

TAIL – Workflow para a gestão de dados de investigação

João Aguiar Castro, Yulia Karimova, João Rocha da Silva, Cristina Ribeiro
INESC TEC—Faculdade de Engenharia da Universidade do Porto



TAIL Team



Cristina
Ribeiro



João Rocha
da Silva



João Aguiar
Castro



Ricardo
Amorim



Yulia
Karimova



Carla
Lopes



Artur
Rocha



João
C. Lopes



Gabriel
David



Tito
Vieira



Eugénia
Fernandes



Ângela
Lomba



João
Honrado



Gestão de dados de investigação

- Motivação:
 - Gestão de dados requerida pelas agências de financiamento
 - Implementação de infraestruturas RDM nas instituições
- Desafios:
 - Complexidade dos recursos
 - Processo de publicação pouco claro
 - Investigadores não familiarizados com descrição de dados

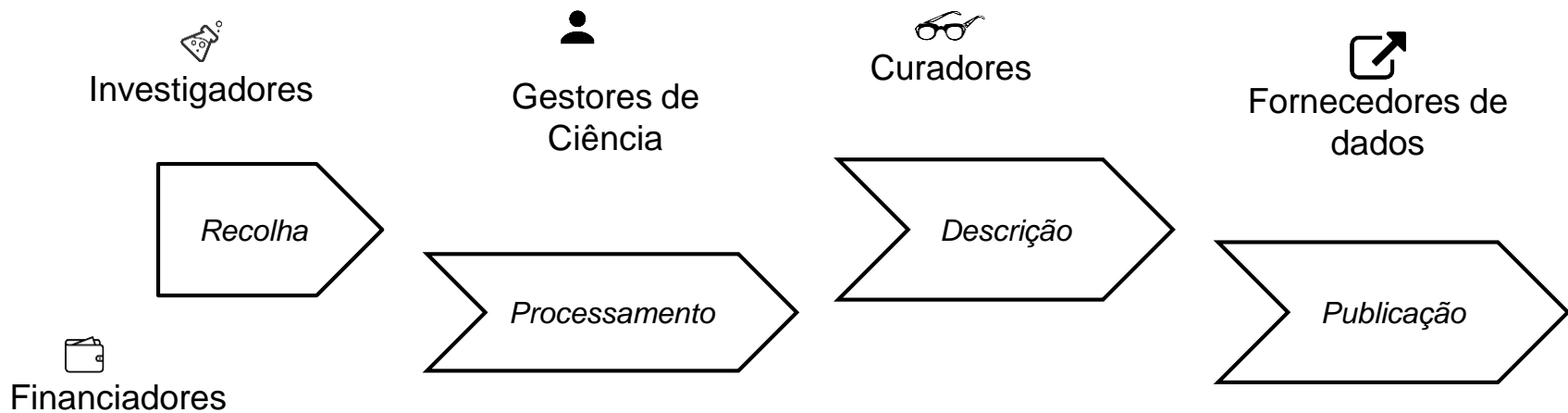
Gestão de dados de investigação

- Oportunidades:
 - Visibilidade para investigadores e instituições
 - Promover reutilização de dados
 - Conformidade com Planos de Gestão de Dados
 - Agilização do fluxo de trabalho

Abordagem RDM na Universidade do Porto

- Foco nos investigadores
 - Criam os requisitos
 - Sabem mais sobre os dados

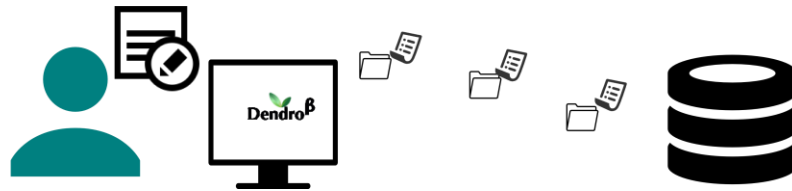
- Foco em ferramentas
 - Para curadores
 - Para investigadores



Programadores

dendro.fe.up.pt

Instituições



Dendro

- Organização e descrição de dados
- Ontologias carregadas para cada domínio
- Integração com repositórios de dados

Dendro

Folder



Selection



- Up to biodata
- meta
- scala_2014.jpg
- observations.xls
- ontTraffic.pdf

Information | Change log

Description progress 0%

SAVE UNDO COPY FROM PARENT IN MANUAL MODE CLEAR

Description

B I H [Icons] PREVIEW

Studies of the endangered species of felid living mainly in the Iberian Peninsula in southwestern Europe. They are categorized as endangered by many institutions, including the International Union for the Conservation of Nature

Coverage



Descriptor Selection

ALL AUTO

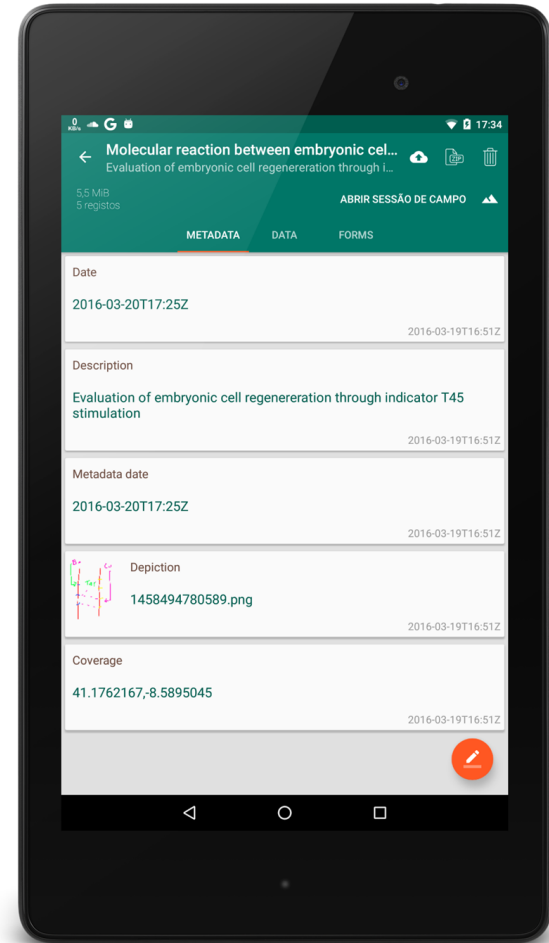
Find descriptors

Type to find descriptors...

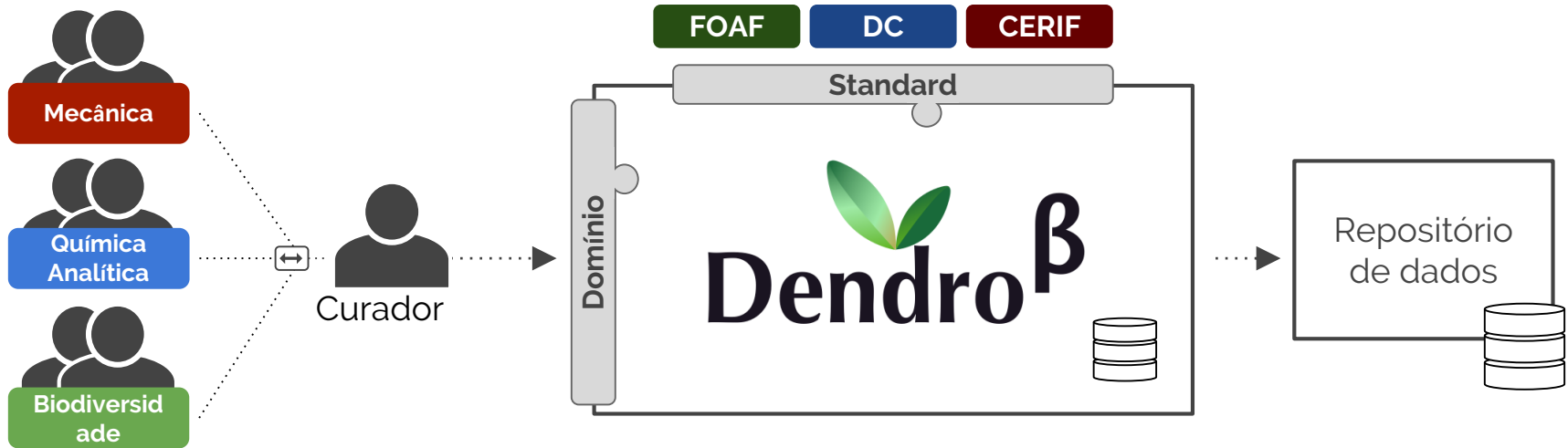
- DEPICTION: A depiction of some thing.
- DESCRIPTION: An account of the resource.
- DATE: A point or period of time associated with an event in the lifecycle of the resource.
- COVERAGE: The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
- SUBJECT

LabTablet

- Metadados “automáticos”
- Metadados de sensores
- Ontologias: do Dendro



Ontologias genéricas e específicas a domínio



- Reutilização de modelos de metadados
- Ontologias de domínio: colaboração entre curador e investigador

Domínios e tipos de dados

Domínio	Tipo de dados	
Fratura de Materiais	Experimental	
Produção de Hidrogénio		
Química Analítica	Experimental	Observacional
Gravimetria		
Oceanografia		
Biodiversidade	Observacional	
Ciências Sociais		
Corte e Empacotamento	Simulação	
Simulação de Veículos		
Dinâmica de Fluídos		

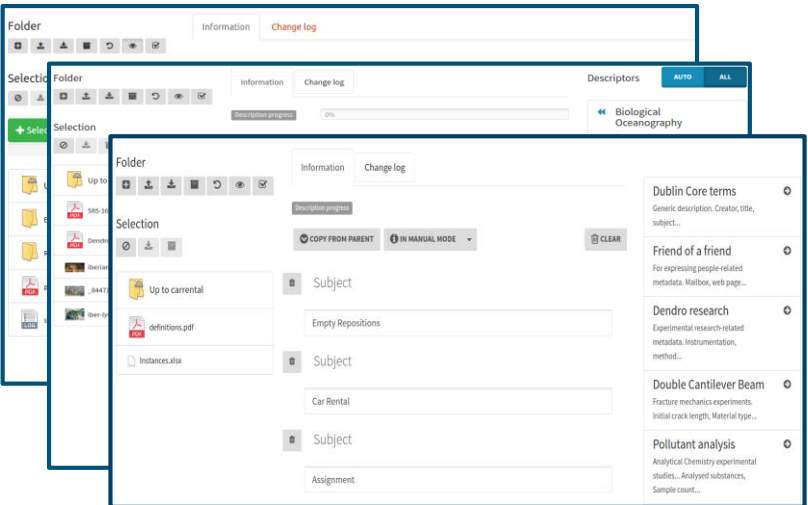


- Colaboração RDM@FEUP - EUDAT
 - **DataPublication@UPorto**: descrição de dados multidisciplinar e workflow de depósito entre o Dendro, e a Infraestrutura Europeia *EUDAT*

<https://eudat.eu/communities/datapublication-uporto>



Dendro & EUDAT



Vehicle reservation assignment in car rental

Beatriz Oliveira
University of Porto

Abstract: Data for the vehicle-reservation assignment problem in a car rental company, including vehicle and reservation information, car transfers time and cost (collected for Portugal), and upgrading and downgrading policies. Data used in Oliveira, B.B., Carravilla, M.A., Oliveira, J.F. and Toledo, F. M.B., 'A relax-and-fix-based algorithm for the vehicle-reservation assignment problem in a car rental company', European Journal of Operational Research, Volume 237, Issue 2, 1 September 2014, Pages 729-737, ISSN 0377-2217, <http://dx.doi.org/10.1016/j.ejor.2014.02.018>.

Files ▾

Name	Date	Size	
files.zip	10 Jul 2016	2.1 MB	Download
metadata.rdf	10 Jul 2016	1.9 kB	Download
metadata.txt	10 Jul 2016	1.7 kB	Download

Export
Export as [BibTeX](#), [MARC](#), [MARCXML](#), [DC](#), [EndNote](#), [NLM](#), [RefWorks](#)

Metadata

PID: <http://hdl.handle.net/11304/12dc669e-6355-4353-b021-8887fed6528c>

Publication: University of Porto

Licence: Attribution-ShareAlike 4.0 International

Uploaded by: dendrordm@gmail.com

Domain: generic

Checksum: 8c5a2ed9bd664ca14121a957d9764a046aaeaba8cb615151225doadb4e128f73

Rate this document:

☆☆☆☆☆

(Not yet reviewed)
[Report abuse](#)

Repositório de Dados INESC

The screenshot shows the homepage of the INESC TEC Research Data Repository. The header is dark blue with the INESC TEC logo on the left and navigation links for 'Datasets', 'Organizations', 'Groups', and 'About' on the right. A search bar is located in the top right corner. The main content area features a large white box with a blue header that reads 'INESC TEC Research Data Repository' and a sub-header 'This is a featured section'. To the right of this box is a 'Search data' section with a search input field containing 'E.g. environment' and a magnifying glass icon. Below the search field are 'Popular tags' for 'recommender systems', 'radon', and 'playlists'. The bottom section contains two data cards. The left card is for 'CS: Computer Science' and lists datasets like 'Music streaming and playlisting activity from music streaming service' and 'Gamma radiation monitoring at the Azores (ENA-ARM station) and Quality-Contro...'. The right card is for 'INESC TEC' and lists datasets like 'Music streaming and playlisting activity from music streaming service' and 'Albergaria Fire Fighters ECG 2010'.

Welcome to the INESC TEC research data repository.

This data repository showcases datasets produced or used by INESC TEC researchers and their partners. It is an embodiment of our institutional commitment to Open Data in research.

INESC TEC
Research Data
Repository

This is a featured section

Search data

E.g. environment

Popular tags: recommender systems, radon, playlists

CS: Computer Science
The Computer Science Cluster mission is to...

Music streaming and playlisting activity from music streaming service
This dataset is taken from the portuguese music social network "Palco Principal", that gathers non-mainstream...

Gamma radiation monitoring at the Azores (ENA-ARM station) and Quality-Contro...
Data from the gamma radiation monitoring campaign at the ENA-ARM facility (Graciosa, Azores), including both raw and...

INESC TEC
The Institute for Systems and Computer...

Music streaming and playlisting activity from music streaming service
This dataset is taken from the portuguese music social network "Palco Principal", that gathers non-mainstream...

Albergaria Fire Fighters ECG 2010
Monitoring of Albergaria Firefighters ECG during emergency events during the summer of 2010 and correspondent event...

Repositório de Dados INESC

[Dataset](#) [Groups](#) [Activity Stream](#)

Music streaming and playlisting activity from music streaming service

This dataset is taken from the portuguese music social network "Palco Principal", that gathers non-mainstream musicians with fans. The website allows free music streaming and users can organize their favorite music tracks in personal playlists. The dataset consists of three files. Two of them are music streaming logs (one line per play), and one file is a music track playlisting log (users adding music tracks to their personal playlist).

Data and Resources



README.txt

Dataset description file

[Explore](#)



listened_tracks_1.tsv

Music streaming log From 2007-06-15 to 2010-05-01 1.171.849 plays Fields...

[Explore](#)

Additional Info

Field	Value
Source	Palco Principal (http://www.palcoprincipal.com)
Author	João Vinagre
Last Updated	13 de Março de 2017, 10:47 (UTC+00:00)
Created	10 de Março de 2017, 10:42 (UTC+00:00)
dc.Contributor	Ubbin Labs, Lda
dc.Coverage.Temporal	listened_tracks_1.tsv - From 2007-06-05 to 2010-05-01; listened_tracks_2.tsv - From 2010-10-20 to 2011-03-22; playlisted_tracks.tsv - From 2009-06-09 to 2013-03-25
dc.Date	Gathered March 2013
dc.Format	*.tsv, *.txt
dc.Format.Extent	46,5MB
dc.Relation	J. Vinagre, A. M. Jorge and J. Gama. Collaborative filtering with recency-based negative feedback. In proceedings of the 30th ACM/SIGAPP Symposium on Applied Computing (SAC'2015), Salamanca, Spain, April 13-17, pages 963-965. ACM, 2015.; P. Matuszyk, J. Vinagre, M. Spiliopoulou, A. M. Jorge, J. Gama. Forgetting methods for incremental matrix factorization in recommender systems. In proceedings of the 30th ACM/SIGAPP Symposium on Applied Computing (SAC'2015), Salamanca, Spain, April 13-17, pages 947-953. ACM, 2015.; J. Vinagre, A. M. Jorge, and J. Gama. Fast incremental matrix factorization for recommendation with positive-only feedback. In proceedings of the 22nd Conference on User Modeling, Adaptation and Personalization, UMAP'2014, Aalborg, Denmark, July 7-11. Volume 8538 of Lecture Notes in Computer Science, pages 459-470. Springer, 2014.
dc.Type	listened_tracks_1.tsv - music streaming, tab-separated; listened_tracks_2.tsv - music streaming, tab-separated; playlisted_tracks.tsv - music playlisting, tab-separated

Dendro

&

Repositório de dados

- RDM cedo no workflow

- Desde o início de projecto
- Envolvimento de investigadores

- Ambiente colaborativo

- Revisão por pares
- Recomendações de investigadores séniores

- Identificadores persistentes

- Preservação de dados

- Conformidade com normas

- Disseminação

- Recolha automática
- Protocolos de disseminação

Publicações relacionadas

❑ TPDFL 2016

- ❑ **Usage-driven Dublin Core descriptor selection—A case study using the Dendro platform for research dataset description**

❑ UAIS, Springer 2016

- ❑ A comparison of research data management platforms: architecture, flexible metadata and interoperability

❑ iPRES 2014

- ❑ **The Dendro research data management platform: applying ontologies to long-term preservation in a collaborative environment**

❑ Digital Libraries 2014

- ❑ **Creating lightweight ontologies for dataset description: Practical applications in a cross-domain research data management workflow**

Visite

Publicações e novidades

<http://dendro.fe.up.pt/>

O Dendro é *open source* no GitHub!

<https://github.com/feup-infolab-rdm>