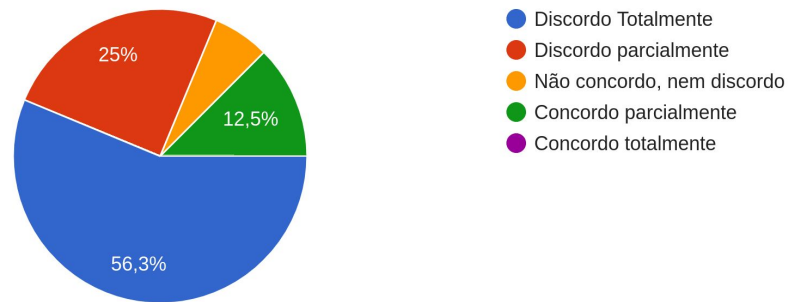


Figura 5.15: Resultados da pergunta QF3 na plataforma DendroShare

A plataforma usada tem uma interface user-friendly

16 respostas

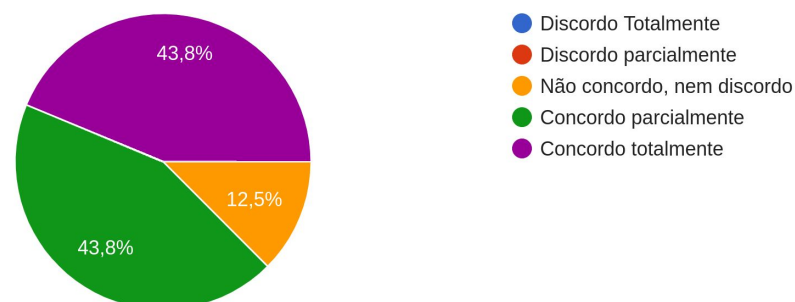


Figura 5.16: Resultados da pergunta QF4 na plataforma DendroShare

Utilizaria esta plataforma no meu grupo de investigação

16 respostas

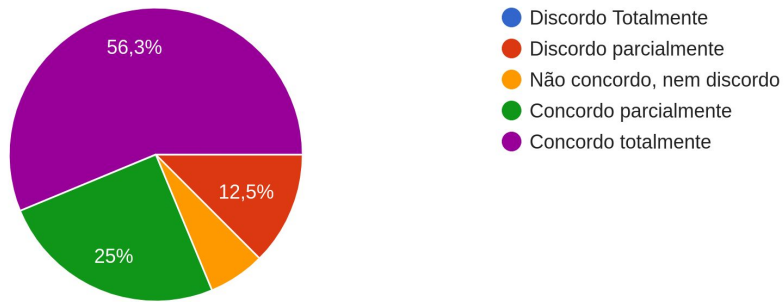


Figura 5.17: Resultados da pergunta QF5 na plataforma DendroShare

A plataforma é capaz de satisfazer as necessidades de partilha de dados no meu grupo de investigação

16 respostas

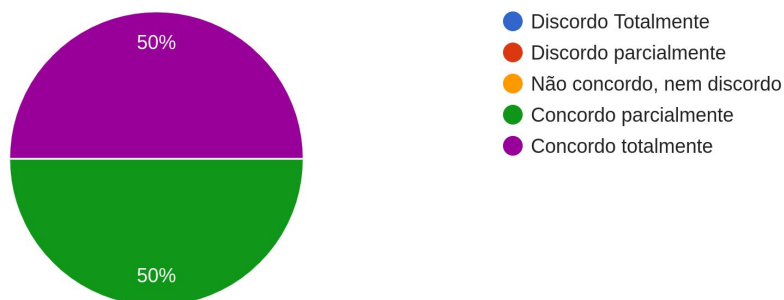


Figura 5.18: Resultados da pergunta QF6 na plataforma DendroShare

A plataforma é capaz de satisfazer as necessidades de citação de dados no meu grupo de investigação

16 respostas

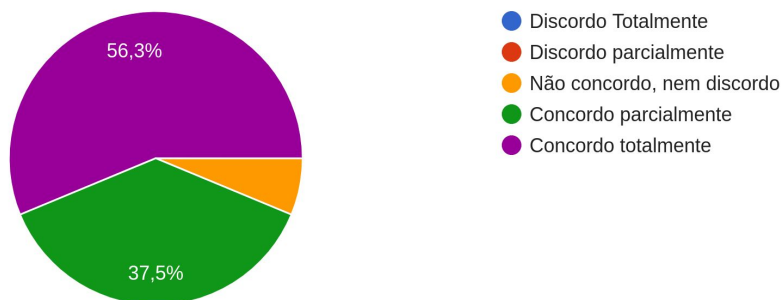


Figura 5.19: Resultados da pergunta QF7 na plataforma DendroShare

Apesar dos escassos comentários obtidos na questão **QF8**, deu para perceber que os utilizadores apenas referenciaram o *snippet* de citação. Mencionado que o *snippet* deveria estar indentado e possuir maior destaque de visibilidade. Os comentários a esta questão encontram-se na Tab 5.6.

Resultados do Questionário de pós-experiência para a plataforma Dataverse

Nos resultados obtidos obtidos à questão **QF1**, a maioria dos utilizadores não demonstrou dificuldade na realização das tarefas, no entanto, 25% dos utilizadores demonstrou alguma dificuldade. Os fatores a esta dificuldade estão de certa forma relacionados com a interface ruidosa da plataforma, será uma questão a abordar na questão **QF4**. Os resultados da questão **QF1**, encontram-se na Fig 5.20.

Nos resultados à questão **QF2**, pode-se verificar que os utilizadores conseguiram concluir as tarefas propostas. Os resultados estão localizados na Fig 5.21.

A plataforma Dataverse possui na página do conjunto de dados um botão onde é possível obter o *snippet* de citação do depósito, no entanto, a dificuldade que os utilizador tiveram na obtenção do *snippet* de citação na plataforma, deve-se ao facto de o utilizador ter dificuldade a encontrar a informação na plataforma. Os resultados da questão **QF3** encontram-se representados na Fig 5.22.

Tabela 5.6: Respostas à pergunta QF8 na plataforma DendroShare

ID	Descrição
COMMENT1	"Gostava de ver a citação devidamente indentada (opcionalmente expansível/colapsável)."
COMMENT2	"Destacar melhor a parte da citação."

Tive dificuldade na realização das tarefas

16 respostas

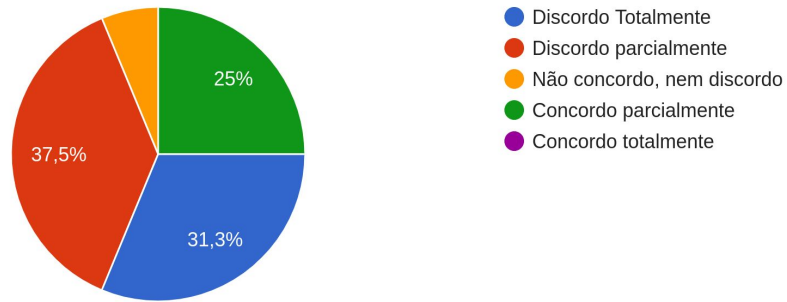


Figura 5.20: Resultados da pergunta QF1 na plataforma Dataverse

Consegui concluir as tarefas propostas

16 respostas

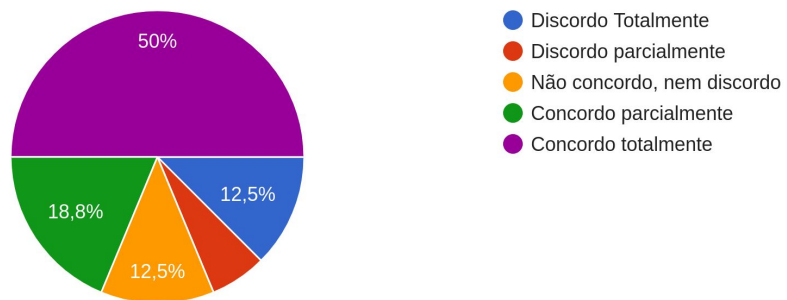


Figura 5.21: Resultados da pergunta QF2 na plataforma Dataverse

Tive dificuldade na criação da citação do conjunto de dados

16 respostas

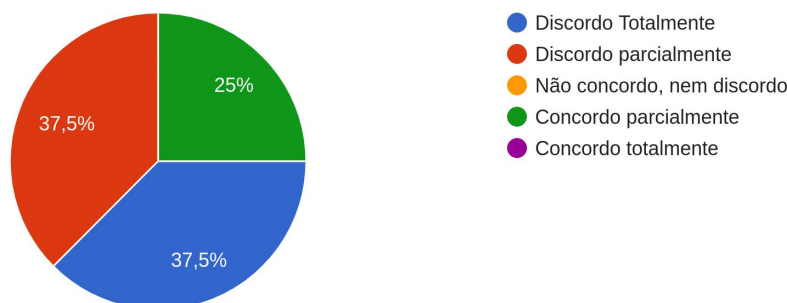


Figura 5.22: Resultados da pergunta QF3 na plataforma Dataverse

Os resultados da questão **QF4**, permitem evidenciar que um dos problemas da plataforma Dataverse, é não possuir uma interface *user-friendly*, mencionado por 25% dos utilizadores. Esta plataforma apresenta demasiada informação para o utilizador e não é dado destaque à informação que é mais relevante para o utilizador. Os resultados a esta questão encontram-se na Fig 5.23.

Na questão **QF5** a maioria dos utilizadores responderam que utilizaria o dataverse no seu grupo de investigação, no entanto, mais uma vez 25% dos utilizadores não utilizariam esta plataforma para o seu grupo de investigação. Os resultados encontram-se localizados na Fig 5.17.

Nos resultados obtidos à questão **QF6**, permite-se verificar que o Dataverse satisfaz as necessidades de partilha de dados no grupo de investigação do utilizador. Os resultados a esta questão encontram-se representados na Fig 5.25.

Os resultados à questão **QF7**, estão relacionados com os resultados da questão **QF3**, daí um utilizador utilizar esta plataforma no caso de a plataforma possibilitar o *snippet* de citação do conjunto de dados de forma automática. Os resultados a esta questão estão representados na Fig 5.26.

Apesar dos reduzidos comentários obtidos à questão **QF8**, pode-se verificar que os utilizadores se focaram na crítica à interface do Dataverse, mencionando não ser intuitiva e possuir muito ruído na interface. Os comentários a esta pergunta encontram-se representados na Tab 5.7.

Resultados do Questionário de pós-experiência para a plataforma CKAN

Nos resultados obtidos à questão **QF1**, a maioria dos utilizadores tiveram dificuldades na realização das tarefas, tal facto, se deve à dificuldade da realização da **TAREFA6** das tarefas de avaliação (Obter o *snippet* de citação do conjunto de dados de investigação). Os resultados a esta questão podem ser encontrados na Fig 5.27.

A plataforma usada tem uma interface user-friendly

16 respostas

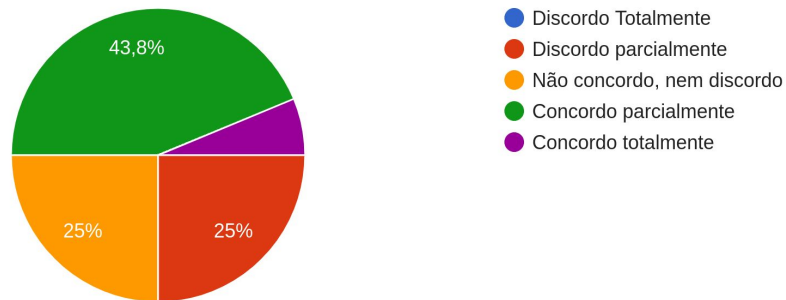


Figura 5.23: Resultados da pergunta QF4 na plataforma Dataverse

Utilizaria esta plataforma no meu grupo de investigação

16 respostas

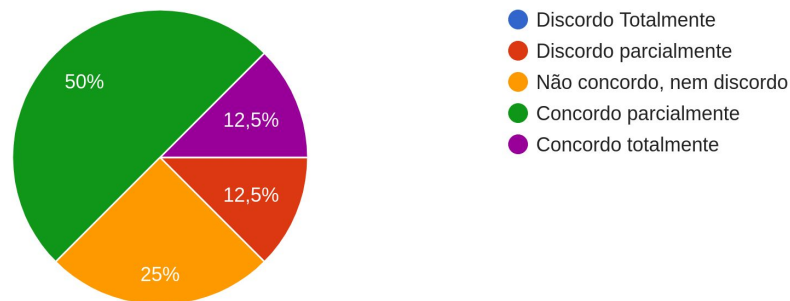


Figura 5.24: Resultados da pergunta QF5 na plataforma Dataverse

Tabela 5.7: Respostas à pergunta QF8 na plataforma Dataverse

ID	Descrição
COMMENT1	"A UI é extremamente ruidosa, o que torna a familiarização com o Dataverse complicada. O design é agradável, mas a quantidade de funcionalidades apresentada ao utilizador de uma só vez torna o sistema irritante."
COMMENT2	"Não achei a plataforma muito intuitiva."

A plataforma é capaz de satisfazer as necessidades de partilha de dados no meu grupo de investigação

16 respostas

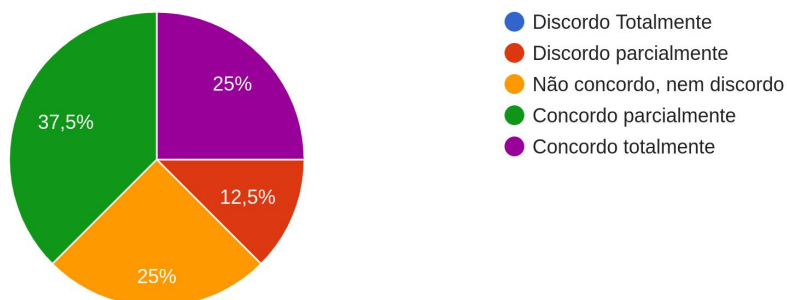


Figura 5.25: Resultados da pergunta QF6 na plataforma Dataverse

A plataforma é capaz de satisfazer as necessidades de citação de dados no meu grupo de investigação

16 respostas

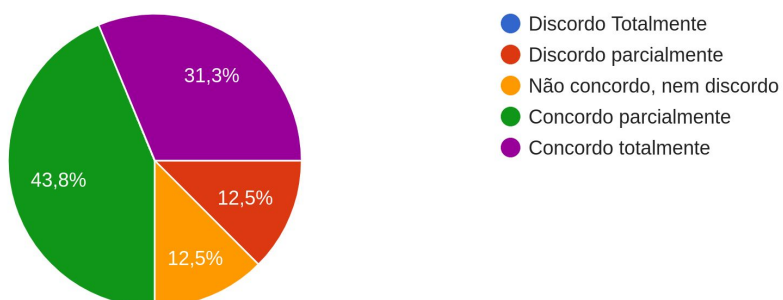


Figura 5.26: Resultados da pergunta QF7 na plataforma Dataverse

Tive dificuldade na realização das tarefas

16 respostas

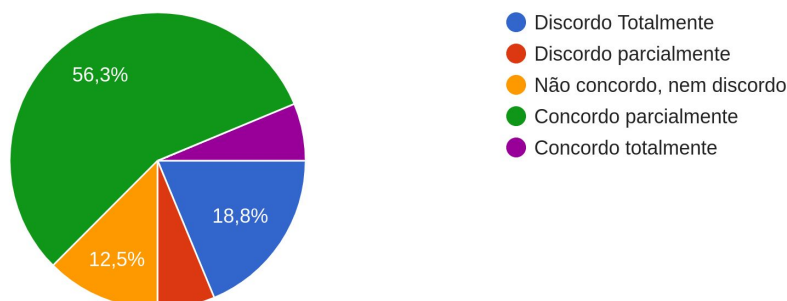


Figura 5.27: Resultados da pergunta QF1 na plataforma CKAN

Apesar de a maioria dos utilizadores afirmar ter concluído as tarefas propostas, a maioria dos utilizadores não terminaram a **TAREFA6** das tarefas de avaliação. Devido a isto, os resultados a esta questão apresentam altas percentagem nas diferentes respostas Fig 5.28.

Como já foi mencionado anteriormente, os utilizadores tiveram grandes dificuldades na criação do snippet de citação do conjunto de dados de investigação (**QF3**), pois este tinha que ser criado de forma manual pelo utilizador, ao contrário das outras plataformas. Os resultados a esta questão podem ser encontrados na Fig 5.29.

Na questão **QF4**, os utilizadores concordaram que o CKAN apresentava uma interface *user-friendly*. Tais resultados podem ser consultados na Fig 5.30.

Nos resultados à questão **QF5**, a maioria dos utilizadores não discordaram nem concordaram em relação à utilização do CKAN como plataforma do seu grupo de investigação, tal facto se deve à plataforma não possuir o *snippet* de citação do conjunto de dados de investigação. Os resultados podem ser consultados na Fig 5.31.

Na questão **QF6**, os utilizadores não concordaram nem discordaram quanto à capacidade de o CKAN satisfazer as necessidades de partilha de dados no seu grupo de investigação, tais resultados podem ser consultados na Fig 5.32.

Nos resultados obtidos à questão **QF7**, a maioria dos utilizadores concordaram que o CKAN não satisfaz as necessidade de citação de dados de investigação, tal facto se deve à plataforma não possuir a geração automática do *snippet* de citação do conjunto de dados armazenado na plataforma. Os resultados a esta questão podem ser consultados na Fig 5.33.

Na questão **QF8**, os utilizadores mencionaram que a plataforma apresentava uma boa interface, o que facilitava a realização das tarefas, mas não apresentava a funcionalidade de geração do *snippet* de citação, o que prejudicava em muito o desempenho da plataforma. Os comentários dos utilizadores podem ser consultados na Tab 5.8.

Consegui concluir as tarefas propostas

16 respostas

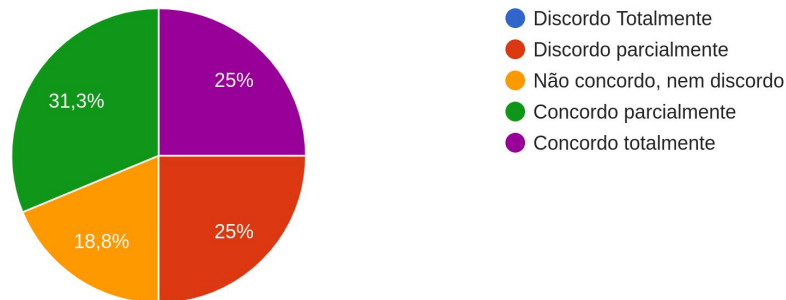


Figura 5.28: Resultados da pergunta QF2 na plataforma CKAN

Tive dificuldade na criação da citação do conjunto de dados

16 respostas

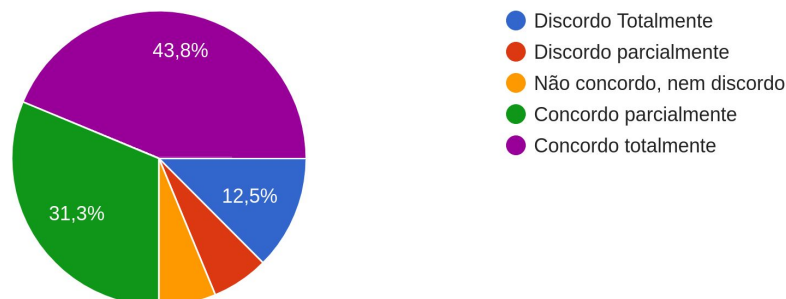


Figura 5.29: Resultados da pergunta QF3 na plataforma CKAN

Tabela 5.8: Respostas à pergunta QF8 na plataforma CKAN

ID	Descrição
COMMENT1	"Este sistema não parece oferecer muitas opções, em particular a citação dos dados seria uma funcionalidade de alta prioridade."
COMMENT2	"O CKAN foi a plataforma com que tive mais facilidade em executar as tarefas."

A plataforma usada tem uma interface user-friendly

16 respostas

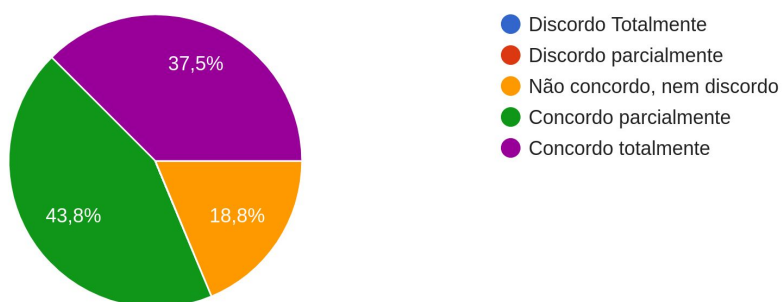


Figura 5.30: Resultados da pergunta QF4 na plataforma CKAN

Utilizaria esta plataforma no meu grupo de investigação

16 respostas

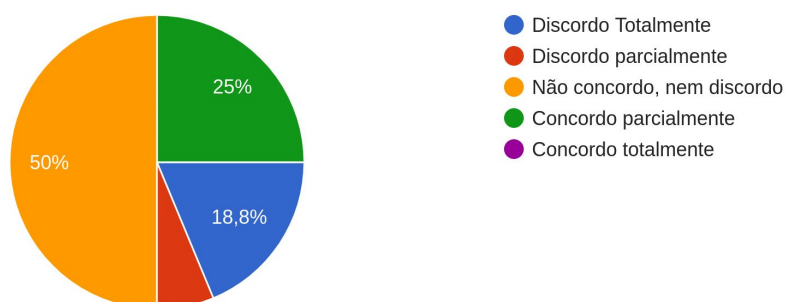


Figura 5.31: Resultados da pergunta QF5 na plataforma CKAN

A plataforma é capaz de satisfazer as necessidades de partilha de dados no meu grupo de investigação

16 respostas

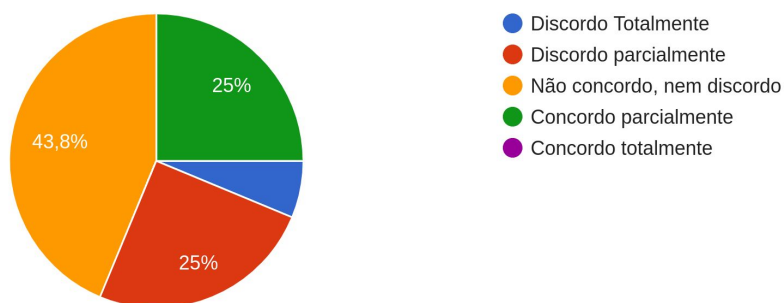


Figura 5.32: Resultados da pergunta QF6 na plataforma CKAN

A plataforma é capaz de satisfazer as necessidades de citação de dados no meu grupo de investigação

16 respostas

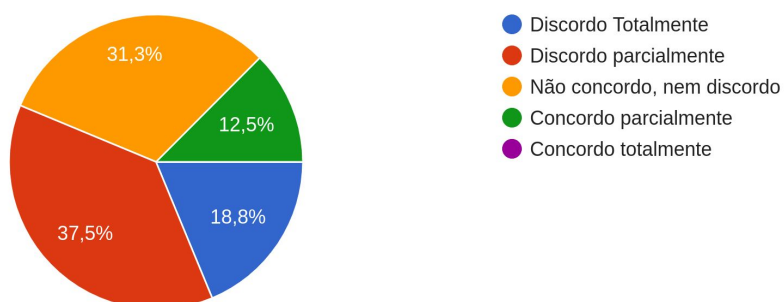


Figura 5.33: Resultados da pergunta QF7 na plataforma CKAN

Tabela 5.9: Desempenho das plataformas

Característica / Funcionalidade	DendroShare	Dataverse	CKAN
Desempenho Geral	BD	BD	MD
Citação de dados	BD	BD	FD
Interface da plataforma	BD	FD	BD
Pesquisa de depósitos por descritores específicos	BD	FD	FD

5.7 Conclusões do desempenho das 3 plataformas e o grau de satisfação dos utilizadores

Nesta secção serão apresentadas as conclusões retiradas em função das respostas dadas pelos utilizadores ao questionário pós-experiência e comparados os tempos de realização das tarefas em cada plataforma.

Começando pelo desempenho obtido pelos utilizadores em cada uma das plataformas, os utilizadores conseguiram desempenhos de tempos idênticos entre as plataformas nas diferentes tarefas. No entanto, na TAREFA6, o CKAN apresenta um desempenho bastante pior em relação às outras plataformas, pois os utilizadores não conseguiram obter o *snippet* de citação do conjunto de dados pretendido, embora tenham gasto praticamente o mesmo tempo de execução da tarefa nas outras plataformas. Passando agora para os resultados obtidos no questionário de pós-experiência, apesar de o guião para a execução das tarefas identificar convenientemente onde os utilizadores devem clicar ou inserir determinados valores em caixas de texto específicas, notou-se uma grande dificuldade por parte dos utilizadores a encontrarem o mencionado no guião na plataforma Dataverse. Esta apresenta demasiada informação para o utilizador, sendo que muita desta informação não é relevante ao interesse do utilizador. Para além disto, os utilizadores ficaram bastante aliados com a interface da plataforma CKAN, pois apresentava apenas a informação necessária para o utilizador. Por outro lado, os utilizadores ficaram bastantes descontentes pela plataforma não possuir a geração do *snippet* de citação do depósito de dados de forma automática. Na plataforma DendroShare, os utilizadores ficaram bastante impressionados com a possibilidade de pesquisa de depósitos pelos descritores específicos do domínio que estes guardam no seu conjunto de metadados. Os utilizadores também mencionaram que a plataforma apresentava uma interface *user-friendly* e satisfazia todas as necessidades de partilha e citação de dados de investigação.

Para facilitar a representação dos resultados do desempenho das 3 plataformas foi realizada uma tabela, representada pela Tab 5.9, onde é especificado as características que cada uma das plataforma obteve **Bom Desempenho (BD)**, **Médio Desempenho (MD)** ou **Fraco Desempenho (FD)**.

Capítulo 6

Conclusões e trabalho futuro

Este capítulo abordará um breve balanço do trabalho realizado, e em suma, apresentará possíveis futuras melhorias na plataforma.

6.1 Publicações resultantes deste trabalho

As principais contribuições resultante deste trabalho foram

- i. Um artigo com a descrição e detalhes de implementação das funcionalidades implementadas na plataforma DendroShare. Este artigo foi submetido e aceite para publicação na conferência TPD2019 a realizar-se em Setembro de 2019: Costa, L. & Rocha, J. (2019), "Dendro: a FAIR, open-source data sharing platform": 23rd International Conference on Theory and Practice of Digital Libraries (TPDL 2019), Oslo, Norway, September 9 - 12, 2019, Proceedings. (Aceite para publicação)
- ii. Uma Demo com a demonstração das funcionalidades da plataforma DendroShare. Esta Demo foi submetida e aceite para apresentação na conferência OSFAIR2019 a realizar-se em Setembro de 2019: Costa, L. & Rocha, J. Dendro: a FAIR data management and repository software (2019), "Dendro: a FAIR data management and repository software": Open Science FAIR, Porto, Portugal, September 16-18, 2018. (Demo) (Disponível em <https://www.opensciencefair.eu/programme-2>)

6.2 Balanço do trabalho realizado

Durante o período de desenvolvimento deste projeto foram cumpridos todos os objetivos propostos. O processo de desenvolvimento foi dividido em duas fases: uma fase mais focada na pesquisa de informação e recolha de soluções para o problema e outra fase, focada no levantamento de requisitos, implementação da solução e na experiência de avaliação da solução implementada.

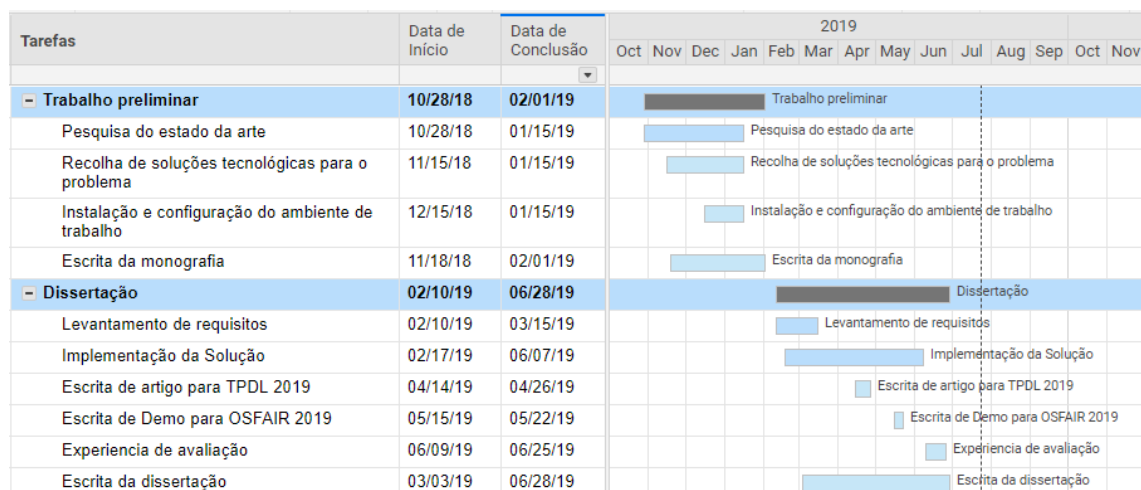


Figura 6.1: Diagrama de Gantt

No diagrama da Fig 6.1 estão representadas as tarefas realizadas, assim como a calendarização do tempo despendido em cada tarefa.

Numa primeira fase, foram realizadas pesquisas para a introdução aos vários tópicos em que o projeto é focado. Nesta fase, foram abordados os tópicos dos objetivos e recomendações da *Open Science*, iniciativas e projetos para promover a *Open Science*, a citação e partilha de publicações e dados de investigação. Por último, foram analisadas algumas plataformas cuja finalidade vai de encontro ao pretendido com este projeto.

Numa segunda fase, foi realizada uma recolha de soluções tecnológicas para o problema, com a análise às funcionalidades de valor em cada uma das plataformas existentes com características semelhantes às pretendidas e a realização de uma planificação de novas funcionalidades, para permitir a partilha e citação de dados acabados no Dendro.

Numa terceira fase, foi realizada uma instalação e configuração do ambiente de trabalho e desenvolvimento do Dendro.

Numa quarta fase foi realizado o levantamento de requisitos para a realização do projeto, que envolveu o planeamento das funcionalidades da plataforma Dendro, a realização dos mockups de interface e o fluxo de navegação entre as páginas existentes.

Numa quinta fase, foi realizada a implementação da solução e escrita de testes para a validação das funcionalidades implementadas.

Por último, foi realizado a experiência de avaliação da eficiência da abordagem realizada. A experiência de avaliação contou com um conjunto de utilizadores para a realização de tarefas similares em 3 plataformas distintas: Dendro, Dataverse e CKAN.

6.2.1 Conclusões do trabalho realizado

Os objetivos deste trabalho eram os objetivos representados na árvore de objetivos da Fig 1.1. Dos objetivos propostos para este projeto apenas não foi concluído o objetivo 1.2.2.2, em que era necessário assegurar a qualidade dos descritores preenchidos. Para demonstrar melhor os objetivos

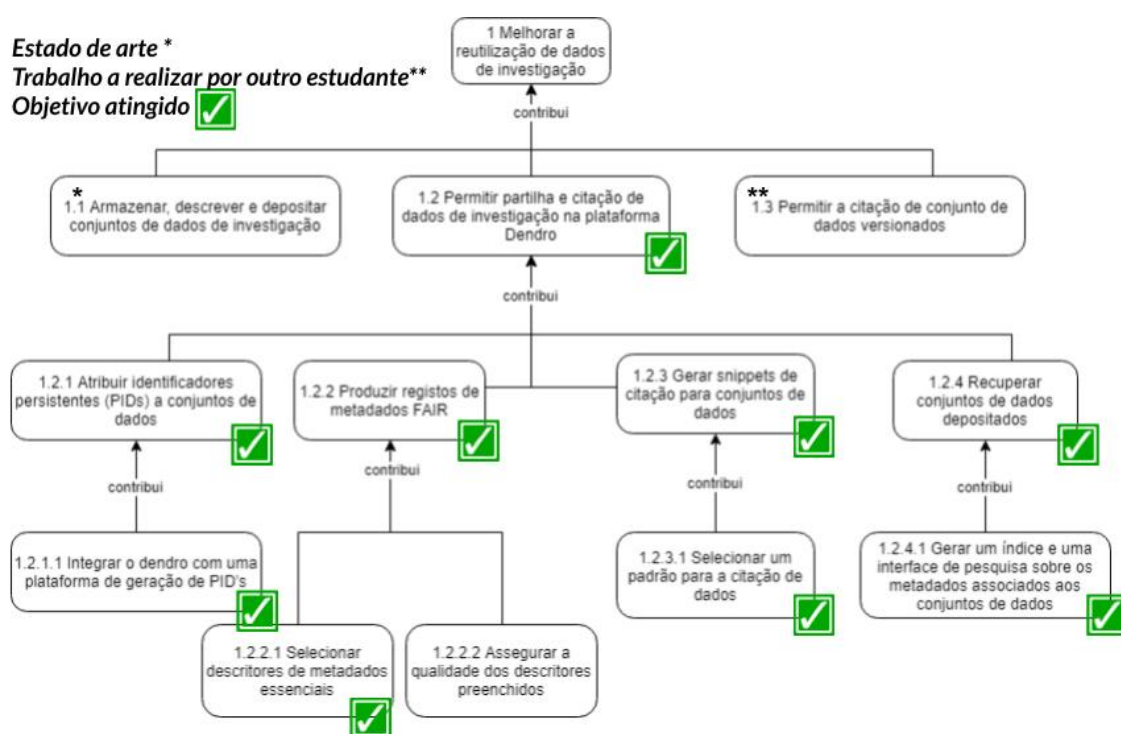


Figura 6.2: Balanço dos objetivos alcançados

atingidos, o estado da arte e o trabalho a realizar por outro estudante foi reformulada a árvore de objetivos representada na figura anterior, onde pode ser consultada na Fig 6.2.

A plataforma Dendro antes da realização desta extensão focava-se em apoiar as operações de recolha, descrição, análise e novas recolhas, para construir e amadurecer os dados para publicação noutra plataforma.

Com a implementação desta extensão de partilha e citação de dados acabados, a plataforma Dendro aproximou-se das características dos princípios FAIR, pois os dados são armazenados e identificados por identificadores únicos, são acessíveis através dos resultados da pesquisa, tem a capacidade de ser interoperável com outras plataformas e os dados podem ser reutilizáveis através das citações.

A extensão DendroShare possibilitou à plataforma Dendro ser capaz de descrever, armazenar e permitir a reutilização dos dados de investigação através da criação da citação de dados. Ao contrário das restantes plataformas de partilha e citação de dados, no Dendro é possível inserir metadados específicos do domínio nos dados inseridos na plataforma e por sua vez realizar a pesquisa facetada nos respetivos metadados associados aos dados.

Com a implementação desta extensão na plataforma é possível atribuir crédito aos autores dos dados, reduzir o esforço da reprodução de dados duplicados e aumentar a fiabilidade dos dados através da facilidade de reprodução dos resultados obtidos nas investigações por parte dos investigadores.

6.3 Trabalho futuro

O Dendro é uma plataforma que se encontra em desenvolvimento e tem como objetivo adicionar várias extensões, não apenas na parte do depósito, armazenamento e partilha de dados, mas em toda a plataforma.

Perante a extensão implementada, uma melhoria que vai em conta ao trabalho realizado é permitir o versionamento de depósitos e a citação das várias versões, um ponto que já tinha sido abordado na árvore de objetivos da Fig 1.1.

Para além disso, dentro de um projeto no Dendro, existe a função de cortar um ficheiro de uma localização do projeto e colar o ficheiro noutra localização no mesmo projeto. De futuro, pretende-se implementar uma funcionalidade de copiar-colar (ou seja, duplicação) de ficheiros e pastas. Para tal, bastará reutilizar o código já implementado para a cópia do conteúdo de uma pasta de um projeto para a pasta raiz de um novo depósito aquando de uma exportação.

Por último, outra melhoria passará por reduzir o tempo de resposta de um resultado de pesquisa. No casos em que os utilizadores inserem poucos valores aos descritores, a resposta é bastante rápida. Contudo, com o aumento do número de valores dos descritores, este tempo de resposta aumenta.

Atualmente, o DendroShare interroga sempre a base de dados para recuperar toda a informação dos depósitos que respeitem as restrições assim que um valor do descritor da pesquisa é alterado—um processo que se torna dispendioso.

Uma possível solução passará por realizar uma primeira *query* para obter apenas o Uri dos depósitos que respeitem as restrições, e numa segunda *query* para obter a informação dos depósitos. A informação dos depósitos seria então guardada em *cache* sempre que fossem realizados pedidos para obter toda a informação de um depósito. Ao fim de algumas interações com a base de dados, esta a *cache* estaria preenchida com informação de vários depósitos, incluindo o Uri. Deste modo, quando um valor de um descritor fosse alterado, seria realizado um pedido à base de dados para verificar os Uri's que respeitem as condições impostas, e no caso da informação desse Uri estar guardada em *cache*, a informação desse depósito era mostrada ao utilizador sem realizar um segundo pedido à base de dados para recuperar os dados dos depósitos com as Uris que correspondem aos critérios de pesquisa, reduzindo-se assim o tempo de resposta da pesquisa de depósitos.

ff

Referências

- [1] GO FAIR. <https://www.go-fair.org/>, Online, accessed 10-12-2018, 2018.
- [2] Liz Allen, Jo Scott, Amy Brand, Marjorie Hlava, e Micah Altman. Publishing: Credit where credit is due. *Nature*, 2014.
- [3] Micah Altman e Mercè Crosas. The Evolution of Data Citation: From Principles to Implementation. *IAssist quarterly*, 37, 2013.
- [4] R.C. Amorim, J.A. Castro, J. Rocha da Silva, e C. Ribeiro. A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Universal Access in the Information Society*, 16(4), 2017.
- [5] Ricardo Carvalho Amorim, João Aguiar Castro, João Rocha da Silva, e Cristina Ribeiro. *A comparative study of platforms for research data management: Interoperability, metadata capabilities and integration potential*. Springer, 2015.
- [6] Keith Baggerly. Disclose all data in publications. *Nature*, 2010.
- [7] Alex BALL e Monica Duke. *How to Cite Datasets and Link to Publications*. Digital Curation Centre, 2015.
- [8] Theodora Bloom, Emma Ganley, e Margaret Winker. Data Access for the Open Access Literature: PLOS's Data Policy. *PLoS Medicine*, 2014.
- [9] Christine L. Borgman. The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 2012.
- [10] Grant R. Brammer, Ralph W. Crosby, Suzanne J. Matthews, e Tiffani L. Williams. Paper Mâché: Creating dynamic reproducible science. *Procedia Computer Science*, 4:658–667, 2011.
- [11] Donatella Castelli, Paolo Manghi, e Costantino Thanos. A vision towards Scientific Communication Infrastructures: On bridging the realms of Research Digital Libraries and Scientific Data Centers. *International Journal on Digital Libraries*, 2013.
- [12] European Commission. H2020 Programme - Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in in Horizon 2020. Relatório técnico, European Commission, 2017.
- [13] Mark J. Costello. Motivating Online Publication of Data. *BioScience*, 2009.
- [14] Terence Critchlow. *Data-Intensive Science*. Chapman & Hall/CRC, 1 st edição, 2013.

- [15] João Rocha da Silva, Cristina Ribeiro, e João Correia Lopes. Ontology-based multi-domain metadata for research data management using triple stores. Em *Proceedings of the 18th International Database Engineering & Applications Symposium on - IDEAS '14*, 2014.
- [16] João Rocha da Silva, Cristina Ribeiro, e João Correia Lopes, da Silva, João Rocha and Ribeiro, Cristina and Lopes, João Correia. Ranking Dublin Core descriptor lists from user interactions: a case study with Dublin Core Terms using the Dendro platform, 2018.
- [17] Nicola Ferro. Reproducibility Challenges in Information Retrieval Evaluation. *Journal of Data and Information Quality*, 2017.
- [18] DataCite Metadata Working Group. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. *Datacite*. http://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel_v4.0.pdf, 2016.
- [19] Sarah Higgins. The dcc curation lifecycle model. *International journal of digital curation*, 3(1), 2008.
- [20] Sarah Higgins e Standards Advisor. The International Journal of Digital Curation The DCC Curation Lifecycle Model. Relatório técnico, Digital Curation Centre, University of Edinburgh, 2008.
- [21] Hans-Werner Hilse e Jochen Kothe. *Implementing Persistent Identifiers. Overview of concepts, guidelines and recommendations*. Libraries, Consortium of European Research, London, 2006.
- [22] Simon Hodson, Sarah Jones, Sandra Collins, Françoise Genova, Natalie Harrower, Leif Laaksonen, Daniel Mietchen, Rūta Petrauskaite, e Peter Wittenburg. Interim report from the European Commission Expert Group on FAIR data. Em *Turning FAIR data into reality*, 2018.
- [23] Jyrki Ilva, Markku Antero Laitinen, e Jarmo Saarti. The costs of open and closed access: Using the finnish research output as an example. *LIBER Quarterly*, 2016.
- [24] Gary King. An introduction to the dataverse network as an infrastructure for data sharing. *Sociological Methods and Research*, 2007.
- [25] Michael J. Kurtz. *For Attribution – Developing Data Attribution and Citation Practices and Standards*. The National Academies Press, 2012.
- [26] Sabina Leonelli, Daniel Spichtinger, e Barbara Prainsack. Sticks and carrots: encouraging open science at its source. *Geo: Geography and Environment*, 2015.
- [27] Charles Lipson. *Cite right: a quick guide to citation styles—MLA, APA, Chicago, the sciences, professions, and more*. University of Chicago Press, 2011.
- [28] Liz Lyon. Dealing with Data: Roles, Rights, Responsibilities and Relationships Consultancy Report Document details Acknowledgement to contributors Acknowledgement to funders. Relatório técnico, UKOLN, University of Bath, 2007.
- [29] Hailey Mooney e Mark Newton. The Anatomy of a Data Citation: Discovery, Reuse, and Credit. *Journal of Librarianship and Scholarly Communication*, 2012.

- [30] Marcus R. Munafò, Brian A. Nosek, Dorothy V.M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric Jan Wagenmakers, Jennifer J. Ware, e John P.A. Ioannidis. A manifesto for reproducible science. *Nature Human Behaviour*, 2017.
- [31] Jakob Nielsen. *Designing Web Usability: The Practice of Simplicity*. New Riders Publishing, Thousand Oaks, CA, USA, 1999.
- [32] Andrew M Odlyzko. Condensed version to appear in Notices Amer. Relatório técnico, AT&T Bell Laboratories, 1994.
- [33] Task Group on Data Citation Standards e CODATA-ICSTI Practices. Out of cite, out of mind: The current state of practice, policy, and technology for the citation of data. *Data Science Journal*, 12:CIDCR1–CIDCR75, 2013.
- [34] Norman Paskin. Digital Object Identifiers for scientific data. *Data Science Journal*, 2005.
- [35] Richard Pears e Graham J Shields. *Cite them right: the essential referencing guide*. Palgrave Macmillan, 2016.
- [36] Heather A. Piwowar, Roger S. Day, e Douglas B. Fridsma. Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2007.
- [37] Heather A. Piwowar e Todd J. Vision. Data reuse and the open data citation advantage. *PeerJ*, 2013.
- [38] Nancy Pontika, Petr Knoth, Matteo Cancellieri, e Samuel Pearce. Fostering open science to research using a taxonomy and an eLearning portal. Em *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business - i-KNOW '15*, 2015.
- [39] Cristina Ribeiro e João Rocha da Silva, Ribeiro, Cristina and Silva, João Rocha da. EUDAT as a long-term repository for the University of Porto. <https://eudat.eu/communities/eudat-as-a-long-term-repository-for-the-university-of-porto>, Online, accessed 31-01-2019, 2006.
- [40] J. Rocha, C. Ribeiro, e J. Correia Lopes. Ontology-based multi-domain metadata for research data management using triple stores. Em *Proceedings of the 18th International Database Engineering & Applications Symposium*, 2014.
- [41] Álvaro Rocha, Ana Maria Correia, Hojjat Adeli, Luís Paulo Reis, e Marcelo Mendonça Teixeira. New advances in information systems and technologies. *Advances in Intelligent Systems and Computing*, 445, 2016.
- [42] Tony Ross-Hellauer, Arvid Deppe, e Birgit Schmidt. Survey on open peer review: Attitudes and experience amongst editors, authors and reviewers. *PLoS ONE*, 2017.
- [43] Gianmaria Silvello. Theory and practice of data display. *CoRR*, abs/1706.0, 2017.
- [44] Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, e Mike Frame. Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 2011.

Anexo A

Guião da experiência

A.1 Descrição da experiência

A plataforma Dendro, desenvolvida no InfoLab da FEUP, tem como principal objetivo suportar a gestão de documentos e dados de investigação de forma colaborativa. A preparação para o seu depósito apresenta-se num repositório de dados. O DendroShare, que é o Dendro com a implementação de uma extensão, tem como objetivo possibilitar o depósito e catalogação de conjunto de dados de investigação na própria plataforma, deixando de ser obrigatória a existência de um repositório para os conjuntos de dados acabados. De forma a avaliar a eficiência da plataforma DendroShare com as plataformas CKAN e Dataverse, serão realizadas um conjunto de tarefas nas 3 plataformas, onde será avaliado a eficiência do armazenamento de ficheiros, a descrição de ficheiros ou datasets, a pesquisa de datasets na plataforma e a obtenção da citação de conjuntos de dados de investigação existentes nessa plataforma. Cada tarefa seguirá o workflow típico de cada plataforma.

A.2 Questionário pré-experiência

https://docs.google.com/forms/d/e/1FAIpQLSe_6c59uWpVDCToW0K66w_q3YCGoXH55_yPH7_JU8PM6vjZ-A/viewform

A.3 Tarefas de Avaliação

Em seguida serão descritas as tarefas de avaliação a realizar pelo utilizar, em cada uma das plataformas

Tarefa 1.A - Criar um Projeto no DendroShare

- 1 - Assim que entrar na página inicial do DendroShare, clique no item "My" localizado na parte superior da página e clique em "Projects". Irá entrar na página dos seus projetos.

2 - Para criar um novo projeto clique no botão "+NEW"

3 - Insira os seguintes valores nos respetivos descritores:

Handle: urbansenseXY. Onde X e Y é o dia e o mês da sua data de nascimento.

Title: Urban Sense XY. Onde X e Y é o dia e o mês da sua data de nascimento.

Description: Estes dados são resultados do projeto UrbanSense, pertencem ao domínio ambiental.

4 - Após o preenchimentos destes metadados iniciais, clique no botão "Create", para criar o projeto.

Tarefa 1.B - Criar um Dataset no Dataverse

1 - Assim que entrar na página inicial do Dataverse, crie um novo Dataset. Para isso, clique no botão "Add Data" e escolha a opção "New Dataset".

2 - Insira os seguintes valores nos respetivos descritores.

Title: Urban Sense

Description: Estes dados são resultados do projeto UrbanSense, pertencem ao domínio ambiental.

Subject: escolha a opção Computer and Information Science

3 - Após o preenchimentos destes metadados iniciais, clique no botão "Save Dataset", para criar o dataset.

Tarefa 1.C - Criar um Dataset no CKAN

1 - Assim que entrar na página inicial do CKAN, clique no item "Datasets" localizado na parte superior da página, onde poderá visualizar os Datasets da plataforma.

2 - Para criar um novo dataset clique no botão "+ Add Dataset".

3 - Insira os seguintes valores nos respetivos descritores.

Title: Urban Sense XY. Onde X e Y é o dia e o mês da sua data de nascimento.

Description: Estes dados são resultados do projeto UrbanSense, pertencem ao domínio ambiental.

4 - Após o preenchimentos destes metadados iniciais, clique no botão "Next: Add Data".

5 - Visto nesta plataforma ser necessário a realização de um upload na fase inicial, vamos só reservar um nome para o ficheiro que irá ser carregado para a plataforma na tarefa posterior. Insira no descritor **Name** um título à sua escolha.

6 - Clique no botão "Finish" para criar o dataset.

Tarefa 2.A - Upload de um ficheiro no DendroShare

Agora que criou um Projeto no DendroShare vamos realizar o upload de um ficheiro.

- 1 - Neste momento está localizado na página dos seus projetos, vamos entrar no projeto que criou. Clique no projeto que criou.
- 2 - Inicialmente vamos adicionar uma pasta ao projeto. Clique na seta ao lado do botão "Upload" e escolha a opção "Add folder".
- 3 - Insira "USense_FEUP" como nome da pasta e clique em "OK"
- 4 - Selecione a pasta que criou e clique no botão "Upload".
- 5 - Escolha um ficheiro disponível no seu ambiente de trabalho e faça o upload.

Tarefa 2.B - Upload de um ficheiro no Dataverse

Agora que criou um Dataset no Dataverse vamos realizar o upload de um ficheiro.

- 1 - Neste momento está localizado na página principal do Dataset que criou, vamos clicar no botão "Upload Files", para realizar o upload de um ficheiro.
- 2 - Clique no botão "Select Files to Add" e escolha um ficheiro disponível no seu ambiente de trabalho e faça o upload.
- 3 - Clique no botão "Save Changes" para guardar as alterações.

Tarefa 2.C - Upload de um ficheiro no CKAN

Agora que criou um Dataset no CKAN vamos realizar o upload de um ficheiro.

- 1 - Neste momento está localizado na página inicial do Dataset que criou, vamos clicar no título do ficheiro que escolheu na tarefa anterior.
- 2 - Clique no botão "Manage".
- 3 - Selecione o botão "Upload" e escolha um ficheiro disponível no seu ambiente de trabalho e faça o upload.
- 4 - Insira uma descrição ao ficheiro no caso de pretender.
- 5 - Clique no botão "Update Resource" para guardar as alterações.
- 6 - Só é possível fazer upload de ficheiros para o directório raiz do Dataset, por esse motivo, este Dataset não terá a mesma estrutura em árvore do Projeto criado no DendroShare. Clique no nome do dataset ("Urban Sense") situado na parte superior da página, para voltar à página principal do dataset.

Tarefa 3.A - Adicionar metadados ao Projeto no DendroShare

Encontra-se localizado dentro do projeto "Urban Sense", onde pode visualizar uma pasta no diretório raiz do projeto, vamos adicionar alguns metadados ao projeto.

- 1 - Para editar ou alterar os metadados do projeto ative o botão "EDIT MODE ON", visível na parte superior direita da página.
- 2 - Clique na pasta que criou.
- 3 - Selecione o botão "EDIT MODE ON".
- 4 - Na caixa de texto "Search for descriptor", localizada no lado direito da página, insira o valor `Methodology`.
- 5 - Selecione o descritor retornado com o valor "Methodology".
- 6 - Escreva na caixa de texto localizada no centro da página o seguinte texto `A Plataforma UrbanSense é composta por sensores, denominados Data Collection`
- 7 - Clique no botão "Save", para guardar as alterações efetuadas.

Tarefa 3.B - Adicionar metadados ao Dataset no Dataverse

Encontra-se localizado na página do Dataset "Urban Sense", vamos adicionar metadados ao Dataset.

- 1 - Clique no elemento "Metadata" da tab central da página, onde pode visualizar os metadados do dataset.
- 2 - Clique no botão "Add + Edit Metadata", situado no lado direito da página, para editar os metadados do Dataset.
- 3 - Insira os seguintes valores nos respetivos descritores.

Language: Escolha a opção English.

Production Date: Insira a data do dia de hoje.

Notes: Insira `Methodology:A Plataforma UrbanSense é composta por sensores, denominados Data Collection; Working State:Desativo; Relation: Artigos relacionados com a plataforma UrbanSense.`

- 4 - Clique no botão "Save Changes", para guardar as alterações realizadas.
- 5 - Por fim, clique na palavra "Demo Dataverse", que se encontra localizado na parte superior esquerda da página.

Tarefa 3.C - Adicionar metadados ao Dataset no CKAN

Encontra-se localizado na página do Dataset "Urban Sense", vamos adicionar alguns metadados ao Dataset.

- 1 - Clique no botão "Manage".
- 2 - Nesta plataforma apenas podemos inserir o título e a descrição a um ficheiro. Para inserirmos alguns metadados específicos do domínio vamos preencher os elementos do descritor **Custom Field** com os seguintes valores:

Key: Methodology **Value:** A Plataforma UrbanSense é composta por sensores, denominados Data Collection.

Key: Region **Value:** Porto

- 3 - Após o preenchimento destes valores, clique no botão "Update Dataset", para guardar as alterações realizadas.
- 4 - Por último, clique na palavra "Datasets" que se encontra localizada na parte superior esquerda da página, para voltar à lista de datasets da plataforma.

Tarefa 4.A - Exportar um depósito de dados acabados no DendroShare

Encontra-se localizado dentro do Projeto, vamos exportar um conjunto de dados acabados para o DendroShare.

- 1 - Para realizar esta operação necessita de ter a pasta selecionada, por isso caso não a tenha selecionado, seleccione-a.
- 2 - Clique no botão "Share" localizado no canto superior direito.
- 3 - Clique no botão "SELECT A DESTINATION REPOSITORY" e seleccione um elemento da lista.
- 4 - Insira o valor `Urban Sense` no título e insira uma descrição à escolha.
- 5 - Escolha a opção **Public** para a privacidade do depósito.
- 6 - Clique no botão "Send", para exportar o depósito para o DendroShare.
- 7 - Por último, navegue até à página inicial do DendroShare, ao clicar no ícone da plataforma que se encontra na parte superior esquerda da página.

Tarefa 5.A - Pesquisar um depósito no DendroShare

No lado direito são mostrados todos os depósitos da plataforma na forma de listagem paginada no lado esquerdo é apresentada uma interface de pesquisa avançada. Vamos procurar por um dataset na plataforma. O depósito tem algures na descrição a palavra "new", a privacidade é "Public" e é da região do "Porto".

- 1 - Insira o valor `new` no descritor "Description".
- 2 - Selecione apenas a checkbox "Public" no descritor "Privacy".
- 3 - Na caixa de texto "Search for descriptor" insira o valor `Region`, e selecione o valor retornado.
- 4 - Escreva `Porto` na nova caixa de texto localizada por baixo do valor inserido da alínea anterior.
- 5 - Como resultado da pesquisa irá aparecer um depósito que respeitará as condições impostas.

Tarefa 5.B - Pesquisar um depósito no Dataverse

Agora que se encontra na página inicial do Dataverse, pode visualizar todos os Datasets e Dataverses da plataforma na forma de lista paginada, vamos procurar por um dataset. O dataset tem algures na descrição a palavra "new", tem como valor do descritor "Producer" o valor "Comissão Europeia".

- 1 - Para realizar a pesquisa nesta plataforma, é necessário inserir valores na barra de pesquisa, valores estes que serão verificados em todos os descritores do depósito. Insira o valor `new` na barra de pesquisa.
- 2 - Obtivemos vários resultados, por isso insira `Comissão Europeia` na barra de pesquisa, em vez do valor inserido anteriormente.
- 3 - Obtivemos um depósito que satisfaz as condições impostas.

Tarefa 5.C - Pesquisar um depósito no CKAN

Nesta página encontram-se listados na forma de paginação todos os datasets da plataforma.

Vamos procurar por um dataset na plataforma. O depósito tem algures na descrição a palavra "new", o título é "New dataset in CKAN" e tem como contribuidor a "Comissão Europeia".

No entanto, apesar de termos inserido alguns descritores específicos do domínio no dataset que criamos, estes não fazem parte dos descritores de pesquisa. Os descritores que fazem parte da pesquisa são o título e a descrição de cada ficheiro e no dataset os descritores fixos da plataforma. Deste modo, para este depósito apenas podemos procurar pelos valores da descrição e do título.

- 1 - Insira o valor `New` na barra de pesquisa.

- 2 - Foram obtidos vários valores por isso, insira o valor `New dataset in CKAN` na caixa de pesquisa em vez do valor pesquisado anteriormente.
- 3 - Obtivemos um depósito que satisfaz as condições impostas.

Tarefa 6.A - Obter o snippet de citação de um depósito no DendroShare

Depois de ter encontrado o dataset pretendido, vamos obter o snippet de citação desse dataset.

- 1 - Clique no título do depósito que aparece sublinhado no resultado de pesquisa.
- 2 - Encontra-se localizado dentro do dataset, pode visualizar os dados e os metadados do mesmo. Selecione o snippet de citação do Dataset.
- 3 - Crie um ficheiro no ambiente de trabalho do seu computador, cole o snippet de citação do depósito e guarde o ficheiro.

Tarefa 6.B - Obter o snippet de citação de um depósito no Dataverse

Depois de ter encontrado o dataset pretendido, vamos obter o snippet de citação desse dataset.

- 1 - Clique no ícone do ficheiro localizado por baixo do título do dataset.
- 2 - Encontra-se localizado dentro do dataset, pode visualizar os dados e os metadados do mesmo. Clique no botão "Cite Dataset" localizado na parte superior direita da página e escolha a opção "BibTex", para fazer o download no snippet de citação para a sua área de transferência.

Tarefa 6.C - Obter o snippet de citação de um depósito no CKAN

Depois de ter obtido o dataset pretendido, vamos obter o snippet de citação desse dataset, caso seja possível.

- 1 - Clique no título do dataset que aparece no resultado de pesquisa.
- 2 - Encontra-se localizado dentro do dataset, pode visualizar os dados e os metadados do mesmo. No entanto não é possível obter o snippet de citação do dataset.
- 3 - Pode tentar concluir esta tarefa escrevendo um snippet de citação deste depósito ou simplesmente abandonar esta tarefa.

A.4 Questionário pós-experiência

https://docs.google.com/forms/d/e/1FAIpQLSfFd2M-BB2x9on-77oQfh_nGj5vFJ4-jNg97viewform

Anexo B

Questionários

B.1 Questionário de pré-experiência

Inquérito Inicial sobre armazenamento, descrição e partilha de dados de investigação

*Obrigatório

1. Estou familiarizado com o conceito de partilha e citação de dados de investigação *

Marcar apenas uma oval.

- Discordo totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

2. Estou a par dos benefícios da reutilização de dados de investigação *

Marcar apenas uma oval.

- Discordo totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

3. Regularmente descrevo e armazeno dados de investigação *

Marcar apenas uma oval.

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

4. Regularmente cito dados de investigação *

Marcar apenas uma oval.

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

5. Regularmente partilho dados de investigação produzidos por mim ou pelo meu grupo de investigação *

Marcar apenas uma oval.

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

6. Tenho dificuldades na criação da citação do conjunto de dados *

Marcar apenas uma oval.

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

7. Nos projetos nos quais participo existe a obrigatoriedade de partilha de dados de investigação em regime de acesso aberto *

Marcar apenas uma oval.

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

Com tecnologia



B.2 Questionário de pós-experiência

Inquérito Final sobre armazenamento, descrição e partilha de dados de investigação

Este inquerito tem como objetivo obter feedback sobre a utilização das plataformas Dendro, CKAN e Dataverse para gestão de dados de investigação.

É composto por 3 partes, cada uma relativa a uma plataforma.

Responda às questões tendo em conta as experiências que acabou de realizar.

***Obrigatório**

Tarefas realizadas na plataforma DendroShare

1. Tive dificuldade na realização das tarefas *

Marcar apenas uma oval.

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

2. Consegui concluir as tarefas propostas *

Marcar apenas uma oval.

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

3. Tive dificuldade na criação da citação do conjunto de dados *

Marcar apenas uma oval.

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

4. A plataforma usada tem uma interface user-friendly **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

5. Utilizaria esta plataforma no meu grupo de investigação **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

6. A plataforma é capaz de satisfazer as necessidades de partilha de dados no meu grupo de investigação **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

7. A plataforma é capaz de satisfazer as necessidades de citação de dados no meu grupo de investigação **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

8. Caso queira adicionar mais alguma informação relativamente à plataforma, indique-a neste espaço abaixo.

Tarefas realizadas na plataforma Dataverse

9. Tive dificuldade na realização das tarefas **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

10. Consegui concluir as tarefas propostas **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

11. Tive dificuldade na criação da citação do conjunto de dados **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

12. A plataforma usada tem uma interface user-friendly **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

13. Utilizaria esta plataforma no meu grupo de investigação **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

14. A plataforma é capaz de satisfazer as necessidades de partilha de dados no meu grupo de investigação **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

15. A plataforma é capaz de satisfazer as necessidades de citação de dados no meu grupo de investigação **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

16. Caso queira adicionar mais alguma informação relativamente à plataforma, indique-a neste espaço abaixo.

Tarefas realizadas na plataforma CKAN**17. Tive dificuldade na realização das tarefas ****Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

18. Consegui concluir as tarefas propostas **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

19. Tive dificuldade na criação da citação do conjunto de dados **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

20. A plataforma usada tem uma interface user-friendly **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

21. Utilizaria esta plataforma no meu grupo de investigação **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

22. A plataforma é capaz de satisfazer as necessidades de partilha de dados no meu grupo de investigação **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

23. A plataforma é capaz de satisfazer as necessidades de citação de dados no meu grupo de investigação **Marcar apenas uma oval.*

- Discordo Totalmente
- Discordo parcialmente
- Não concordo, nem discordo
- Concordo parcialmente
- Concordo totalmente

24. Caso queira adicionar mais alguma informação relativamente à plataforma, indique-a neste espaço abaixo.



Anexo C

Formulário de Avaliação

Avaliação das tarefas realizadas pelos utilizadores nas Plataformas DendroShare, Dataverse e CKAN

*Obrigatório

1. Ocupação do utilizador *

2. Sexo *

Marcar apenas uma oval.

Masculino

Feminino

3. Idade *

Tarefa 1 - Criar um projeto ou um dataset na plataforma

Tarefa 1.A (Criar um Projecto no DendroShare)

4. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

5. Número de ajudas? *

6. Número de erros? *

7. Tempo gasto na experiência? *

Tarefa 1.B (Criar um Dataset no Dataverse)

8. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

9. Número de ajudas? *

10. Número de erros? *

11. Tempo gasto na experiência? *

Tarefa 1.C (Criar um Projecto no CKAN)

12. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

13. Número de ajudas? *

14. Número de erros? *

15. Tempo gasto na experiência? *

Tarefa 2 - Fazer upload de ficheiros na plataforma

Tarefa 2.A (Upload de ficheiros no DendroShare)

16. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

17. Número de ajudas? *

18. Número de erros? *

19. Tempo gasto na experiência? *

Tarefa 2.B (Upload de ficheiros no Dataverse)

20. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

21. Número de ajudas? *

22. Número de erros? *

23. Tempo gasto na experiência? *

Tarefa 2.C (Upload de ficheiros no CKAN)

24. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

25. Número de ajudas? *

26. Número de erros? *

27. Tempo gasto na experiência? *

Tarefa 3 - Adicionar mais alguns metadados ao Projeto ou Dataset criado

Tarefa 3.A (Adicionar metadados ao projeto no DendroShare)

28. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

29. Número de ajudas? *

30. Número de erros? *

31. Tempo gasto na experiência? *

Tarefa 3.B (Adicionar metadados ao dataset no Dataverse)

32. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

33. Número de ajudas? *

34. Número de erros? *

35. Tempo gasto na experiência? *

Tarefa 3.C (Adicionar metadados ao projeto no CKAN)

36. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

37. Número de ajudas? *

38. Número de erros? *

39. Tempo gasto na experiência? *

Tarefa 4 - Exportar um depósito de dados acabados no DendroShare

40. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

41. Número de ajudas? *

42. Número de erros? *

43. Tempo gasto na experiência? *

Tarefa 5 - Pesquisar um deposito na plataforma

Tarefa 5.A (Pesquisar um deposito no DendroShare)

44. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

45. Número de ajudas? *

46. Número de erros? *

47. Tempo gasto na experiência? *

Tarefa 5.B (Pesquisar um deposito no Dataverse)

48. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

49. Número de ajudas? *

50. Número de erros? *

51. Tempo gasto na experiência? *

Tarefa 5.C (Pesquisar um deposito no CKAN)

52. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

53. Número de ajudas? *

54. Número de erros? *

55. Tempo gasto na experiência? *

Tarefa 6 - Obter o snippet de citação de um deposito da plataforma

Tarefa 6.A (Obter o snippet de citação de um deposito no DendroShare)

56. Conclui a tarefa? *

Marcar apenas uma oval.

Sim

Não

57. Número de ajudas? *

58. Número de erros? *

59. Tempo gasto na experiência? *

Tarefa 6.B (Obter o snippet de citação de um deposito no Dataverse)

60. **Conclui a tarefa? ***

Marcar apenas uma oval.

Sim

Não

61. **Número de ajudas? ***

62. **Número de erros? ***

63. **Tempo gasto na experiência? ***

Tarefa 6.C (Obter o snippet de citação de um depósito no CKAN)

64. **Conclui a tarefa? ***

Marcar apenas uma oval.

Sim

Não

65. **Número de ajudas? ***

66. **Número de erros? ***

67. **Tempo gasto na experiência? ***

Com tecnologia

