

Joint event with the Portuguese “Fórum Gestão de Dados de Investigação”

TAIL Research data management from creation to deposit and sharing *final workshop*

FEUP, INESC TEC, CIBIO, University of Porto



TAIL Goals and Achievements

- Goals

- Develop software tools for supporting storage, description, publication and search
- Involve research groups in RDM
- Support researchers in data publication
- Connect with work in roadmap infrastructures

- Achievements

- Dendro
- LabTablet
- Set of domain-specific ontologies
- Set of deposited datasets
- Case studies
- Connection with EUDAT European infrastructure as a pilot



Dendro, an integrated data management environment *(João Rocha da Silva)*

The screenshot displays the Dendro web interface for a project named 'musicstreaming'. The interface is organized into several sections:

- Header:** Includes the Dendro logo, user profile 'João Rocha', and search functionality.
- Project Navigation:** Shows 'My Projects > musicstreaming' and tabs for 'About', 'Base Data', 'Changes', 'Project Changes', and 'Project Stats'.
- Folder Management:** A 'Folder' section on the left shows a tree view with 'My Projects' and 'Base Data'.
- Abstract Editor:** A central text area with a rich text editor (B, I, H, etc.) containing the text: "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)."
- Metadata Fields:** A list of fields for project metadata, including:
 - Contributor: Ubbin Labs, Lda
 - Creator: João Vinagre
 - Date: 03/10/2017
 - Format: *.tsv, *.txt
 - Relation: (empty)
- Actions:** Buttons for 'EDIT MODE ON', 'SHARE', 'SAVE', 'UNDO', 'COPY FROM PARENT', and 'IN MANUAL MODE'.
- Right Panel:** A 'SUGGESTIONS' section with a search bar and a list of expandable categories: ABSTRACT, ACCESS RIGHTS, ACCRUAL METHOD, ACCRUAL PERIODICITY, ACCRUAL POLICY, ALTERNATIVE TITLE, AUDIENCE, AUDIENCE EDUCATION LEVEL, BIBLIOGRAPHIC CITATION, CONFORMS TO, and CONTRIBUTOR.
- Footer:** Copyright information: © FEUP InfoLab. Version: dendro Dev v.0.2. Active Branch: master-epi-404. Commit: 28722070980390600804401767023070 Date: Mon Apr 2 20:40:42 2017 -0200.



What is it?

- “Dropbox” for research data
- A data portal for research groups
- Ontology-based metadata
- Fully built on a graph model
- Scalable storage
- Covered by integration tests
- Open-Source (BSD License)
- Available via [🔗 GitHub](#), [🔗 Docker image](#) and [🔗 installation scripts](#)



Features

Data Staging

- Collaboration
- Data Storage “like Dropbox”
- Domain-specific metadata
- User access controls
- External deposit in any repository

Data Repository

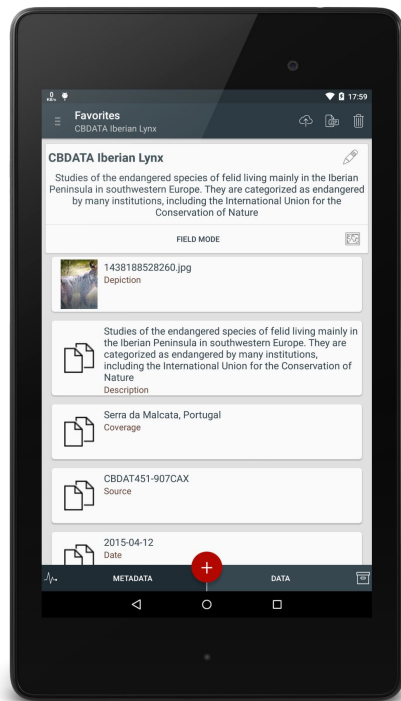
- Internal deposit
- Faceted search
- DOI attribution
- Request access
- BibTeX snippets



Demo

LabTablet, a mobile data and metadata collection application

(*João Rocha da Silva*)



- Android app
- Electronic Laboratory Notebook
- “Automatic” metadata
- Metadata from onboard sensors
- Laboratory notebook notes = great metadata!
- Integrates with Dendro
- Available on the [Play Store](#) and [GitHub](#) (GPL v3 license)

Demo

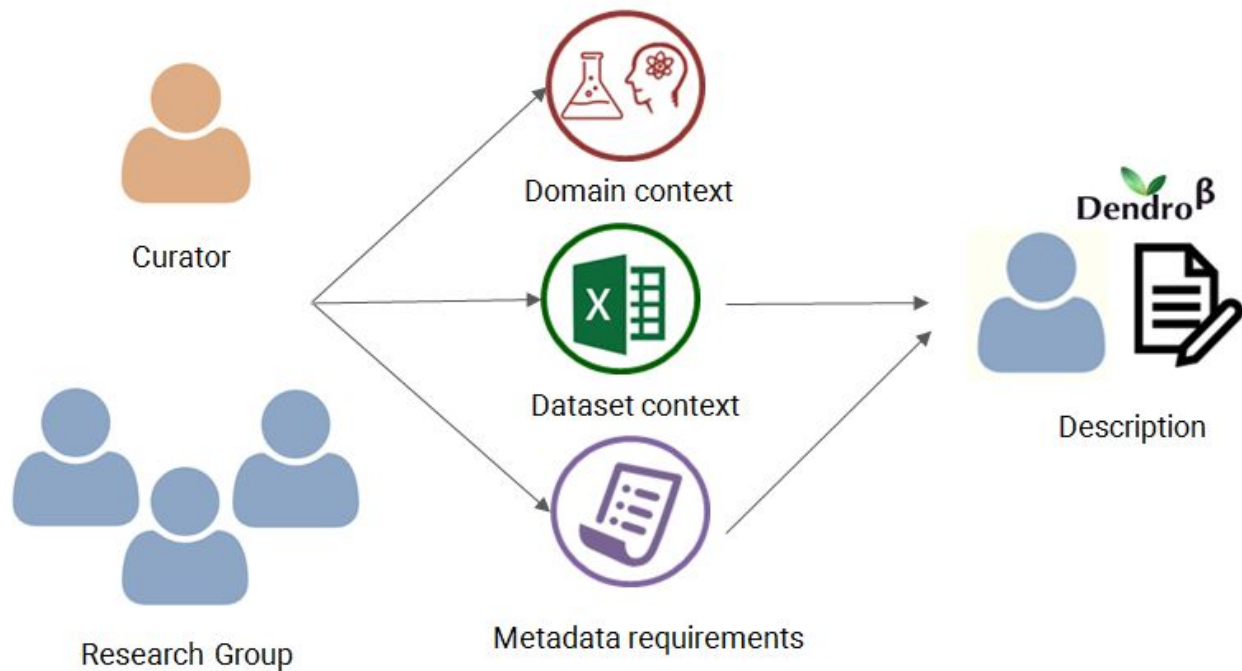
Demo



Engaging researchers in research data management *(João Aguiar Castro)*

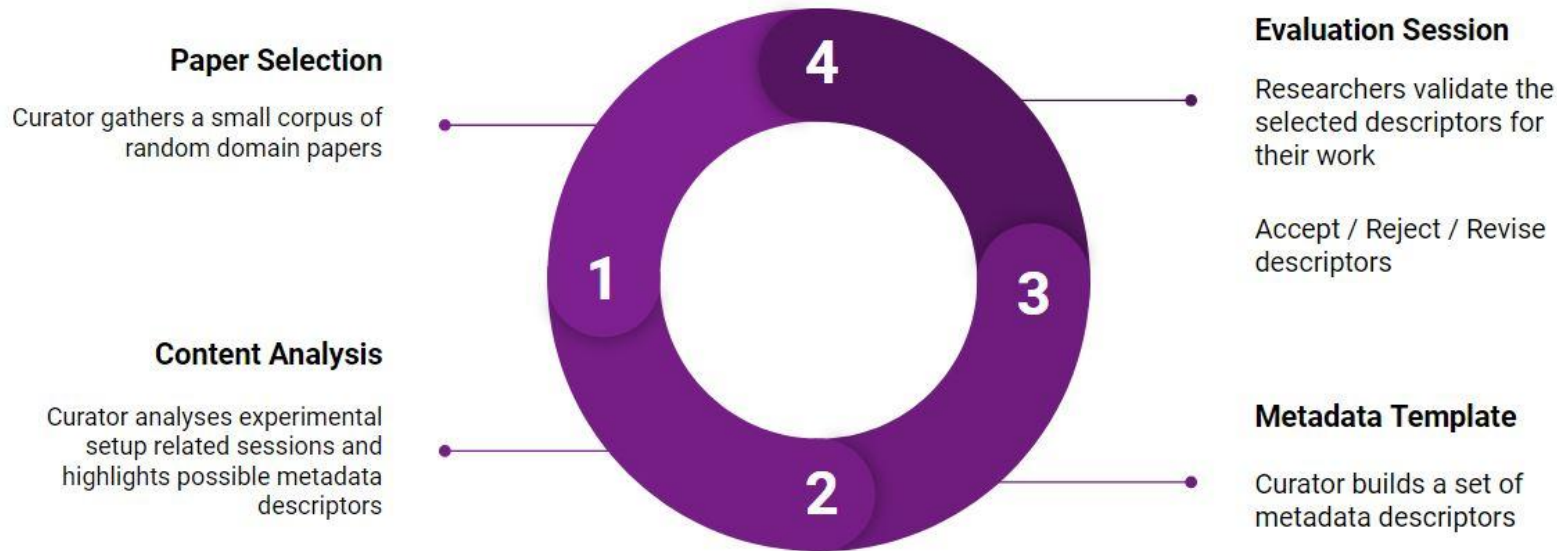


Researchers feedback is essential to define domain-specific metadata models *(João Aguiar Castro)* - activities to engage them



Content analysis to improve communication with researchers

(*João Aguiar Castro*) - before the meeting with researchers the curator has already selected domain descriptors



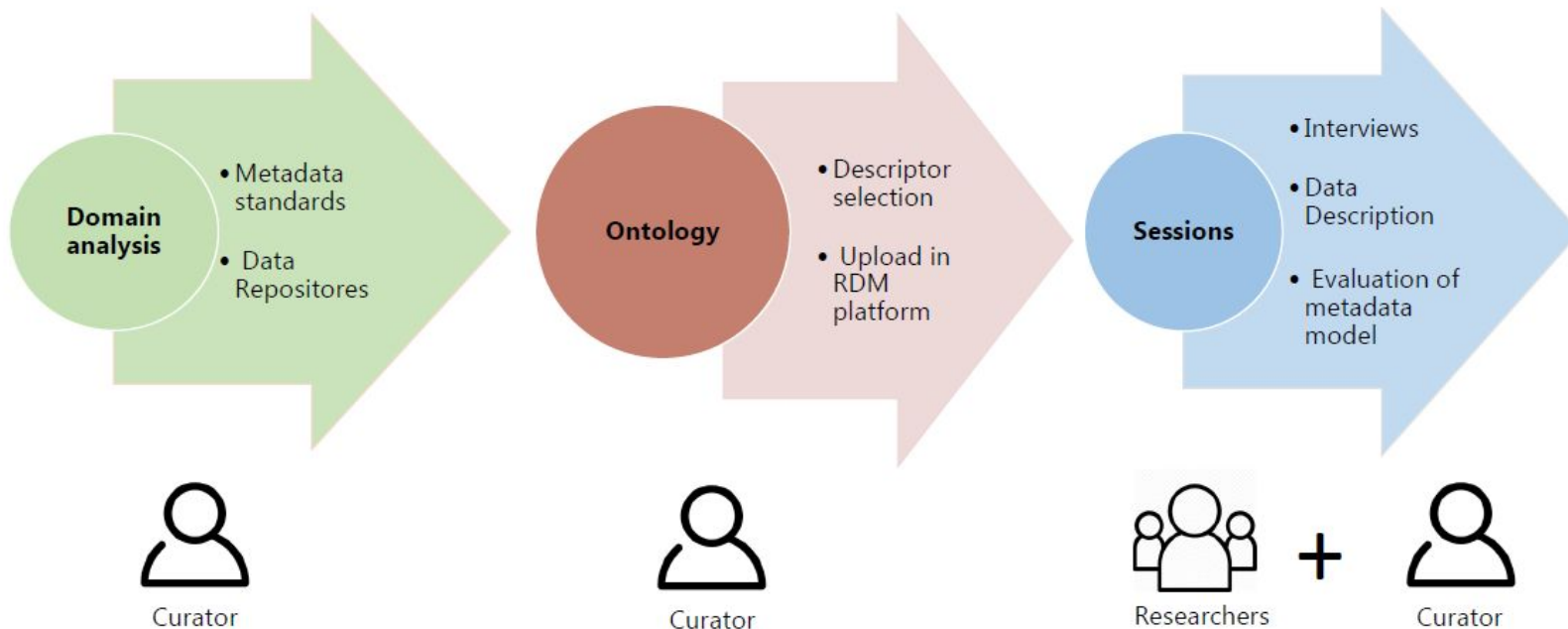
Content analysis to improve communication with researchers

(*João Aguiar Castro*) – researchers evaluate the proposed descriptors

| Descriptor evaluation | Sustainable Chemistry | Photovoltaic Generation | Nanoparticle Synthesis |
|-------------------------|---|--------------------------|--------------------------------|
| Proposed by the curator | 60 | 56 | 23 |
| Directly accepted | 38 | 42 | 18 |
| Revised | 15 | 3 | 3 |
| New suggestions | Catalyst wavelength | Dielectric constant real | |
| Not needed | 3 | 11 | Reducing agent |
| Not understood | 3 | | Passive molecule concentration |
| Overall acceptance rate | 84 per cent descriptors directly accepted | | |

Starting with available scientific metadata standards *(João Aguiar Castro)*

- a descriptor subset is selected by the curator



Starting with available scientific metadata standards *(João Aguiar Castro)*

- Selection and representation of 30 descriptors from the MIBBI standard (biomedics)

| Category | Descriptors |
|------------|--|
| Sample | Organism, Disease, Organism Part, Age, Sex, Ethnicity, Developmental Stage, Tissue, Cell Line, Cell Type, Sample Size, Molecule, Sample Type |
| Methods | Assay Type, Collection Date, Measurement, Method, Sample Collection Protocol, Treatment Protocol, Temperature, Study Design |
| Materials | Material, Drug Usage, Reagent |
| Technology | Instrument Name, Instrument Type, Software |
| Others | Experimental Factor, Environmental Factor, Study Domain |

Starting with available scientific metadata standards *(João Aguiar Castro)*

- Data descriptions sessions in the biomedical domain

| Descriptor | Researcher 1 | Researcher 2 | Researcher 3 | Researcher 4 |
|--------------------------------|---|---|--|---------------------------------|
| Disease | gastric carcinoma | hypertension | gastric cancer | cancer |
| Instrument Name | Illumina HISEQ (2500) | Illumina | Ion Torrent Sequencer (ThermoFischer, City, Country) | Flow Cytometer |
| Material | Truseq | whatmann paper | RPMI and Bovine Serum | collection tubes |
| Method | Protocol Reference | Protocol Reference | Stop infection Remove medium Wash 2x with RPMI medium Add new medium - 200 uL R10 Add RTK inhibitors - 2uL per each 96 well (dil 1:1000) | Staining for immunofluorescence |
| Organism | Homo sapiens | Homo sapiens | Homo Sapiens | Human |
| Organism Part | stomach | blood | stomach | gut |
| Software | GraphPad | Sequencher | GraphPad v8 (statistical analysis) — IDEAS software v3 (imaging analysis) | FlowJo |
| Study Domain | Disease susceptibility | Genetic Diversity | Oncology | Stem cells and cancer |
| Recommended descriptors | Replicate Count, Replicate Type, Country of Origin and Study Type | Sample Identifier, Instrument Manufacturer, Study Type and Protocol | Clinical Trial Description, Clinical Trial Phase, Clinical Trial Type, Collection Site | X |

Lightweight ontologies *(João Aguiar Castro)*

owl:topDataProperty

- Age
- 'Assay Type'
- 'Cell Line'
- 'Cell Type'
- 'Collection Date'
- 'Developmental Stage'
- Disease
- 'Drug Usage'
- 'Environmental Factor'
- Ethnicity
- 'Experimental Factor'
- 'Instrument Name'
- 'Instrument Type'
- Material
- Measurement
- Method
- Molecule
- Organism
- 'Organism Part'
- Reagent
- 'Sample Collection Protocol'
- 'Sample Size'
- 'Sample Type'
- Sex
- Software
- 'Study Design'
- 'Study Domain'
- Temperature
- Tissue
- 'Treatment Protocol'



SUGGESTIONS MANUAL SELECTION

Search for descriptor...

| | |
|--|---|
| Dublin Core terms | ⊕ |
| Friend of a friend | ⊕ |
| Dendro research | ⊕ |
| Biodiversity evolution studies | ⊕ |
| Hydrogen Generation | ⊕ |
| Data Documentation Initiative (DDI) | ⊕ |
| DDI-RDF Discovery Vocabulary | ⊕ |
| Chemistry | ⊕ |
| Minimum Information for Biological and Biomedical Investigations | ⊕ |

Vocabulary for the description of data resulting from experiments in the biology/biomedical domains: samples used, methods, etc...

Lightweight ontologies *(João Aguiar Castro)*

Universe

Todos os Programas dos Governos Constitucionais da República Portuguesa entre 1996 e 2013

Analysis Unit

Políticas governamentais para as áreas do ambiente e energia

Independent Dimension

Contexto político e económico

- + DATA COLLECTION DATE
- + DATA COLLECTION METHODOLOGY
- + DATA COLLECTION SOFTWARE
- + DATA SOURCE
- + EXTERNAL AID
- + KIND OF DATA
- + METHODOLOGY
- + SAMPLE SIZE
- + SAMPLING PROCEDURE
- + UNIVERSE

Concept extraction (Carla Teixeira Lopes)

Project documents



Curator



Creation



Domain familiarization

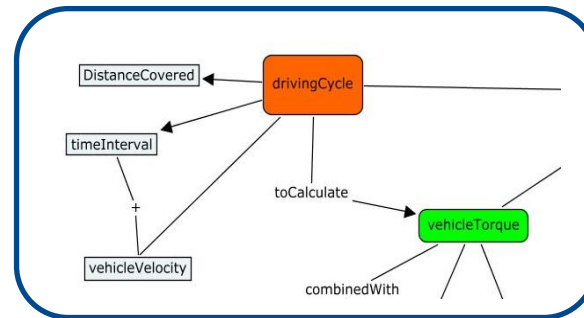


Researcher

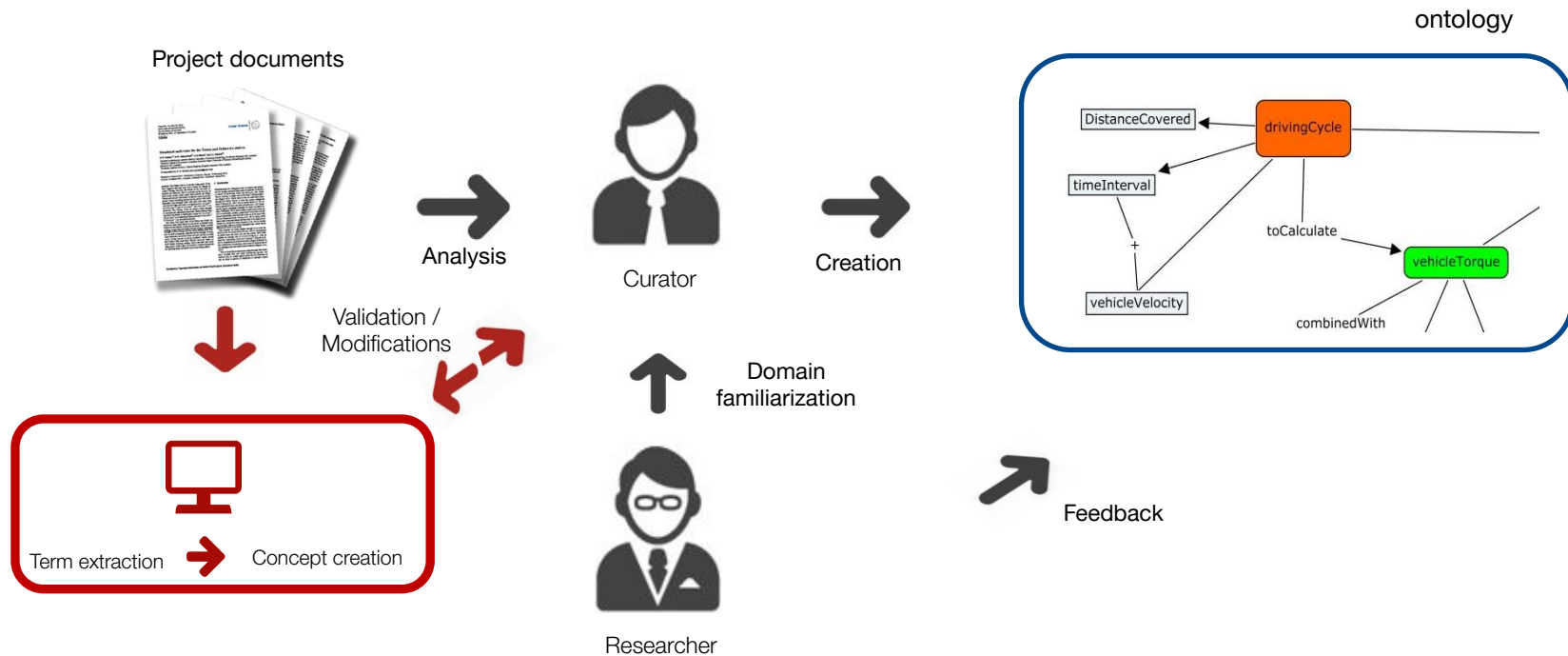


Feedback

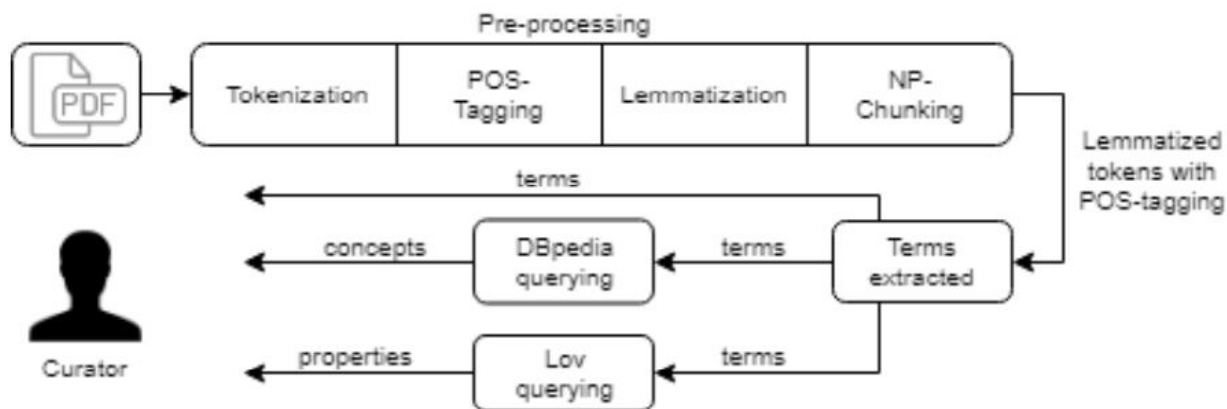
ontology



Concept extraction *(Carla Teixeira Lopes)*



Concept extraction *(Carla Teixeira Lopes)*



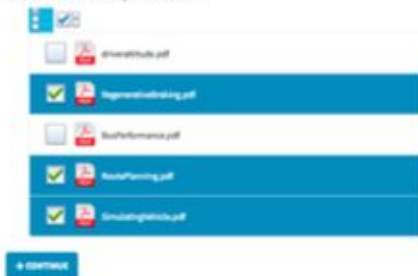
Concept extraction (Carla Teixeira Lopes)



Editing project "Vehicle Simulation"

Metadata People Settings Storage Ontology Learning

Select files for concept extraction



List of files within the project



Editing project "Vehicle Simulation"

Metadata People Settings Storage Ontology Learning

Select terms to query DBpedia



List of extracted terms ordered by score

Concept extraction *(Carla Teixeira Lopes)*

The screenshot shows the Dendro web interface for editing a project named "Vehicle Simulation". At the top, there is a search bar with the text "Search Dendro" and a "SUBMIT" button. Below the search bar, the project name "Editing project 'Vehicle Simulation'" is displayed. A navigation menu includes "Metadata", "People", "Settings", "Storage", and "Ontology Learning". The main content area is titled "Select terms to query DBpedia" and contains a list of terms. A blue button labeled "Terms List" is visible in the top right corner of the list area.

Select terms to query DBpedia

- slips model
- performance model
- microscopic model
- system model
- simulink model
- super capacitor bus model
- bus model
- single simulation model
- simulation model
- rule object model
- hia object model
- federation object model
- object model
- high level model
- car following model
- mobility system
- traffic system
- transportation system
- real system
- transmission system
- brake system
- battery system
- entire battery system

Example of clustering in Vehicle Simulation

The screenshot shows a search input box containing the text "kinetic energy; vehicle model; controller efficiency". To the right of the input box is a blue button labeled "AB". Below the input box, there is a line of text: "Extra terms that might be important but are not presented in the current list". At the bottom left of the input box area is a blue button labeled "BACK", and at the bottom right is a blue button labeled "CONTINUE".

kinetic energy; vehicle model; controller efficiency

Extra terms that might be important but are not presented in the current list

BACK CONTINUE

Input box for the addition of new search terms

Concept extraction (Carla Teixeira Lopes)

Dendro[®] My - Find - Search Dendro SUBMIT Claudio Monteiro -

Metadata People Settings Storage **Ontology Learning**

Select concepts to query LOV

| | |
|--------------------------|--|
| <input type="checkbox"/> | Search Term: controller efficiency Label: Uri: Description: |
| <input type="checkbox"/> | Search Term: vehicle model Label: Model military vehicle Uri: http://dbpedia.org/resource/Model_military_vehicle Description: |
| <input type="checkbox"/> | Search Term: kinetic energy Label: Kinetic energy Uri: http://dbpedia.org/resource/Kinetic_energy Description: The kinetic energy of an object is the energy which it possesses due to its motion. It is defined as the work needed to accelerate a body of a given mass from rest to its stated velocity. Having gained this energy during its acceleration, the body maintains this kinetic energy unless its speed changes. The same amount of work is done by the body in decelerating from its current speed to a state of rest. |
| <input type="checkbox"/> | Search Term: deborah perrotta Label: Uri: Description: |
| <input type="checkbox"/> | Search Term: electric vehicle Label: Electric vehicle Uri: http://dbpedia.org/resource/Electric_vehicle Description: An electric vehicle (EV), also referred to as an electric drive vehicle, uses one or more electric motors or traction motors for propulsion. Three main types of electric vehicles exist, those that are directly powered from an external power station, those that are powered by stored electricity originally from an external power source, and those that are powered by an on-board electrical generator, such as an internal combustion engine (a hybrid electric vehicle) or a hydrogen fuel cell. |

DBpedia Label, URI and description for the terms selected

Dendro[®] My - Find - Search Dendro SUBMIT Claudio Monteiro -

Editing project "Vehicle Simulation"

Metadata People Settings Storage **Ontology Learning**

List of terms, descriptions and possible descriptor

| |
|--|
| Search Term: vehicle model Label: Model Uri: http://www.w3.org/2003/12/wif/nsmodel Vocabulary: owl |
| Search Term: kinetic energy Label: Vocabulary used to describe clean energy actors, projects and technologies Uri: http://reegle.info/schematenergyframework Vocabulary: reegle |
| Search Term: electric vehicle Label: vehicle Uri: http://dbpedia.org/ontology/vehicle Vocabulary: dbpedia-owl |
| Search Term: road network Label: road Uri: http://dbpedia.org/ontology/road Vocabulary: dbpedia-owl |

LOV properties based on the search terms

Concept extraction *(Carla Teixeira Lopes)*

Automatic Evaluation

3 ontologies manually created within Dendro + materials used during their creation.
Vehicle simulation, Sustainable Chemistry, Photovoltaic Application

These ontologies were considered the gold standard to which our results were compared.

Although the precision was not high, we were able to provide the user with a right quantity of concepts.

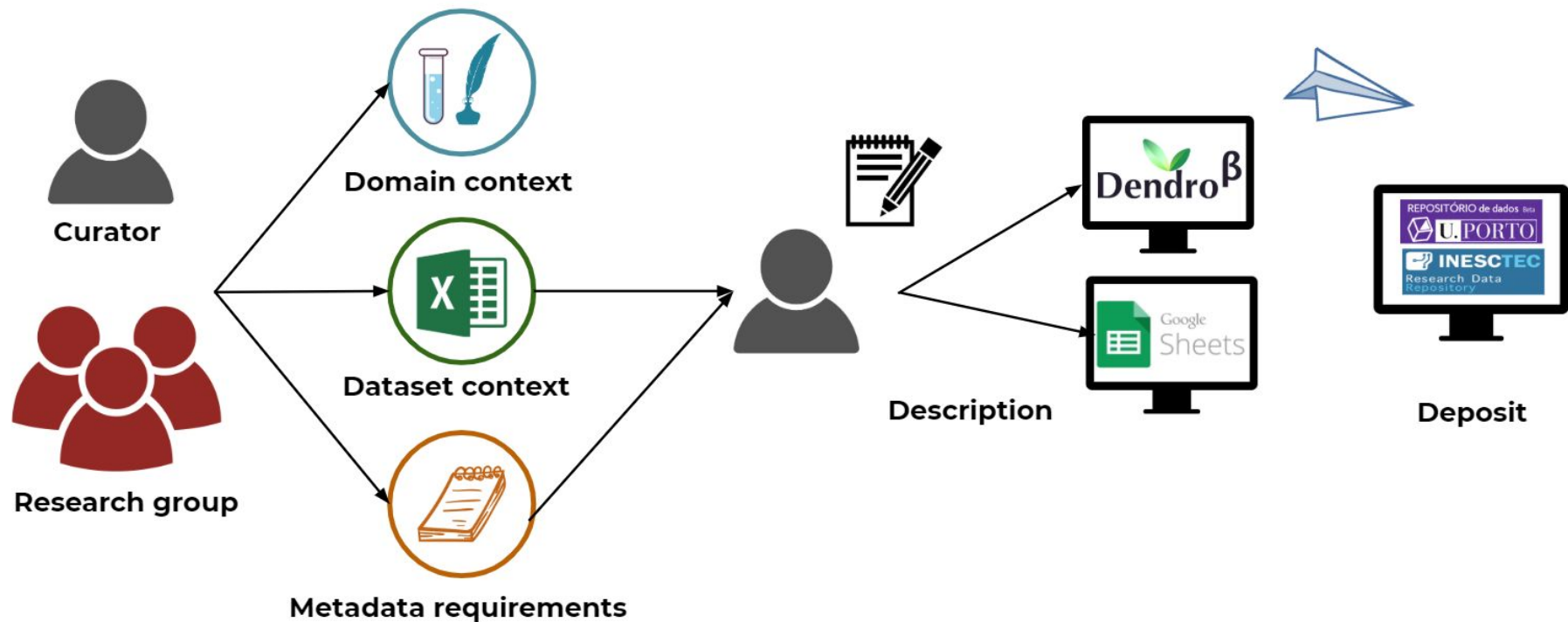
Manual Evaluation

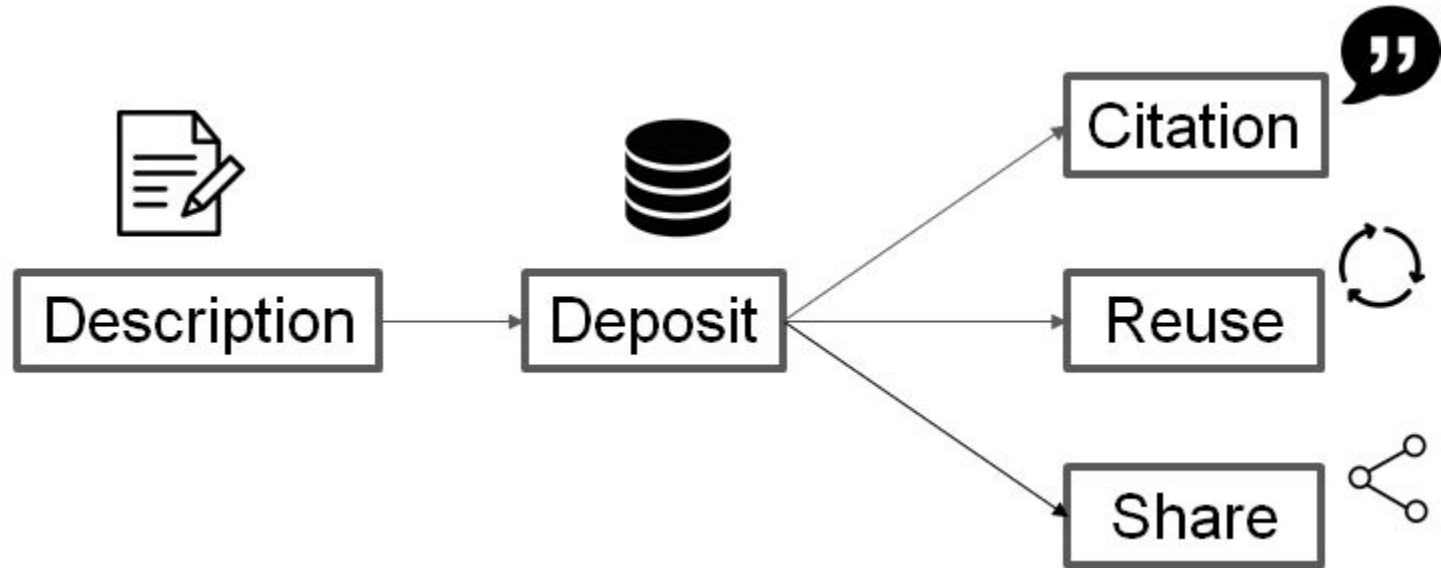
Done with the curators that provided those ontologies.

They were given a selection of scientific documents to be used as input and were guided through the different phases of the tool while providing feedback accordingly.

Curators agree that a tool like Dendro Keywords is a good addition to their work.

Use cases for data deposit *(Yulia Karimova, Joana Rodrigues)*





Groups

What are Groups?

You can use CKAN Groups to create and manage collections of datasets. This could be to catalogue datasets for a particular project or team, or on a particular theme, or as a very simple way to help people find and search your own published datasets.

Search groups...

6 groups found

Order by: Name Ascending



CETAPS - Centre for English, Translation, and Anglo-Portuguese Studies

CETAPS (Centre for English, Translation, and Anglo-Portuguese Studies) is a...

1 Dataset



FEUP - Faculty of Engineering of the University of Porto

The Faculty of Engineering of the University of Porto undertakes activities...



CIBIO - Research Centre in Biodiversity and Genetic Resources

CIBIO is a research centre in Biodiversity and Genetic Resources and the goal...

0 Datasets



I3S - Instituto de Investigação e Inovação em Saúde

The I3S consortium headed...



CPUP - Center for Psychology at University of Porto

The Center for Psychology at University of Porto (CPUP) brings together...

1 Dataset



FCNUP - Faculty of Nutrition and Food Sciences of the University of Porto

The Faculty of Nutrition and Food Sciences of the University of Porto is one...

3 Datasets

Groups / I3S - Instituto de Investigação e Inovação em Saúde



I3S - Instituto de Investigação e Inovação em Saúde

The I3S consortium, headed by the UPorto, brings together four institutions and researchers from several schools of the UPorto, thus consolidating an extensive collaboration... read more

Followers 0 Datasets 2

Organizations

UP - University of ... (2)

Groups

I3S - Instituto de ... (2)

Datasets Activity Stream About

Search datasets...

2 datasets found

Order by: Relevance

Retrieved data for study published on BMJ Open: "Methodological standards, qu...

Excel file with four sheets, the first with data number-coded for SPSS, the second with data labels, the third with data for group size calculation for the preclinical studies...

*.xlsx

Questionnaire data for paper "Researchers' attitudes to the 3Rs – an upturned...

SPSS data file (.sav) with anonymized data concerning questionnaire answers by laboratory animal science course participants

*.sav

Retrieved data for study published on BMJ Open: "Methodological standards, quality of reporting, and regulatory compliance in animal research on amyotrophic lateral sclerosis: a systematic review"

Followers

0

Organization

U. PORTO

UP - University of Porto

Founded in 1911, the University of Porto (U.Porto) is a benchmark institution for Higher Education and Scientific Research in Portugal and one of the top 200 European... read more

Social

Google+

Twitter

Facebook

Dataset Groups Activity Stream

Retrieved data for study published on BMJ Open: "Methodological standards, quality of reporting, and regulatory compliance in animal research on amyotrophic lateral sclerosis: a systematic review"

Excel file with four sheets, the first with data number-coded for SPSS, the second with data labels, the third with data for group size calculation for the preclinical studies subset, and a fourth with data gathered to estimate trends in animal model of choice in ALS research.

Data and Resources



Animal studies on ALS odd years from 2005 to 2015

Excel file with four sheets, the first with data number-coded for SPSS, the...

Explore

ALS amyotrophic lateral... animal research animal testing animal welfare guidelines methods systematic review

Additional Info

| Field | Value |
|-----------------|--|
| Author | Nuno Henrique Franco |
| Last Updated | March 8, 2019, 11:18 AM (UTC+00:00) |
| Created | March 6, 2019, 2:05 PM (UTC+00:00) |
| dc.Author | Nuno Henrique Franco, Joana G Fernandes, Andrew J Grierson, Jan Hultgren, Andrew JW Furley e I. Anna S. Olsson |
| dc.Contributor | i3S - Instituto de Investigação e Inovação em Saúde |
| dc.Date Created | 2013-2016 |
| dc.Format | xlsx file |
| dc.Language | English |

License

Creative Commons Attribution Share-Alike [OPEN DATA](#)

| | |
|---------------------------------|--|
| dc.Publisher | University of Porto Data Repository |
| dc.Relation | Scientific paper: "Methodological standards, quality of reporting, and regulatory compliance in animal research on amyotrophic lateral sclerosis: a systematic review". BMJ |
| dc.Spatial Coverage | Papers published on the SOD-1 mouse model of ALS in odd years between 2005 and 2015, retrieved from the ISI Web of Science database |
| dc.Temporal Coverage | Search performed in Feb 2013 for scientific articles from 2009 and 2011, in August 2013 for scientific articles from 2005, in September 2014 for scientific articles from 2013, in November 2016 for scientific articles from 2015, and in February 2017 for scientific articles from 2007 |
| dc.Type | Data systematically retrieved from reading sample of ALS papers, following predefined data extraction protocol |
| ddi.Data Collection Methodology | Each published study was categorised as either a 'preclinical' (i.e., carried out "to evaluate a drug for use in humans") or 'proof-of-concept' (i.e., aiming "to elucidate the mechanism of the disease"), according to the suggested classification for animal studies on ALS 2.38. Table 1 describes the information retrieved regarding regulatory compliance, animal models, experimental design and animal welfare. This information was retrieved through careful reading of the full papers. The review protocol was defined prior to data collection. No modifications to data collection methods were made during the research, but the period to be covered was extended to include publication year 2015 |
| ddi.Instrument Name | "ISI Web of Science", currently "Clarivate analytics Web of Science" |
| ddi.Research Domain | Systematic review of methodological standards and animal welfare in animal research on ALS |
| ddi.Sample | Scientific articles reporting studies on the SOD-1 mouse model of ALS (N=569) |
| ddi.Sampling Procedure | The sample comprises scientific articles (published in English), retrieved by an advanced search carried out on the ISI Web of Science® database with the query <<TS = ((mice OR mouse) SAME (ALS OR "amyotrophic lateral sclerosis"))>> |
| ddi.Type of Instrument | Scientific citation indexing service search engine and database |

What are Groups?

You can use CKAN Groups to create and manage collections of datasets. This could be to catalogue datasets for a particular project or team, or on a particular theme, or as a very simple way to help people find and search your own published datasets.



4 groups found

Order by: Name Ascending ▼



CS: Computer Science

The Computer Science Cluster mission is to contribute to the understanding of...



II: Industry and Innovation

The Industry and Innovation Cluster is the aggregation of INESC TEC research...



NIS: Networked Intelligent Systems

The Networked Intelligent Systems Cluster aims to create autonomous networked...



PE: Power and Energy

The Power and Energy Cluster aims to assure the

Organizations

INESC TEC (23)

Groups

CS: Computer Science (14)

NIS: Networked Inte... (7)

II: Industry and in... (1)

Tags

environmental radio... (5)

atmosphere (5)

radon (4)

portuguese (4)

text mining (3)

named entity recogn... (3)

gamma radiation (3)

vision (2)

news (2)

models (2)

Show More Tags

Formats

[Add Dataset](#)



23 datasets found

Order by: Relevance ▼

Simple English Wikipedia Link Graph with Clickstream Transitions 2018-12

The Simple English Wikipedia Link Graph with Clickstream Transitions is a gzipped GML file representing the hyperlink graph of the Simple English Wikipedia. It was prepared...

[gml.gz](#) [RData](#) [TXT](#)

Dataset for kidney exchange problems

The common practice is to model the Kidney Exchange Problem (KEP) on directed graph $G = (V, A)$, called compatibility graph, where set of vertices V corresponds to the set of...

[ZIP](#)

Gamma radiation from INESC TEC station (Porto)

The dataset consists on measurements of the total number of gamma rays counted by a NaI(Tl) scintillator on the roof of INESC TEC main building. This dataset has Jupyter...

[TXT](#) [pynb](#)

Atmospheric electric field from INESC TEC station (Porto)

The dataset consists on 1-min measurements of the atmospheric electric field by a CS110 field mill installed on the roof of INESC TEC main building. This dataset has Jupyter...

[TXT](#) [pynb](#)

Simple English Wikipedia Link Graph with Clickstream Transitions 2018-12

The Simple English Wikipedia Link Graph with Clickstream Transitions is a gzipped GML file representing the hyperlink graph of the Simple English Wikipedia. It was prepared using the "pagelinks" and "page" SQL dumps for 2019-01-01 and extended with an edge property called "transitions" based on the Clickstream dump for the English Wikipedia from 2018-12. It was designed to be used as a ground truth to evaluate node ranking metrics, like PageRank, but it can be useful for Network Science in general, or for Machine Learning and Information Retrieval to compute features over a medium-sized, complete Wikipedia link graph.

Data and Resources



simplewiki_link_graph-201812.gml.gz

A gzipped GML of the Simple English Wikipedia hyperlinks among pages in the...

[Explore](#)


simplewiki_link_graph-201812.RData

A serialized R object named 'g', based on igraph 1.2.2, with a preloaded...

[Explore](#)


README

Text file with the general description of the dataset, its data preparation...

[Explore](#)
[Clickstream](#)
[Graph](#)
[Graph-Based Metrics](#)
[Hyperlinks](#)
[Network Analysis](#)
[Node Centralities](#)
[Node Ranking](#)
[Simple English](#)
[Transitions](#)
[Wikipedia](#)

Additional Info

| Field | Value |
|----------------------|---|
| Source | "https://dumps.wikimedia.org/enwiki/20181101/enwiki-20181101-pagelinks.sql.gz https://dumps.wikimedia.org/enwiki/20181101/enwiki-20181101-page.sql.gz https://dumps.wikimedia.org/other/clickstream/2018-12/clickstream-enwiki-2018-12.tsv.gz " |
| Author | José Devezas |
| Last Updated | 1 de Outubro de 2019, 14:27 (UTC+01:00) |
| Created | 6 de Março de 2019, 10:23 (UTC+00:00) |
| Cite As | DEVEZAS, José, NUNES, Sérgio. Simple English Wikipedia Link Graph with Clickstream Transitions 2018-12 [dataset]. 06 mar 2019. INESC TEC research data repository. DOI: https://doi.org/10.25747/83vk-zt74 |
| DOI | https://doi.org/10.25747/83vk-zt74 |
| dc.License | GNU Free Documentation License (GFDL)+ CC BY-SA 3.0 (https://dumps.wikimedia.org/legal.html) |
| dc.Contributor | Sérgio Nunes |
| dc.Coverage.Temporal | 2018-12 |
| dc.Date | 2019-02-01T12:04 |
| dc.Format | Graph Modeling Language (GML); GZip |
| dc.Format.Extent | Total: 166 MB; GML: 35 MB (compressed); 897,577 nodes; 6,986,460 edges; RData: 131 MB (igraph 1.2.2; 'g' variable) |
| dc.Language | EN |
| dc.Publisher | INESC TEC |
| dc.Type | Simple English Wikipedia Link Graph for network analysis. |
| ddi.Software | R; Gephi; igraph; NetworkX |

Connecting with EUDAT B2SHARE *Yulia Karimova*



[HELP](#) [COMMUNITIES](#) [UPLOAD](#) [CONTACT](#)

[Login](#)

RECORDS » 7C429C0EC56D4DB590FAA7AEDA7D7C22

Latest Version - Dec 22, 2016

Dendro lightweight ontologies

by [João Aguiar Castro](#); [João Rocha da Silva](#); [Cristina Ribeiro](#); [Ricardo Amorim](#); [Yulia Karimova](#); [Rubia Gattelli](#);

Dec 22, 2016

Last updated at Jan 11, 2018

Abstract: A collection of domain-specific lightweight ontologies for research data description. These ontologies can be combined with each other, or with others, like Dublin Core or FOAF, for a more comprehensive description. The ontologies were created in collaboration with researchers from scientific domains such as Computational Fluid Dynamics, Analytical Chemistry, Biodiversity among others.

Keywords: [Ontology](#); [Research Data Management](#); [Data Description](#); [Dendro](#);

PID: [11304/6e54d7ca-8a24-11e5-gbb4-2b0aad496318](#)



Files

| Name | Size |
|---|---------|
| Dendro_lightweightOntologies.zip | 26.59KB |
| Checksum: md5: 3f642432b41493a63b82f8a95918d7d4 | |
| PID: 11304/66257555-b7e7-4e16-ab28-cfc2f8c2db52 <input type="button" value="Copy"/> | |

Basic metadata

| | |
|------------------|--|
| Open Access | True <input checked="" type="checkbox"/> |
| License | Creative Commons Attribution-NonCommercial-ShareAlike (CC-BY-NC-SA) |
| Contact Email | joaoaguiarcastro@gmail.com |
| Publication Date | 2015-11-13 |

Connecting with EUDAT B2NOTE, *Yulia Karimova*

HAREM NER Models for OpenNLP, Stanford CoreNLP, spaCy,

by [André Pires](#);

Jun 22, 2017

Keywords: [named entity recognition](#); [models](#); [text mining](#); [portuguese](#);

DOI: [XXXX/b2share.a4906773dc1f42f882bd03be0c9846c3](#) [Copy](#)

PID: [0000/a4906773dc1f42f882bd03be0c9846c3](#) [Copy](#)

| Files | |
|---|----------|
| Name | Size |
| > HAREM NER MODELS.json | 7.01KB |
| > HAREM NER MODELS.rdf | 2.04KB |
| > HAREM NER MODELS.txt | 2.10KB |
| > HAREM NER MODELS.zip | 111.61MB |
| > nltk.zip | 5.76MB |
| ▼ open-nlp.zip | 2.41MB |
| Checksum: md5: 677f1c21c83c7fbf5420626f5370a447 | |
| PID: 0000/open-nlp.zip Copy | |
| Annotate in BzNote 1 | |
| > spacy.zip | 15.72MB |
| > stanford-corenlp.zip | 87.75MB |

| Basic metadata | |
|----------------|-------|
| Open Access | True |
| Publisher | INESC |
| Language | PT |

B2NOTE
Yulia

Semantic tag | Free-text keyword

Comment

Type-in to select a tag. [Create](#)

▼ All my annotations

▼ All annotations about this file

[Let us know what you think](#)

[Edit Record](#)

[Report Abuse](#)

open-nlp.zip
Checksum: md5:677f1c21c83c7fbf5420626f5370a447
PID: 0000/open-nlp.zip
Annotate in BzNote 1

B2NOTE
Yulia

Semantic tag | Free-text keyword

Comment

Type-in to select a tag. Create

All my annotations

All annotations about this file

Let us know what you think

2

B2NOTE
Yulia

The comment annotation was created. OK

NER models for OpenNLP, for the categories, types and subtypes of HAREM, with the best

All my annotations

All annotations about this file

3

B2NOTE
Yulia

Type-in to select a tag. Create

All my annotations

- Semantic tag 2
- Free-text keyword 3
- Comment 2

All annotations about this file

- Semantic tag 0 0
- Free-text keyword 0 0
- Comment 0 1

Let us know what you think

4

B2NOTE
Yulia

No semantic tag annotation exists about this file.

Free-text keyword ^{^top}

No free-text keyword annotation exists about this file.

Comment ^{^top}

| Label | Actions |
|-------------------------|---------|
| NER models for OpenN... | |

5

B2NOTE Yulia

Search

Free-text keyword

OpenNLP

+

Q

6

B2NOTE Yulia

Export search results

Query: OpenNLP

New search Generate JSON export

Exact match

- Files (select all)
- <https://trng-b2share.eudat.eu/api/files/aa3b2653-00b5-44a3-a540-2787802a4d25/open-nlp.zip>

Related match (synonyms)

No file retrieved on synonym match.

Download this as JSON-LD 8 Download as RDF/XML

7

B2NOTE nelsonpereira1991

Comment

This is a really useful dataset. I recommend it.

By: nelsonpereira1991

Used by me to annotate this file:

<https://trng-b2share.eudat.eu/api/files/aa3b2653-00b5-44a3-a540-2787802a4d25/open-nlp.zip>

9

B2NOTE nelsonpereira1991

Annotations

Semantic tag 1 0

Free-text keyword 1 1

Comment 1 1

Semantic tag ^{top}

| Label | Nb Files | Actions |
|-----------------|----------|---------|
| Models, Genetic | 1 | |

Free-text keyword ^{top}

| Label | N Files | Actions |
|---------|---------|---------|
| CoreNLP | 1 | |
| OpenNLP | 1 | |

Comment ^{top}

| Label | Actions |
|----------------------------|---------|
| This is a really use... 10 | |
| NER models for OpenNLP | |

Edit is not possible. This annotation was not created by me.

HAREM NER Models for OpenNLP, Stanford CoreNLP, spaCy, NLTK

by André Pires;

Jun 22, 2017

Keywords: named entity recognition; models; text mining; portuguese;

DOI: [XXXX/b2share.a4906773dc1f42f882bd03be0c9846c3](https://doi.org/10.1111/b2share.a4906773dc1f42f882bd03be0c9846c3)

PID: [0000/a4906773dc1f42f882bd03be0c9846c3](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63862-p0000-a4906773dc1f42f882bd03be0c9846c3)

Files

| Name | Size |
|--|----------|
| ▶ HAREM NER MODELS.json | 7.01KB |
| ▶ HAREM NER MODELS.rdf | 2.04KB |
| ▶ HAREM NER MODELS.txt | 2.10KB |
| ▶ HAREM NER MODELS.zip | 111.61MB |
| ▶ nltk.zip | 5.76MB |
| ▼ open-nlp.zip | 2.41MB |
| Checksum: md5:677f1c21c83c7fbf5420626f5370a447 | |
| PID: 0000/open-nlp.zip <input type="button" value="Copy"/> | |
| <input type="button" value="Annotate in B2Note"/> | |
| ▶ spacy.zip | 15.72MB |
| ▶ stanford-corenlp.zip | 87.75MB |

Basic metadata

| | |
|-------------|--|
| Open Access | True <input checked="" type="checkbox"/> |
| Publisher | INESC TEC |
| Language | PT |

B2NOTE
nelsonpereira1991

Annotations

Semantic tag: 1 0

Free-text keyword: 1 1

Comment: 1 1

Semantic tag ^{^top}

| Label | Nb Files | Actions |
|-----------------|----------|---|
| Models, Genetic | 1 | <input type="button" value="edit"/> <input type="button" value="delete"/> |

Free-text keyword ^{^top}

| Label | N Files | Actions |
|---------|---------|---|
| CoreNLP | 1 | <input type="button" value="edit"/> <input type="button" value="delete"/> |
| OpenNLP | 1 | <input type="button" value="edit"/> <input type="button" value="delete"/> |

Comment ^{^top}

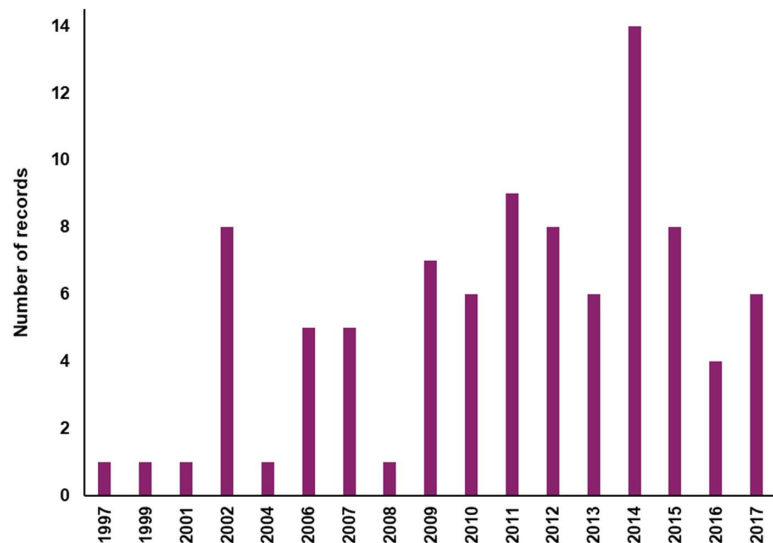
| Label | Actions |
|-------------------------|---|
| This is a really use... | 5 <input type="button" value="edit"/> <input type="button" value="delete"/> |
| NER models for OpenNLP | <input type="button" value="edit"/> <input type="button" value="delete"/> |

Edit is not possible. This annotation was not rated by r.o.e.

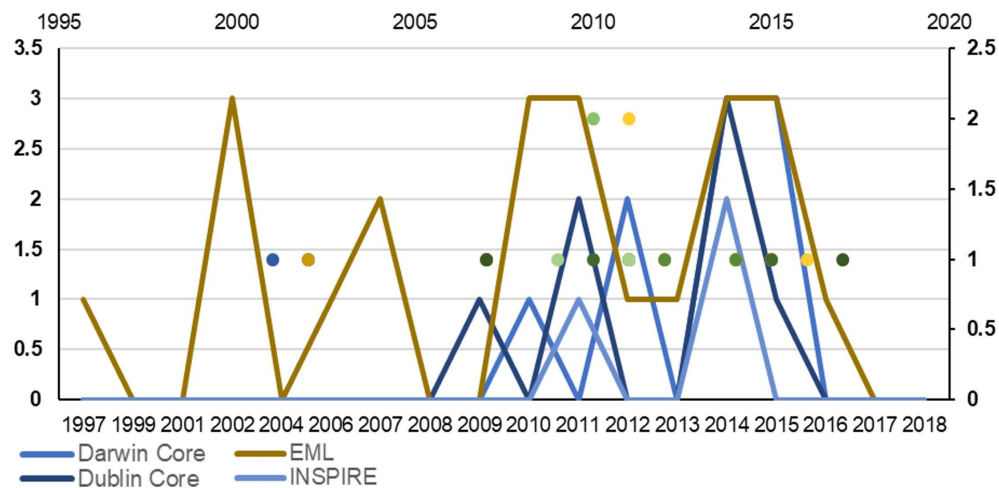
Research data management in the domain of Ecology (Ângela Lomba)

- Meta-analysis of existing metadata models in the domain of Ecology

Temporal overview



Diversity of tools



Research data management in the domain of Ecology (Ângela Lomba)

- Meta-analysis of existing metadata models in the domain of Ecology

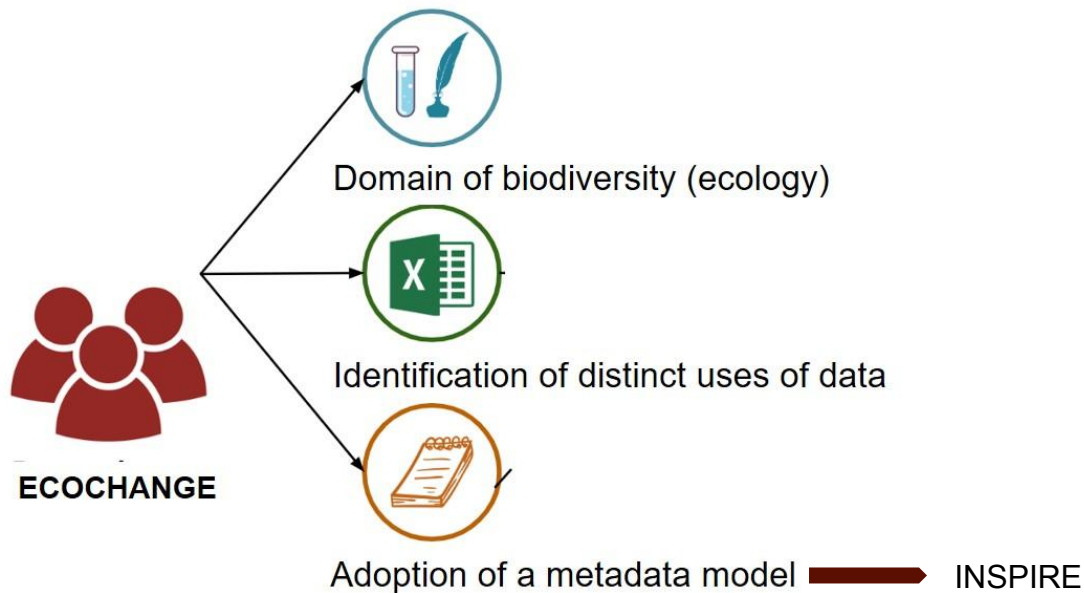
| Platforms/Projects | Metadata Standard |
|---|-------------------|
| Long-term Ecological Research (LTER) | EML |
| The Knowledge Network for Biocomplexity (KNB) | EML |
| Taxonomic Databases Working Group (TDWG) | |
| Map of Life Project | Darwin Core |
| Global Biodiversity Information Facility (GBIF) | |

The screenshot shows the homepage of The Long Term Ecological Research Network. It features a navigation menu with options like NETWORK, SITES, RESEARCH, EDUCATION, DATA, HISTORY, and SCIENCE UPDATE. Below the navigation, there is a section for 'LTER Site Profiles' with a brief description and contact information for the LTER Network Communications Office. A map of the United States is displayed, showing the locations of various LTER sites. To the right of the map, a list of LTER sites is provided, including Andrews Forest LTER (AND), Arctic LTER (ARC), Baltimore Ecosystem Study (BES), Beaufort Lagoon Ecosystem (BLE), Bonanza Creek LTER (BNZ), California Current Ecosystem LTER (CCE), Cedar Creek Ecosystem Science Reserve (CDR), Central Arizona - Phoenix LTER (CAP), Coweeta LTER (CWT), Florida Coastal Everglades LTER (FCE), Georgia Coastal Ecosystems LTER (GCE), Harvard Forest LTER (HFR), Hubbard Brook LTER (HBR), Jornada Basin LTER (JBN), Kolloid Biological Station LTER (KBS), Konza Prairie LTER (KNZ), LTER Network Communications Office (NCO), Leopold LTER (LEO), McHardo Dry Valley LTER (MCH), and Moorea Coral Reef LTER (MCR).

Below the map and list, there is a section titled 'What is GBIF?' with a sub-header 'GBIF—the Global Biodiversity Information Facility—is an open data research infrastructure funded by the world's governments and aimed at providing anyone, anywhere access to data about all types of life on Earth.' Below this text is a photograph of a colorful macaw parrot standing in a grassy field.

Research data management in the domain of Ecology *(Ângela Lomba)*

- The case of the ECOCHANGE research group



Research data management in the domain of Ecology *(Ângela Lomba)*

- The case of the ECOCHANGE research group / the catalog

| | C | D | E | F | G | H | I |
|---|---|---|----------------|--------------------|-----------------|---------------------|-----------------------------|
| 1 | Resource title | Resource abstract | Topic category | Spatial resolution | Temporal extent | Date of publication | Geographic coordinates |
| 2 | Statistical units | Units for dissemination or use of statistical information. | | | | | |
| 3 | BGRI 2001 | Portuguese Basis for Geographic Information Referencing for the north region of Portugal. Information from de Portuguese National Statistical Institute (INE). | Boundaries | 25k | | 2001 | |
| 4 | BGRI 2001 [SP] | Portuguese Basis for Geographic Information Referencing for the north region of Portugal. Information from de Portuguese National Statistical Institute (INE). | boundaries | 1:25000 | 01/01/2001 | 01/01/2001 | -8.4 41.6 |
| 5 | Buildings | Geographical location of buildings. | | | | | |
| 6 | Buildings | Buildings for the PNPG work area. Information from the official National M888 Series of Military Map of Portugal. | Structure | 25k | 1936 - 1937 | 1937 | |
| 7 | Other buildings | Other buildings for the PNPG work area. Information from the official National M888 Series of Military Map of Portugal (churches, chapels, mills, ruins, etc.). | Structure | 25k | 1936 - 1937 | 1937 | |
| 8 | Buildings and interest points for the Melgaço municipality | Contains information regarding buildings and interest points, namely: lodging sites, hotels, touristic infrastructures, culture and local heritage, commerce, religious sites, leisure and other infrastructures. | Structure | 10k | 2010 | 2010 | -163 143 2753 2531 |
| | Heritage value maps for the Peneda-Gerês National Park (PNPG) | Heritage value maps for the Peneda-Gerês National Park (PNPG) | Structure | 25k | 2008 | 2008 | 1773 522 |

- Diversity of uses of data, data formats, scales, geographic projections;

- Heterogeneity in the way data producers describe similar information;

-Lack of relevant information (==blank fields);





Wrap-up, *Gabriel David, Ângela Lomba*

TAIL contributions to Research Data Management

*“The project has **two well-identified lines of work**. The first is **technological** and involves the selection and integration of tools for the TAIL data management workflow. The required developments are strong in tool integration and deployment, taking the DENDRO platform for data description, the LabTablet metadata collection app and data repository platforms to provide a friendly data management environment for researchers. The second is **experimental** and uses the TAIL environment and a panel of researchers to perform data management experiments where existing datasets are selected, described and deposited.”*

*“By the end of the 3-year project, there will be a panel of **research groups** that have successfully **managed the datasets** they were creating and who also reaped the benefits of **having their data published, findable and cited**. Such **success stories** are still lacking in the community and their lessons are invaluable for the progress of research data management.”*