



Universidade do Minho  
Serviço de Documentação e Bibliotecas

# UNIVERSITY OF MINHO DATA REPOSITORY

## features to publish & share data and why we chose the dataverse software



Pedro Príncipe | 25 January 2022 | reSEArch-EU Open Data Staff Week





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

<http://hdl.handle.net/1822/63930>



Universidade do Minho  
Serviço de Documentação e Bibliotecas

# DATA REPOSITORY IS A SERVICE INTEGRATED IN THE RDM STRATEGY DEFINED FOR UMINHO

## COMPONENTS OF THE STRATEGY FOR RESEARCH DATA MANAGEMENT AT UNIVERSITY OF MINHO (2018)

### Policies (OA, OS, RDM) at UMinho

#### Infrastructure - systems and services

- Data Box UMinho
- Open Data RepositóriUM
- Data Catalogue
- Software RepositóriUM
- Data vault

#### Tools

- Online Data Management Plans
- License selector
- Metadata and documentation
- Digital Object Identifier register

#### Training

- Library training actions (OS cafés)
- Open Science e-learning courses
- Data Champions best practices sessions
- Practical courses for Data Science
- RDM in post-graduate training

Licenses, data protection and intellectual property (support and consulting)

The screenshot displays the DataRepositórioUM website. At the top, there is a navigation bar with the logo and menu items: Adicionar Dados, Pesquisar, Guia de Utilizador, Suporte, Português, Criar Conta, and Iniciar Sessão. Below the navigation bar, there are statistics for 'Métricas' (1,056 Downloads) and buttons for 'Contacto' and 'Partilhar'. The main content area shows search results for 'Repositório de Dados da Universidade do Minho'. On the left, there are filters for 'Categorias do Dataserve', 'Ano de Publicação', 'Autor: Nome', 'Assunto', 'Palavra-chave: Termo', 'Data de Depósito', 'Produtor: Nome', 'Data de Produção', and 'Colaborador: Nome'. The search results list 11 items, each with a title, date, and a brief description. The bottom of the page features a 'SUPPORT' section with links to 'Guia sobre Gestão de Dados de Investigação', 'Normas Genes', and 'Contact-us', along with a copyright notice for 2021 and a logo for 'Desenvolvido por DataVerse'.



Universidade do Minho  
 Serviço de Documentação e Bibliotecas

# DataRepositórioUM

## Launched on February 17th, 2020

<https://datarepositorium.uminho.pt>



# Data RepositóriUM: why?

(...) how was it introduced?

The Data Repository will not solve all the needs identified in UMinho...

- It will unlock the many limitations felt today by researchers in **research data curation**.
- It will be a self-curated tool with dataset registration and file deposit that will contribute to the **generation of good practices in data management and sharing**.



# *How we spread the word...*



**data**

**RepositórioUM**



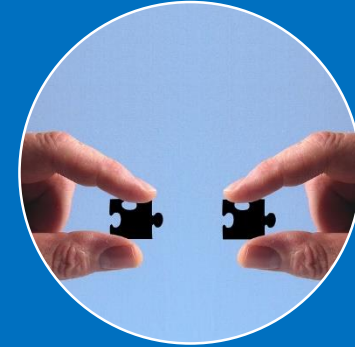
## SHARE

Promote data sharing and the practice of Open Science.



## PUBLISH

Make data available for reuse and generate greater impact.



## MANAGE

Ensure good documentation and data management practices





# Why the Dataverse software...

**Solution:** Use of open source software, within the framework of a flexible and interoperable infrastructure, with best practices in use around the world.

- Analysis of various solutions of software platforms for data repositories (2017/2018):
  - *DSpace, CKAN, Invenio e Dataverse.*

Software	Responsible Organization	No. of Installations	Repository Type
DSpace <a href="https://duraspace.org/dspace/">https://duraspace.org/dspace/</a>	DuraSpace	1767	Hybrid (Data and publications)
INVENIO <a href="https://invenio-software.org/">https://invenio-software.org/</a>	CERN	3 (Invenio V3) 40 (Invenio V1/2)	Hybrid (Data and publications)
CKAN <a href="https://ckan.org/">https://ckan.org/</a>	Open Knowledge Foundation	146	Open Data
Dataverse project <a href="https://dataverse.org">https://dataverse.org</a>	IQSS - Harvard's Institute for Quantitative Social Science	32	Research Data

- Final option: Dataverse software, a platform developed and in use at Harvard University. Open source web application solution for sharing, preserving, citing, exploring and analyzing research data.

# DATAVERSE

## considerations about the selection process



Universidade do Minho  
Serviço de Documentação e Bibliotecas

- ✓ The solution was a result of a brief study of the options on the market but very focused on **pragmatic preferences**.
- ✓ The preference for an open source solution has as main objective to ensure the **sustainability** of the service.
- ✓ The specific features & functionalities of the system were analyzed in line with the researchers **requirements** of previous diagnostic studies.
- ✓ The options are always made at a certain time, taking into consideration the **status of the service** development, uptake and features.
- ✓ In the analysis process there are details that make the difference, in this case the most relevant was the **community** of users around the world.



# DATAVERSE IS A REPOSITORY

Find, share, cite and publish data

# DATAVERSE IS A SOFTWARE

Repository software platform

# DATAVERSE IS A COMMUNITY

Promote data management and sharing worldwide

<https://dataverse.org/>

## Open source research data repository software



Researchers

Enjoy full control over your data. Receive *web visibility*, *academic credit*, and *increased citation counts*. A personal Dataverse collection is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. [Want to set up your personal Dataverse collection?](#)



Journals

Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data*. Participate in the open data movement by using a Dataverse collection as part of your journal data policy or list of repository recommendations. [Want to find out more about journal Dataverse collections?](#)



Institutions

Establish a research data management solution for your community. Federate with a growing list of Dataverse repositories worldwide for increased discoverability of your community's data. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. [Want to install a Dataverse repository?](#)



Developers

Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. [Integrate research analysis, visualization and exploration tools](#), or other research and data archival systems with the Dataverse Project. [Want to contribute?](#)

### DATAVERSE REPOSITORIES - A WORLD VIEW

76 Installations



[View more Metrics](#)





# 10 reasons for our choice of Dataverse:

1. **Flexibility in the organization of collections (dataverses)**, necessary to explicitly include all the Institutional Research Units.
2. **Versatility in the usage of discipline-specific metadata schemas** for each dataverse, from a list of 6 schemas already configured.
3. **Flexibility in the definition of roles and responsibilities in the data curation workflows** and in collection management (8 default roles).
4. **Autonomy in the operation of the curation processes and in system usage**, the user is able to create dataverses and datasets, deposit...
5. **High potential for integration with other systems**, with tools already integrated for data visualization, data organization, georeferencing...



# 10 reasons for our choice of Dataverse:

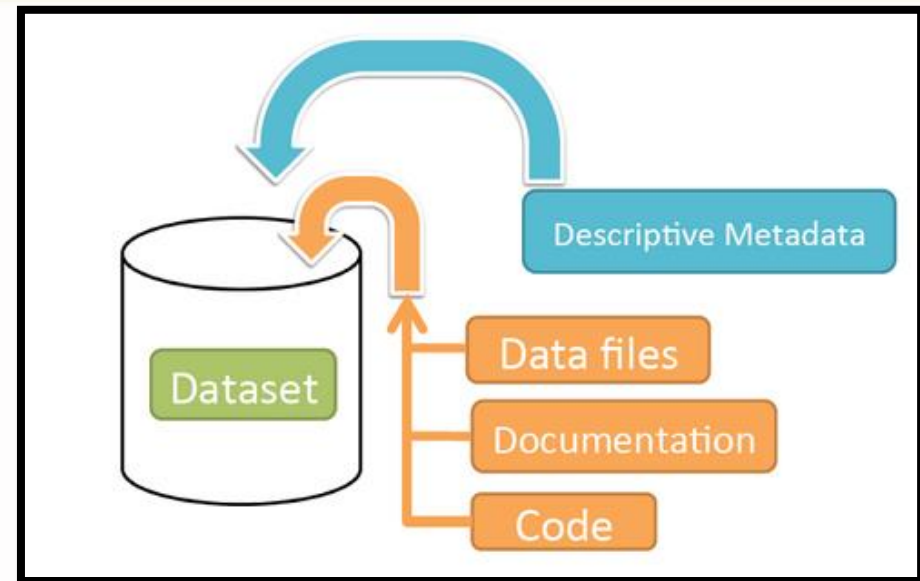
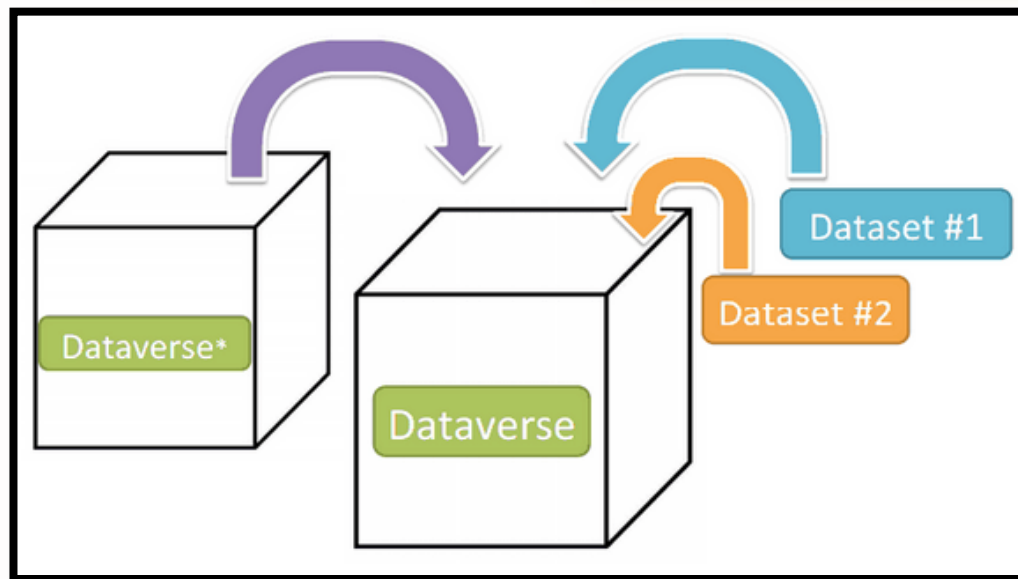
6. **Multiplicity of options in data availability** (open, restricted..., guest book), implementing the “as open as possible, as closed as necessary” principle.
7. **Adaptability of the storage service options**, allowing to use different models or systems (local/cloud) associated with the repository.
8. **Aggregation of data from external sources**, configurable by OAI-PMH and in a dashboard environment.
9. **Global community of research institutions** with relevant software installations, such as the Harvard University.
10. **Transparency in future software developments**, with strong community participation and detailed info about the roadmap and the releases.

# Data RepositóriUM / Dataverse structure

**DATVERSE** = collection / repository / community

**DATASET** = dataset recording with files

**FILE** = files that comprise the datasets



# DataRepositóriUM in practice

The screenshot shows the DataRepositóriUM interface. At the top, there's a navigation bar with 'data RepositóriUM' and links for 'Add Data', 'Search', 'User Guide', 'Support', 'Sign Up', and 'Log In'. Below this is the logo for CECIS (Centro de Estudos de Comunicação e Sociedade) at the University of Minho. The main heading of the dataset is 'A presença dos incêndios 15 de outubro de 2017 nos média online: peças publicadas por 15 meios de comunicação (entre 15 e 31 de outubro) e interações no Facebook'. A 'Dataset Metrics' box shows '17 Downloads'. The 'Description' section provides details about the data collection process. The 'Subject' is 'Social Sciences' and the 'Keyword' is '15 de outubro, incêndios, informação, média online, fires, information, online media'. At the bottom, there's a list of files with download buttons.

DOI

DOI: 10.34622/datarepositorium/QUV6OP

CITATION

Santos, Luís António; Mourão, Marisa; Conrado Filho, Francisco, 2019, "A presença dos incêndios 15 de outubro de 2017 nos média online: peças publicadas por 15 meios de comunicação (entre 15 e 31 de outubro) e interações no Facebook", <https://doi.org/10.34622/datarepositorium/QUV6OP>, Repositório de Dados da Universidade do Minho, V2, UNF:6:IJ9a1z0kLZlr0mRmuhf4Lw== [fileUNF]

Cite Dataset

Learn about Data Citation Standards.

USAGES STATS

Dataset Metrics

17 Downloads

# DataRepositóriUM in practice

**data RepositóriUM**  
Add Data Search User Guide Support Sign Up Log In

CECS Centro de Estudos de Comunicação e Sociedade (Universidade do Minho)  
Repositório de Dados da Universidade do Minho > Centro de Estudos de Comunicação e Sociedade >  
A presença dos incêndios 15 de outubro de 2017 nos média online: peças publicadas por 15 meios de comunicação (entre 15 e 31 de outubro) e interações no Facebook

**A presença dos incêndios 15 de outubro de 2017 nos média online: peças publicadas por 15 meios de comunicação (entre 15 e 31 de outubro) e interações no Facebook**  
Version 2.1

Santos, Luís António, Mourão, Marisa, Conrado Filho, Francisco, 2019, "A presença dos incêndios 15 de outubro de 2017 nos média online: peças publicadas por 15 meios de comunicação (entre 15 e 31 de outubro) e interações no Facebook", <https://doi.org/10.34622/datarepositorium/QUV6OP>; Repositório de Dados da Universidade do Minho, V2, UNF:6:U9a1z0kLZir0mRmuhf4Lw== [file:UNF]

**Dataset Metrics**  
17 Downloads

**Description**  
Foram recolhidas e selecionadas as peças publicadas entre os dias 15 e 31 de outubro de 2017 relacionadas com incêndios de 15 de outubro de 2017, dos seguintes 15 meios de comunicação informativos generalistas com presença online: Correio da Manhã, Destak, Diário de Notícias, Expresso, Jornal de Notícias, Jornal I, Observador, Público, Rádio Renascença, RTP, Sábado, SIC Notícias, TSF, TVI 24 e Visão. Foram capturados todos os links publicados nos sites destes meios entre 15 de outubro a 31 de outubro, o que corresponde a um total de 23.684. Posteriormente, foram selecionadas as peças relativas aos incêndios. Restaram 3.890 links relativamente aos quais recolhemos, no dia 9 de novembro de 2017, junto da API do Facebook, o total de interações de cada um. Para efeitos de análise dos títulos e número de notícias do meio foram, ainda, excluídos todos os links e artigos duplicados e aqueles cujos títulos não foram recolhidos automaticamente e a página já não se encontrava disponível, restando 3.590 links.

**Subject**  
Social Sciences

**Keyword**  
15 de outubro, incêndios, informação, média online, fires, information, online media

Files Metadata Terms Versions

Search this dataset... Find

Filter by  
File Type: All Access: All

1 to 3 of 3 Files

- 2018\_Santos\_Mourao\_ConradoFilho\_incendiosOutubro2017\_NotaMetod  
Plain Text - 2.9 KB - Nov 30, 2019 - 7 Downloads  
MD5: ba54a4e8c7e436a43747b16d56862778  
Breve nota metodológica
- 2018\_Santos\_Mourao\_ConradoFilho\_incendiosOutubro2017\_interacao.t  
Tabular Data - 574.6 KB - Nov 26, 2019 - 5 Downloads  
6 Variables, 3890 Observations - UNF:6:6kPuXozqgSujTvd8JVSJQ==  
Interações (no Facebook) relativas a cada um dos 3.890 links
- 2018\_Santos\_Mourao\_ConradoFilho\_incendiosOutubro2017\_titulo-meio.tab  
Tabular Data - 1019.7 KB - Nov 26, 2019 - 5 Downloads  
107 Variables, 3590 Observations - UNF:6:r5IMV0e/+0mC-4zUSmu+0==  
Títulos e meios de comunicação que publicaram as 3.590 peças (foram excluídos todos os links e artigos duplicados e aqueles cujos títulos não foram recolhidos automaticamente e a página já não se encontrava disponível)

## SET OF FILES

Files Metadata Terms Versions

Search this dataset... Find

Filter by  
File Type: All Access: All

1 to 3 of 3 Files

- 2018\_Santos\_Mourao\_ConradoFilho\_incendiosOutubro2017\_NotaMetod  
Plain Text - 2.9 KB - Nov 30, 2019 - 7 Downloads  
MD5: ba54a4e8c7e436a43747b16d56862778  
Breve nota metodológica
- 2018\_Santos\_Mourao\_ConradoFilho\_incendiosOutubro2017\_interacao.tal  
Tabular Data - 574.6 KB - Nov 26, 2019 - 5 Downloads  
6 Variables, 3890 Observations - UNF:6:6kPuXozqgSujTvd8JVSJQ==  
Interações (no Facebook) relativas a cada um dos 3.890 links

## TERMS OF USE & CONTROLLED ACCESS

Terms of Use

**Waiver**  
Our Community Norms as well as good scientific practices expect that proper credit is given via citation. Please see citation above, generated by the Dataverse.  
No waiver has been selected for this dataset.

**Terms of Use**  
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Guestbook

## VERSIONING

Files Metadata Terms Versions

View Differences

	Dataset	Summary	Contributors	Published
<input type="checkbox"/>	2.1	Terms of Use/Access Changed <a href="#">View Details</a>	Marisa Mourão	Dec 2, 2019
<input type="checkbox"/>	2.0	Citation Metadata: Description (1 Changed), Files (Added: 1); <a href="#">View Details</a>	Marisa Mourão	Nov 30, 2019
<input type="checkbox"/>	1.0	This is the first published version.	Marisa Mourão	Nov 26, 2019

# DataRepositóriUM in practice

The screenshot shows the DataRepositóriUM website interface. At the top, there is a navigation bar with 'data RepositóriUM', 'Add Data', 'Search', 'User Guide', 'Support', 'English', 'Sign Up', and 'Log In'. Below this, the 'EDUPLACES (Universidade do Minho)' logo and 'EDUPLACES home page' link are visible. The main content area displays 'Repositório de Dados da Universidade do Minho > Centro de Investigação em Educação > EDUPLACES' and a search bar with 'Search this dataverse...'. On the left, there are filters for 'Dataverses (4)', 'Datasets (31)', and 'Files (96)'. The main results area shows '1 to 10 of 35 Results' with a 'Sort' button. The first result is 'Perfis de Percursos Académicos Atípicos' by Lúcio, Joana, 2019. Subsequent results include 'Guião da entrevista - Responsáveis das Práticas', 'Entrevista 1 - Responsável da Prática N\_SB\_2', 'Entrevista 2 - Responsável da Prática N\_CB\_2', 'Entrevista 3 - Responsável da Prática NW\_SB\_2', and 'Entrevista 4 - Responsável da Prática NW\_CB'.

## PROJECTS DATAVERSE

<https://datarepositorium.uminho.pt/dataverse/eduplaces>

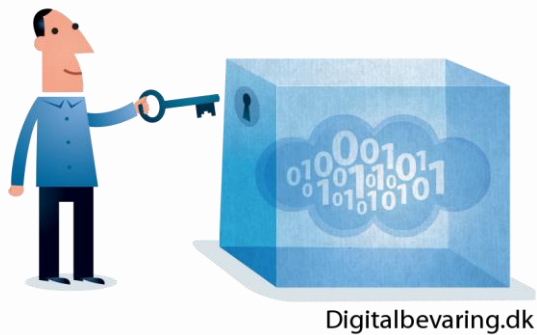
The screenshot shows the 'EDUPLACES (Universidade do Minho)' dataverse page. It features the 'data RepositóriUM' logo, 'EDUPLACES (Universidade do Minho)', and 'EDUPLACES home page' link. The breadcrumb trail reads 'Repositório de Dados da Universidade do Minho > Centro de Investigação em Educação > EDUPLACES'. The page is framed by a yellow border.

## FUNDER IDENTIFICATION

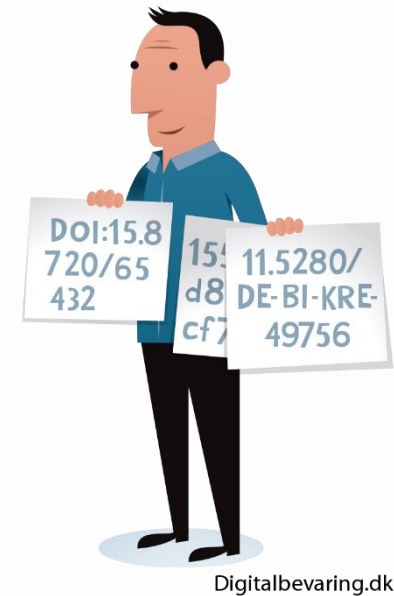
The screenshot shows the 'Entrevista 8 - Responsável da Prática N\_CB\_1' dataset page. It displays the dataset title, author 'Lúcio, Joana, 2019', and the date 'Nov 29, 2019'. Below the dataset information, there is a pagination bar showing '<< < Previous 1'. At the bottom of the page, there are logos for 'CIEE Instituto de Educação - UMINHO', 'CIEE', 'Cetrad', 'CIEO', 'FCT', and 'REPÚBLICA PORTUGUESA'.

# Dataverse highlights

## SECURE DATA MANAGEMENT.



## TRACK CHANGES.



## EFFECTIVE SHARING.

## LONG-TERM ACCESS AND PRESERVATION.

# Dataverse highlights

**ORGANIZATION AND  
COMPATIBILITY.**



**SAVE TIME.**

**INCREASE  
RESEARCH VISIBILITY.**



**MEET GRANT  
REQUIREMENTS.**



# ***FINAL REMARKS***

## **REALITIES OF RDM**

One size does not fit all in scoping the Institutions RDM services bundle.

## **INSTITUTIONAL STRATEGY**

The repository as a facility embedded in an institutional strategy and policy.

## **LONG TAIL DATA**

The repository is central in supporting researchers to manage & share long tail data.

## **RDM GOOD PRACTICES**

The repository is instrumental to improve data documentation and research data publishing.

## **DATA LIFECYCLE**

The repository is a critical service in the data curation life cycle stages.





Questions?

Thanks!



**Universidade do Minho**

Serviço de Documentação e Bibliotecas

Pedro Príncipe

[pedro.principe@usdb.uminho.pt](mailto:pedro.principe@usdb.uminho.pt)

