

Por trás do conteúdo livre: o Wikidata enquanto banco de dados livres

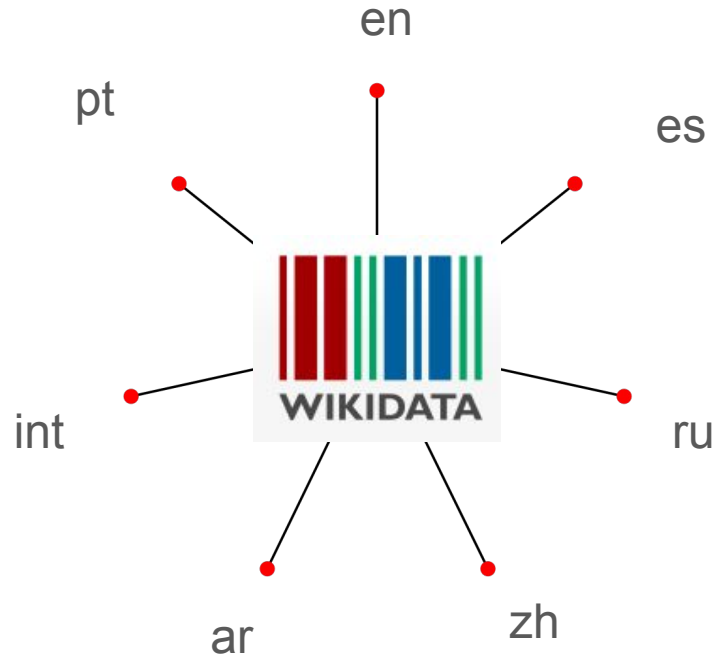
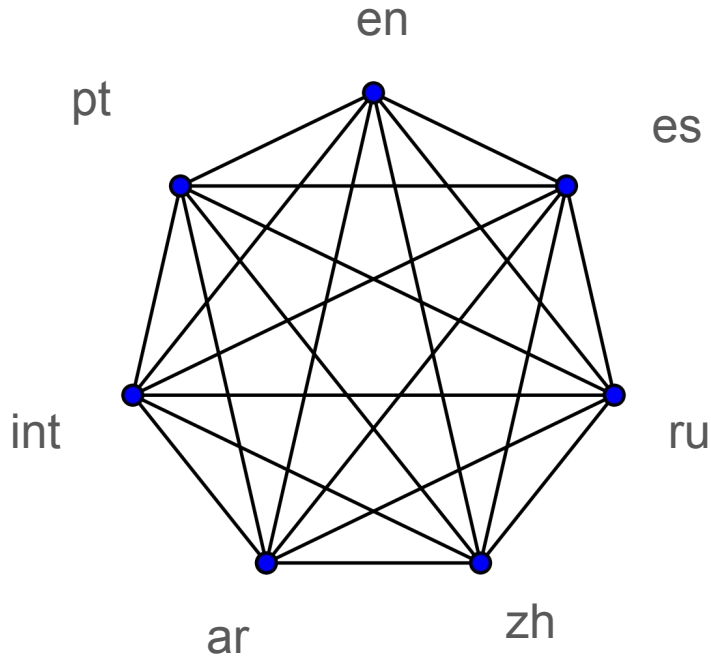
Tiago Lubiana

XI Oficina Wikimedia & Educação

21 de Julho de 2021



O que é o Wikidata?



O que é o Wikidata

The image shows a Wikidata entry for 'The Human Cell Atlas: from vision to reality.' with several annotations:

- label**: Points to the title 'The Human Cell Atlas: from vision to reality.'
- identifier**: Points to the Wikidata ID '(Q47565008) ...'
- property**: Points to the 'instance of' property in the 'Statements' section.
- value**: Points to the value 'scholarly article ...' of the 'instance of' property.
- reference**: Points to the '1 reference' section, which contains a table of bibliographic information.
- statement**: Points to the entire 'Statements' section.

Additional text in the interface includes: 'scientific article published in October 2017', 'Recoim: Most relevant properties which are absent', 'In more languages', and buttons for '+ add reference' and '+ add value'.

property	value
instance of	scholarly article ...

reference	
stated in	Europe PubMed Central
PubMed ID	29072289
retrieved	4 February 2018
reference URL	http://europepmc.org/abstract/MED/29072289

O que são dados?

Podemos pensar em três tipos:

- Dados como dados
 - O primeiro genoma humano completo
 - Os dados coletados pelo Síncrotron Sirius, em Campinas

O que são dados?

Podemos pensar em três tipos:

- Dados como dados
 - O primeiro genoma humano completo
 - Os dados coletados pelo Síncrotron Sirius, em Campinas
- Informação como dados
 - O ano de nascimento de Raul Seixas
 - A população atual de Luanda

O que são dados?

Podemos pensar em três tipos:

- Dados como dados
 - O primeiro genoma humano completo
 - Os dados coletados pelo Síncrotron Sirius, em Campinas
- Informação como dados
 - O ano de nascimento de Raul Seixas
 - A população atual de Luanda
- Conhecimento como dados
 - A velocidade da luz
 - O ponto de ebulição da água

O que são dados?

Podemos pensar em três tipos:

- Dados como dados
 - O primeiro genoma humano completo
 - Os dados coletados pelo Síncrotron Sirius, em Campinas
- **Informação como dados**
 - O ano de nascimento de Raul Seixas
 - A população atual de Luanda
- **Conhecimento como dados**
 - A velocidade da luz
 - O ponto de ebulição da água

O que são dados *livres*?

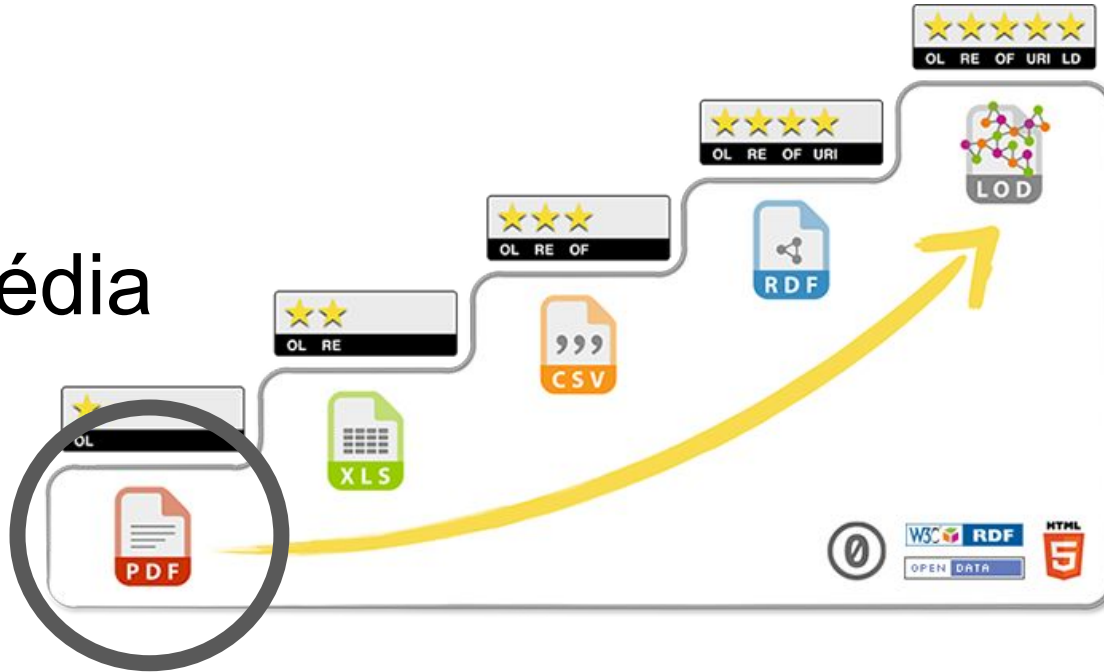
dados 5 estrelas



<https://5stardata.info/pt-BR/>

O que são dados *livres*?

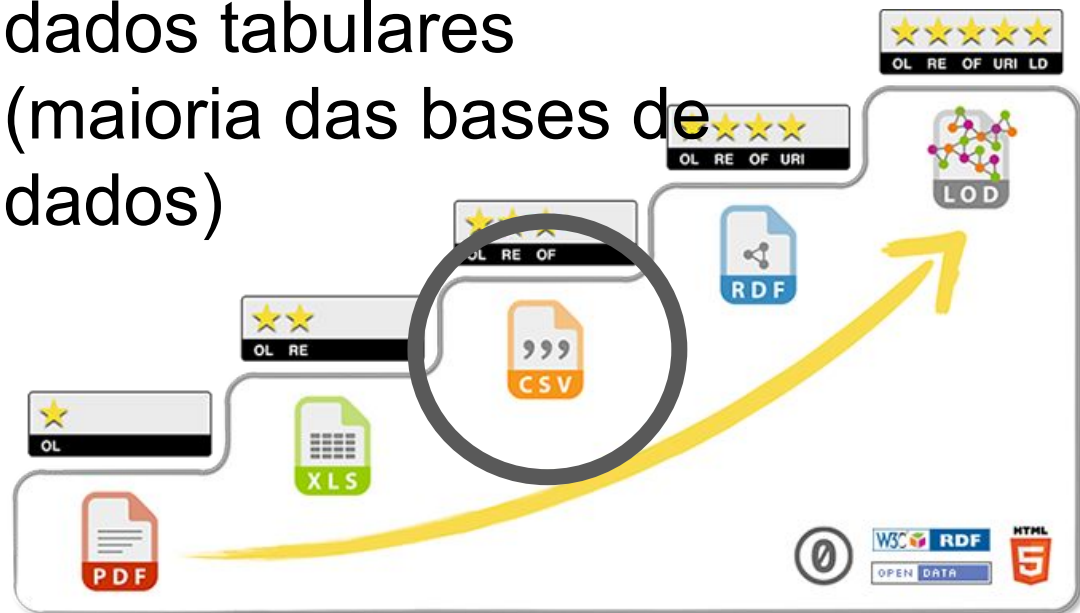
Wikipédia



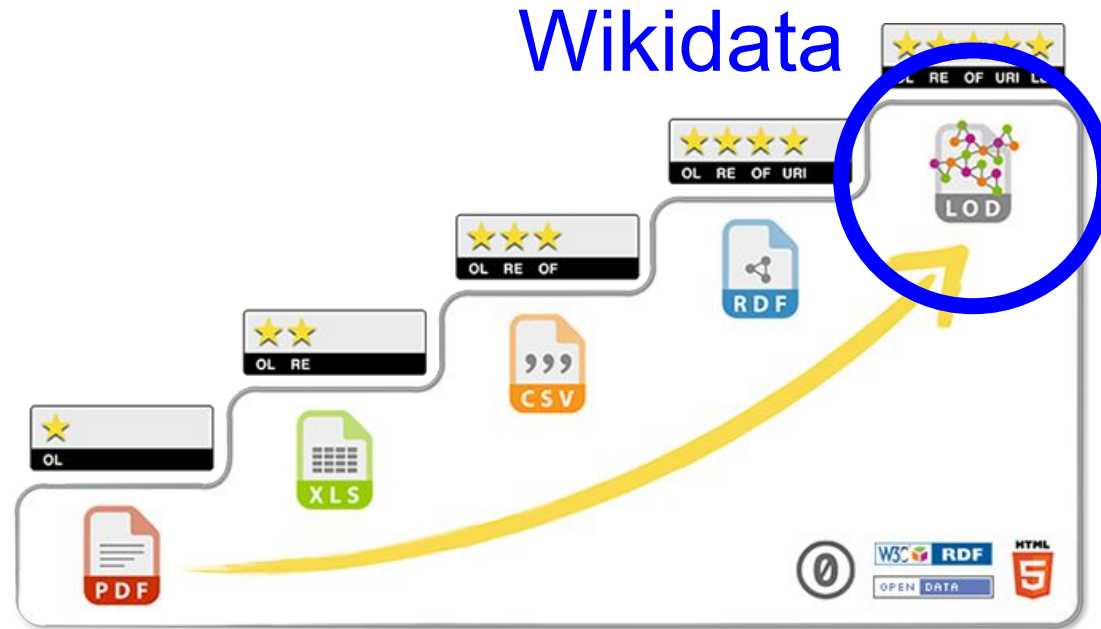
<https://5stardata.info/pt-BR/>

O que são dados *livres*?

dados tabulares
(maioria das bases de
dados)



O que são dados *livres*?



A academia está conectada ao Wikidata

Autores de artigos sobre o Wikidata:



<https://searchy.toolforge.org/search/Q2013>

Wikidata na literatura biomédica



Computational and Systems Biology



Science Forum: Wikidata as a knowledge graph for the life sciences



Andra Waagmeester, Gregory Stupp, Sebastian Burgstaller-Muehlbacher, Benjamin M Good, Malachi Griffith, Obi L Griffith, Kristina Hanspers, Henning Hermjakob, Toby S Hudson [see all](#) »

Methodology article | [Open Access](#) | Published: 22 January 2021

A protocol for adding knowledge to Wikidata: aligning resources on human coronaviruses

[Andra Waagmeester](#), [Egon L. Willighagen](#), [Andrew I. Su](#), [Martina Kutmon](#), [Jose Emilio Labra Gayo](#), [Daniel Fernández-Álvarez](#), [Quentin Groom](#), [Peter J. Schaap](#), [Lisa M. Verhagen](#) & [Jasper J. Koehorst](#)

Wikidata: A large-scale collaborative ontological medical database

[Houcemeddine Turki](#) ^a , [Thomas Shafee](#) ^b , [Mohamed Ali Hadj Taieb](#) ^c , [Mohamed Ben Aouicha](#) ^c , [Denny Vrandečić](#) ^d , [Diptanshu Das](#) ^e , [Helmi Hamdi](#) ^f

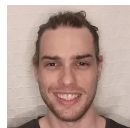
Scholia and scientometrics with Wikidata

Finn Årup Nielsen¹, Daniel Mietchen², and Egon Willighagen³

¹ Cognitive Systems, DTU Compute, Technical University of Denmark, Denmark
² EvoMRI Communications, Jena, Germany

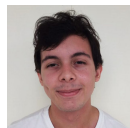
³ Dept of Bioinformatics - BiGCaT, NUTRIM, Maastricht University, The Netherlands

Wikidata for 5-star Linked Open Databases: A case study of PanglaoDB



Tiago Lubiana ([@lubianat](#))

Computational Systems Biology Laboratory - INOVA USP,
University of São Paulo, SP, Brazil



João Vitor Ferreira Cavalcante ([@jvfe](#))

Bioinformatics Multidisciplinary Environment,
Federal University of Rio Grande do Norte, RN, Brazil

ISMB/ECCB 2021
Bio-Ontologies COSI
July 25th, 2021

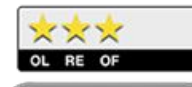
PanglaoDB e 5-star Linked Open Data

espécie	símbolo oficial do gene	Tipo celular
Mm Hs	ACHE	neurônio colinérgico

<http://www.wikidata.org/entity/Q101405051> (neurônio colinérgico humano)

<https://www.wikidata.org/entity/P8872> (é marcada por)

<https://www.wikidata.org/wiki/Q407983> (o gene ACHE humano)

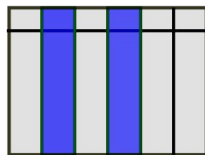


Migrando informação para o Wikidata

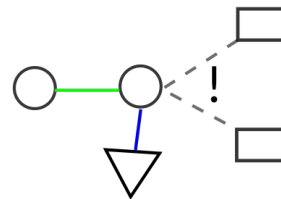
get permission
from developers



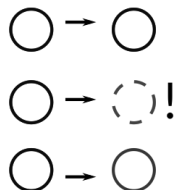
select fields to keep



choose/create
Wikidata properties



match/create
Wikidata classes



add to Wikidata
with citations



make integrative
SPARQL queries



get permission
from developers

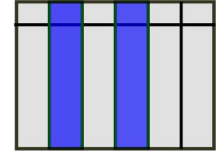
Permissão dos desenvolvedores



- Todo conteúdo no Wikidata é CC0 (Domínio Público)
- Fontes importadas em totalidade precisam estar disponíveis em licença compatível (CC0)

select fields to keep

Selecionando os campos

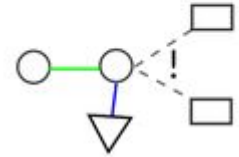


- Base de interesse PanglaoDB tinha 16 campos diferentes
- Somente 3 campos eram de fato interessantes para integrar:
 - Espécie
 - Tipo de célula
 - Símbolo oficial do gene

Mapeando/criando propriedades no Wikidata

- Há uma variedade de relações biomédicas no Wikidata
- Mais relações podem ser criadas se houver demanda

choose/create
Wikidata properties



subclasse de

tipo celular
neutro

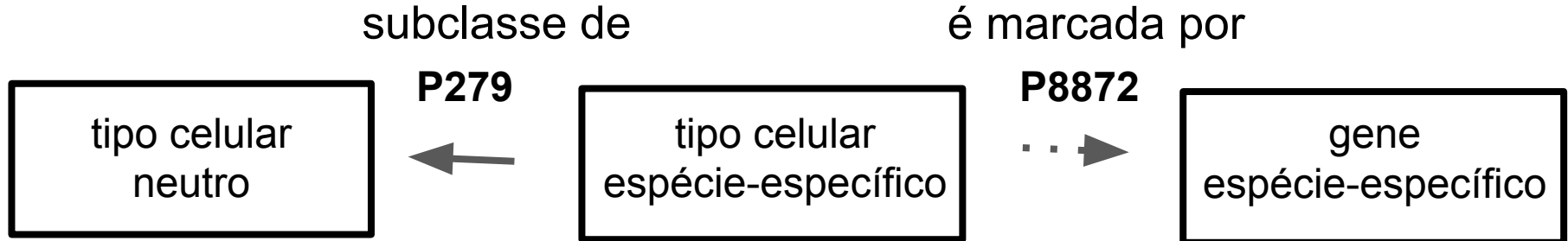
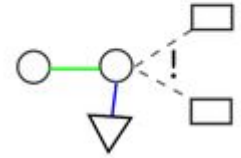
P279
←

tipo celular
espécie-específico

Mapeando/criando propriedades no Wikidata

- Há uma variedade de relações biomédicas no Wikidata
- Mais relações podem ser criadas se houver demanda

choose/create
Wikidata properties



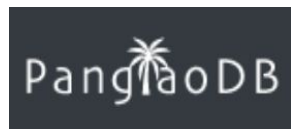
Mapeando/criando items no Wikidata

match/create
Wikidata classes

○-○

○-○!

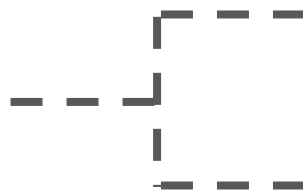
○-○



"B cells"



célula B
(Q188930)



célula B de camundongo
(Q104416288)

célula B humana
(Q68620797)

Adicionando declarações referenciadas

- 15656 declaração com o pacote Wikidata Integrator Python

human cholinergic neuron (Q101405051)

has marker

ACHE

▶ 1 reference

CHAT

▼ 1 reference

retrieved	27 November 2020
reference URL	https://panglaodb.se/markers.html
stated in	PanglaODB

add to Wikidata
with citations



<https://w.wiki/3ZGz>

Buscas SPARQL no Wikidata

relações tipo celular - doença



 human beta cell

↓ has marker

 ?gene

↑ genetic link to

 ?disease


COUNT

- 46 diseases
- 26 genes

Wikidata Query Service

Examples Help More tools

```
1 SELECT ?cellTypeLabel ?diseaseLabel
2 (COUNT(DISTINCT ?diseaseGene) AS ?count)
3 (GROUP_CONCAT(DISTINCT ?geneLabel; SEPARATOR=", ") AS ?genes)
4 WHERE
5 {
6   wd:Q101405087 wdt:P8872 ?diseaseGene .
7   ?disease wdt:P2293 ?diseaseGene .
8
9   wd:Q101405087 rdfs:label ?cellTypeLabel .
10  ?disease rdfs:label ?diseaseLabel .
11  ?diseaseGene rdfs:label ?geneLabel .
12
13  FILTER(LANG(?cellTypeLabel) = "en")
14  FILTER(LANG(?diseaseLabel) = "en")
15  FILTER(LANG(?geneLabel) = "en")
16
17 }
18
19 GROUP BY ?diseaseLabel ?cellTypeLabel
20 ORDER BY DESC(?count)
```




<https://w.wiki/3HjX>

Buscas SPARQL no Wikidata

relações tipo celular - doença



 **human beta cell**

 **has marker**

 **?gene**

 **genetic link to**

 **?disease**

COUNT

• 46 diseases

• 26 genes

tipo celular

doença

genes

célula beta
humana

obesidade

PCSK2, ADCYAP1,
SLC30A8

célula beta
humana

tipo 2 diabetes

SLC30A8, TGFBR3

célula beta
humana

Doença de Parkinson

SH3GL2

<https://w.wiki/3HjX>

Por trás do conteúdo: Wikidata como ponte educacional para a web semântica

Trunfos:

- Ligação direta com a Wikipédia
- Comunidade enorme super diversa em especialidades
- Interface para uso Web e programático
- Acomoda múltiplas línguas
- Muito aberto: chegar e editar

O Wikidata está inserido na academia brasileira

A Wikidata e os desafios da interoperabilidade na era dos dados abertos ligados na Web: uma breve reflexão

Linair Maria Campos

Universidade Federal Fluminense (UFF) - Brasil

Maria Luiza de Almeida Campos

Universidade Federal Fluminense (UFF) - Brasil

Nilson Theobald Barbosa

Universidade Federal Fluminense (UFF) - Brasil

Criando um grafo de conhecimento biológico através do Wikidata com foco no projeto Human Cell Atlas

Processo: 19/26284-1

Linha de fomento: [Bolsas no Bras](#)

Vigência (Início): 01 de agosto de 2019

Vigência (Término): 31 de julho de 2020

Área do conhecimento: [Interdisciplinar](#)

Applying a Multi-Level Modeling Theory to Assess Taxonomic Hierarchies in Wikidata

Freddy Brasileiro¹, João Paulo A. Almeida¹, Victorio A. Carvalho^{1,2}, Giancarlo Guizzardi¹

¹ Ontology & Conceptual Modeling Research Group (NEMO), Informatics Department
Federal University of Espírito Santo (UFES), Vitória, ES, Brazil

² Research Group in Applied Informatics, Informatics Department,
Federal Institute of Espírito Santo (IFES), Colatina, ES, Brazil

PAINEL DE INFORMAÇÃO SOBRE A COVID-19: CONSULTAS SPARQL NA WIKIDATA

Ana Carolina Simionato Arakaki, Fabiano Ferreira de Castro, Felipe Augusto Arakaki

As potencialidades de narrativas estruturadas para o Jornalismo Computacional: competências jornalísticas na elaboração de textos gerados com bancos de dados

A Study of the Quality of Wikidata

Kartik Shenoy^a, Filip Ilievski^{a,*}, Daniel Garijo^b, Daniel Schwabe^c and Pedro Szekely^a

^aInformation Sciences Institute, University of Southern California, USA

^bOntology Engineering Group, Universidad Politécnica de Madrid, Spain

^cDept. of Informatics, Pontifícia Universidade Católica Rio de Janeiro, Brazil

Érica Camillo Azzellini

Centro de Pesquisa, Inovação e Difusão em Neuromatemática (CEPID NeuroMat)

João Alexandre Peschanski

Centro Interdisciplinar de Pesquisa (CIP) da Faculdade Cásper Líbero (FCL)

Fernando Jorge da Paixão

Instituto de Física Gleb Wataghin (IFGW) da Universidade Estadual de Campinas (UNICAMP)

Volume 12
Número 1
janeiro-abril
2019

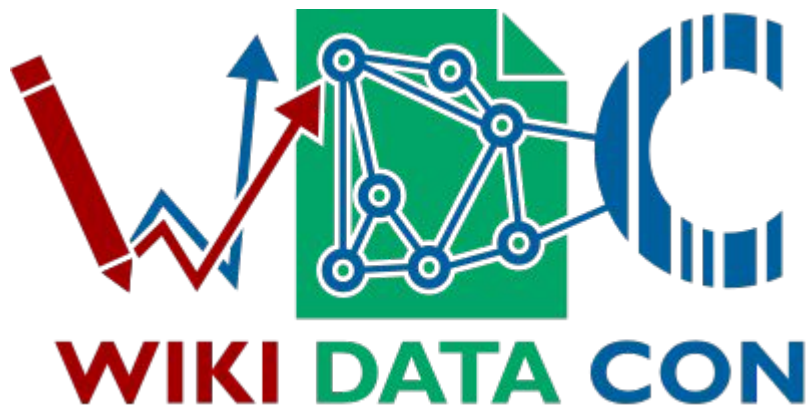


Universidade de Marília
Marília, SP
CEP 13164-900

Textos em
Linguagem e Tecnologia

UFMG

O Wiki Movimento Brasil tem papel central no Wikidata



Organização do WikidataCon 2021
em parceria com WMDE

WIKIDATA LAB XXX

PROCESSO DE RETROALIMENTAÇÃO PARA O TAINACAN

Este treinamento pretende compartilhar recursos e capacidades para a integração do Wikidata com outros projetos Wikimedia, em particular Wikimedia Commons e Wikipédia, mas também organizações externas.

Neste evento, parte da série Wikidata Labs, haverá uma apresentação e uma série de atividades práticas e discussões sobre o processo de retroalimentação de metadados de instituições GLAM utilizando o Tainacan, apresentando as experiências e os desafios envolvidos.

O evento é voltado à capacitação da comunidade de

Organização

Grupo de trabalho
WMB
Wiki Movimento Brasil

Parceria

FUNDAÇÃO
TAINACAN

Realização

USP
FUSP

tainacan

Série de treinamentos premiada,
com bastante coisa em português

Por trás do conteúdo livre: o Wikidata enquanto banco de dados livres

Tiago Lubiana

XI Oficina Wikimedia & Educação
21 de Julho de 2021

