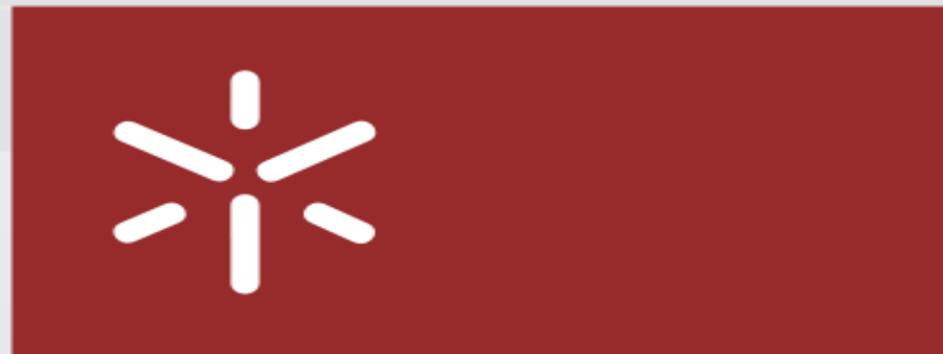


4º FÓRUM
GESTÃO DE DADOS
DE INVESTIGAÇÃO

16 DE NOVEMBRO

Pedro Príncipe e Eloy Rodrigues

pedroprincipe@sdum.uminho.pt



Universidade do Minho
Serviços de Documentação

Data RepositóriUM

projeto de implementação do repositório
de dados para a Universidade do Minho





Os Serviços de Documentação da Universidade do Minho têm previsto no seu plano de atividades para 2018 a **criação de um repositório de dados**.

Esta iniciativa decorre do seu plano de intervenção para apoiar a gestão de dados e enquadra-se na **estratégia institucional** em discussão para a **GDI na UMinho**.

Data RepositóriUM

[projeto de implementação Maio 2018-Fevereiro 2019]





Universidade do Minho

ESTRATÉGIA INSTITUCIONAL PARA A
GESTÃO DOS DADOS DE INVESTIGAÇÃO:
ESTUDO E RECOMENDAÇÕES

Relatório do grupo de trabalho constituído por:

Rui L. Reis, Vice-reitor (coordenador), Eloy Rodrigues (Serviços de Documentação), Eugénio Ferreira (Centro de Engenharia Biológica), Joana Palha (Instituto de Ciências da Vida e da Saúde), José Manuel Meijome (Centro de Física), Manuela Gomes (Grupo de Investigação 3B's), Miguel Gonçalves (Centro de Investigação em Psicologia), Moisés Martins (Centro de Estudos de Comunicação e Sociedade), Natália Monteiro (Núcleo de Investigação em Políticas Económicas), Paulo Novais (Centro Algorítmico), Pedro Príncipe (Serviços de Documentação).

Setembro de 2017



Universidade do Minho
Serviços de Documentação

A Universidade do Minho tem em curso a definição da estratégia institucional para a gestão de dados de investigação que inclui:

- formulação de uma **política** institucional,
- planificação de **infraestrutura** e sistemas para dados,
- produção de **serviços e ferramentas** de apoio à gestão dos dados,
- desenvolvimento de programas de **capacitação**
- reforço de serviços de apoio e consultoria em **questões legais e de proteção de dados.**





Universidade do Minho

ESTRATÉGIA INSTITUCIONAL PARA A GESTÃO DOS DADOS DE INVESTIGAÇÃO: ESTUDO E RECOMENDAÇÕES

Relatório do grupo de trabalho constituído por:

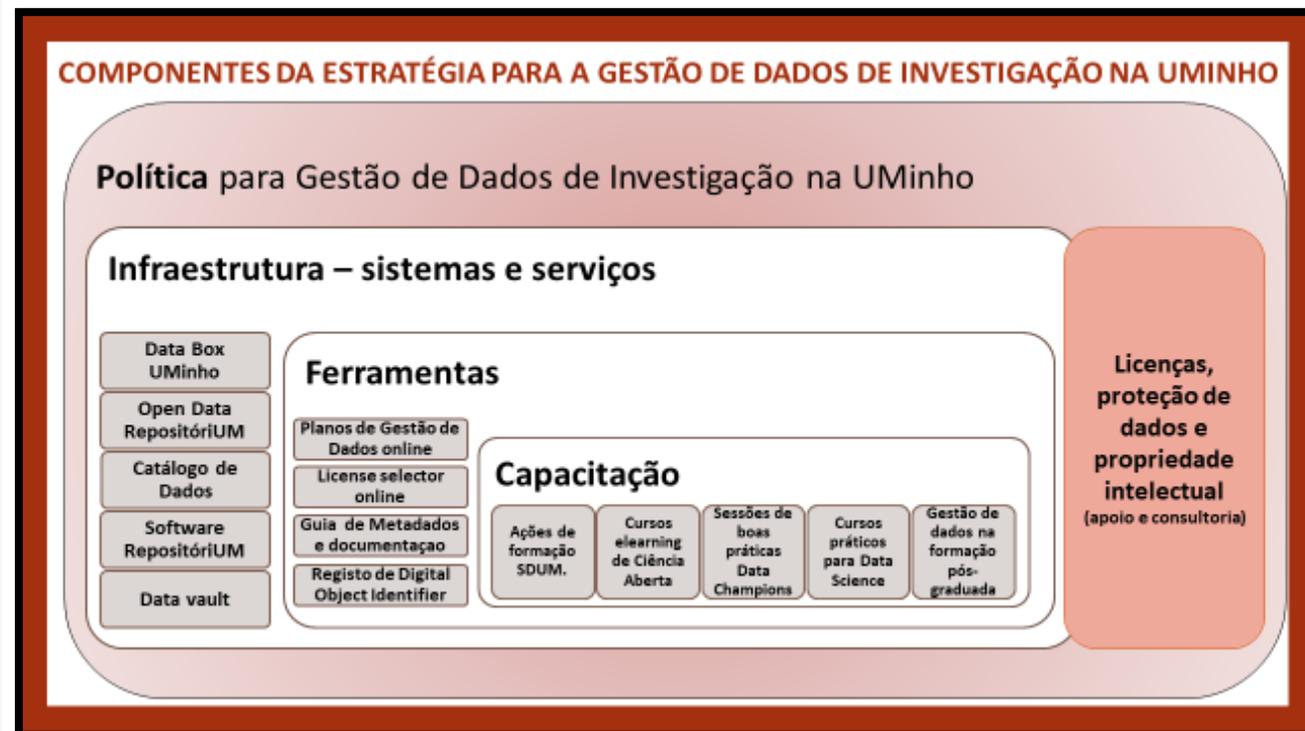
Rui L. Reis, Vice-reitor (coordenador), Eloy Rodrigues (Serviços de Documentação), Eugénio Ferreira (Centro de Engenharia Biológica), Joana Palha (Instituto de Ciências da Vida e da Saúde), José Manuel Meijome (Centro de Física), Manuela Gomes (Grupo de Investigação 3B's), Miguel Gonçalves (Centro de Investigação em Psicologia), Moisés Martins (Centro de Estudos de Comunicação e Sociedade), Natália Monteiro (Núcleo de Investigação em Políticas Económicas), Paulo Novais (Centro Algoritmi), Pedro Príncipe (Serviços de Documentação).

Setembro de 2017



Universidade do Minho
Serviços de Documentação

Disponibilização de uma infraestrutura de **sistemas e serviços integrados** a disponibilizar à comunidade da UMinho para a gestão de dados ao longo do ciclo de vida da investigação:





Data RepositóriUM: porquê?

O Repositório de Dados não resolverá todas as necessidades identificadas na UMinho...

- Servirá de *desbloqueador* a muitas limitações sentidas hoje pelos investigadores no trabalho de curadoria de dados,
- Será uma ferramenta de utilização prática de auto-curadoria com registo de datasets e depósito de ficheiros que contribuirá para geração de boas-práticas na gestão e partilha de dados.





Data RepositóriUM: para quê?

A criação do repositório de dados irá permitir ganhos imediatos para a comunidade de investigadores da UMinho, nomeadamente:

- Registo de identificadores persistentes (**DOI**);
- Viabilizar a **citação** de dados e o **crédito** aos produtores de dados;
- Auto **depósito** de ficheiros por datasets e em coleções organizadas por UOEI;
- **Preservação** dos ficheiros e datasets;
- Apoiar o **cumprimento dos requisitos** Open Data da Comissão Europeia (H2020);
- Auxiliar os autores na disponibilização de dados referidos em artigos (**editores**);
- Permitirem **acesso controlado** a dados e sob identificação prévia dos utilizadores;
- Associar **licenças** de uso dos datasets e explicitar os termos de uso dos ficheiros;
- Documentar os datasets com informação descritiva e administrativa adequada;
- Facilitar a **disseminação** dos resultados de investigação.





Sistema de Repositório de dados

Solução:

- Recurso a software de código aberto, no quadro de uma infraestrutura flexível e interoperável, tendo como modelo as melhores práticas em instituições.
- Análise de várias soluções de plataformas de software para repositório de dados: *DSpace*, *CKAN*, *Invenio* e *Dataverse*.

Software	Organismo responsável	Nº de Instalações	Tipo de Repositório
DSpace https://duraspace.org/dspace/	DuraSpace	1767	Híbrido (Dados e publicações)
INVENIO https://invenio-software.org/	CERN	3 (Invenio V3) 40 (Invenio V1/2)	Híbrido (Dados e publicações)
CKAN https://ckan.org/	Open Knowledge Foundation	146	Dados Abertos
Dataverse project https://dataverse.org	IQSS - Harvard's Institute for Quantitative Social Science	32	Dados de investigação

- Opção final: **software *Dataverse***, plataforma desenvolvida e em utilização pela Univ. de Harvard. Solução open source de aplicação web para partilha, preservação, citação, exploração e análise de dados de investigação.





The screenshot shows the homepage of the Dataverse Project. At the top, there is a navigation menu with the following items: "Dataverse Project" (with a logo), "About", "Community", "Best Practices", "Software", and "Contact". The main heading reads "The Dataverse Project" with a logo consisting of three orange circles connected by lines. Below this, the text says "Open source research data repository software". There are two main content blocks: one for "Researchers" with a green icon of two eyes and text describing full control over data and increased visibility; and one for "Journals" with a blue icon of an open book and text describing seamless management of data associated with published articles.

<http://dataverse.org/>



10 razões para a escolha da plataforma de software Dataverse:



Universidade do Minho
Serviços de Documentação

- 1. Versatilidade na organização de coleções (dataverses)**, necessária para abarcar de forma explícita todas as Unidades Operacionais e de Investigação.
- 2. Polivalência na adaptação dos esquemas de metadados** disciplinares específicos para cada dataverse, de uma lista de 6 esquemas já configurados.
- 3. Flexibilidade na definição de papéis e responsabilidades** nos workflows de curadoria de dados e na administração das coleções (8 papéis).
- 4. Autonomia na operação dos processos de curadoria** e na utilização do sistema, podendo o utilizador ser criador de dataverses e datasets, depositar...
- 5. Escalabilidade e potencial de integração com outros sistemas**, existindo ferramentas integradas para análise de dados, georreferenciação...



10 razões para a escolha da plataforma de software Dataverse:



Universidade do Minho
Serviços de Documentação

6. **Multiplicidade de opções na disponibilização e acesso aos dados**, desde dados abertos, restritos com metadados públicos, *guest book*...
7. **Versatilidade nas opções de serviço de armazenamento**, sendo possível usar diferentes modelos ou sistemas associados ao repositório.
8. **Agregação de dados de fontes externas**, configurável por OAI-PMH e de forma ágil em ambiente de *dashboard*.
9. **Comunidade de instituições** de investigação com instalações dataverse de grande relevância, desde logo os autores do sistema, a Harvard University.
10. **Transparência nos desenvolvimentos futuros do software**, com a apresentação pública e periódica do *roadmap* de atualizações da plataforma.



Data RepositóriUM

[projeto de implementação Maio 2018-Fevereiro 2019]



Universidade do Minho
Serviços de Documentação

PLANIFICAÇÃO - TAREFAS	Maio	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.	Jan.
Instalação dos pré-requisitos									
Instalação do software									
Definição de especificação de storage									
Tarefas de segurança da instalação									
Instalação de sistemas complementares									
Autenticação no sistema (LDAP)									
Conceção da imagem									
Configuração de layout									
Criação da estrutura de dataverses									
Análise dos esquemas de metadados									
Definição de modelos de termos de uso									
Teste do Sistema (dataverse SDUM)									
Definição das tarefas de curadoria									
Formação de curadores SDUM									
Elaboração do guia de utilização									
Integração de comunidades pilotos									
Formadores curadores e comunidades									
Suporte às comunidades piloto									
Relatório final									
Apresentação pública									

The screenshot displays the Dataverse web interface. At the top, there is a navigation bar with the Dataverse logo, a search icon, and links for 'User Guide', 'Support', 'Sign Up', and 'Log In'. Below the navigation bar, the page title is 'UMinho (Universidade do Minho) Repositório de Dados da Universidade do Minho'. A sidebar on the left shows filters for 'Metrics' (1 Download), 'Contact', and 'Share'. The main content area features a search bar with the text 'sdum' and a 'Find' button. Below the search bar, there are three search results. The first result is 'sdum (Serviços de Documentação da Universidade do Minho)' with a publication date of 'Jun 27, 2018'. The second result is 'Test 2' with a publication date of 'Jun 27, 2018' and a description 'José Carvalho, 2018, "Test 2", https://doi.org/10.5072/FK2JZ7REVY, UMinho, V1'. The third result is 'Resultados do questionários SDUM 2017' with a publication date of 'Jun 27, 2018' and a description 'Príncipe, Pedro, 2018, "Resultados do questionários SDUM 2017", https://doi.org/10.5072/FK2JZ7REVY, UMinho, V1'. The footer of the page includes 'Copyright © 2018' and 'Powered by The Dataverse Project v. 4.9 build 882-d9451c5'.



Data RepositóriUM

[projeto de implementação Maio 2018-Fevereiro 2019]

Pedro Príncipe

pedroprincipe@sdum.uminho.pt

Obrigado!

