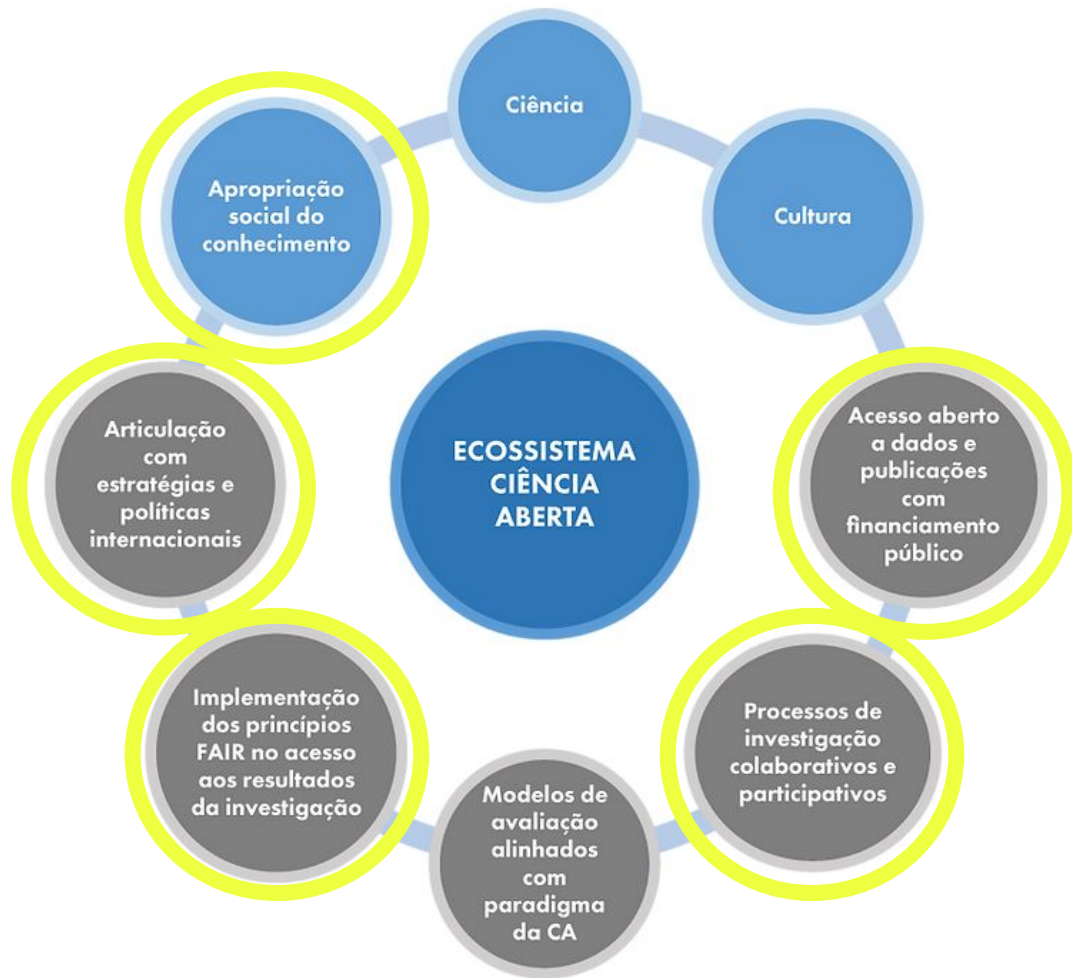




# Como usar a Wikidata em investigação: boas práticas de utilização e partilha de dados abertos na sociedade civil

Rute Correia, Associação Wikimedia Portugal/NOVA FCSH

30 de maio de 2023



Ecosystem of open science, illustrated by PNCA

## Gestão de conteúdo

Edição colaborativa, histórico de edições, flexibilidade de alojamento multimédia

## Gestão de dados e metadados

Interligação a bases de dados e identificadores externos, formatos interoperáveis, princípios FAIR



## Inovação aberta

Acesso aberto a dados e conteúdo; licenças livres e abertas para dados, conteúdo e software

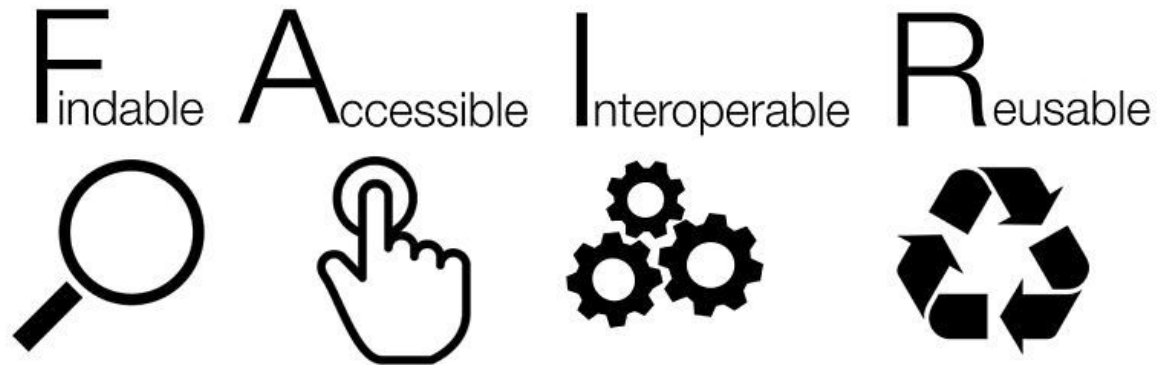
## Divulgação do conhecimento

Integração direta do conhecimento na sociedade sem *gatekeeping*, alargando as possibilidades de ciência cidadã

# Infraestrutura Wikimedia

solução integrada para a **incorporação da ciência e da investigação na sociedade**, seguindo os fundamentos da ciência aberta.





*Imagem de SangyaPundir (CC BY-SA, via Wikimedia Commons)*

A integração de dados no Wikidata está alinhada com os [Princípios FAIR](#) de gestão de dados científicos.

# Wikidata

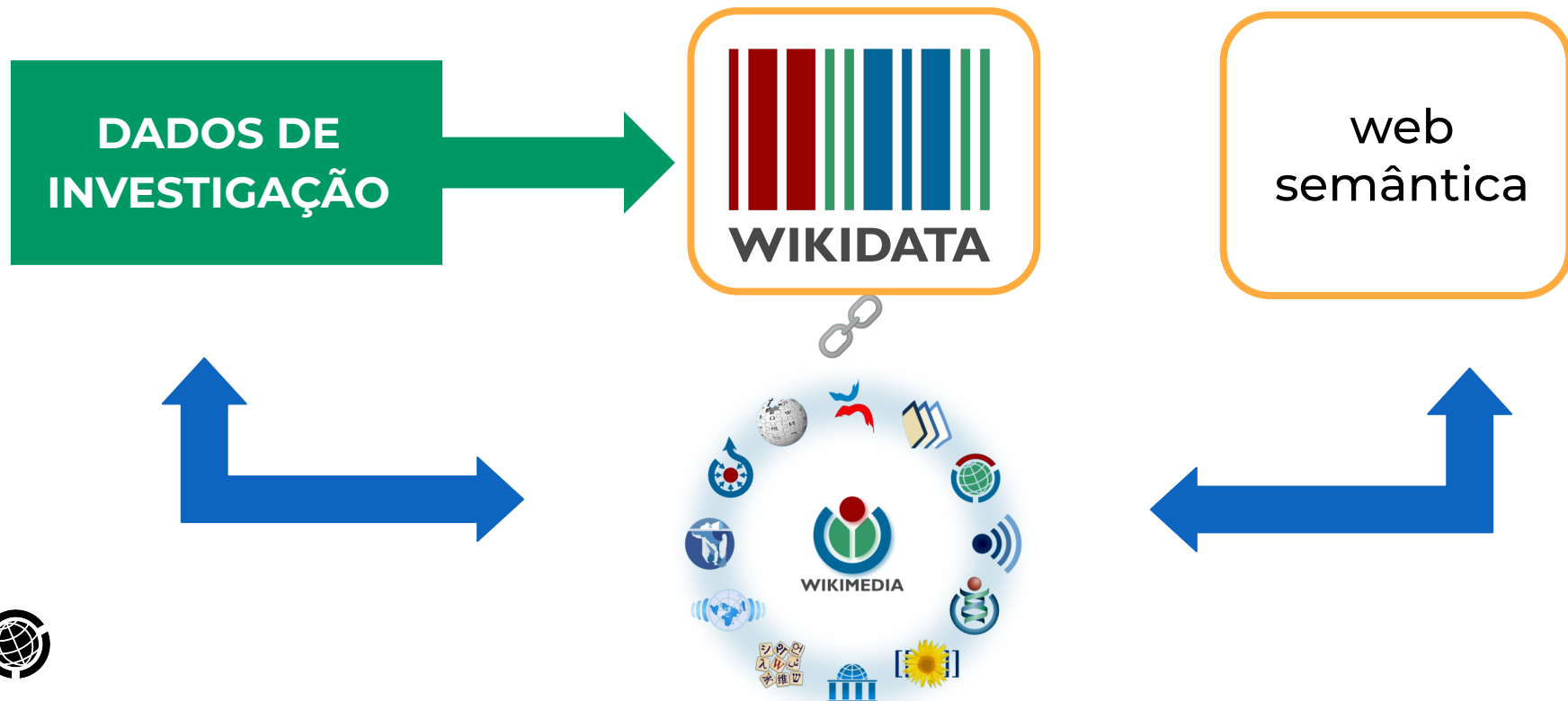
Base de dados interligados e estruturados, livre e aberta, que pode ser lida e editada por humanos e por máquinas

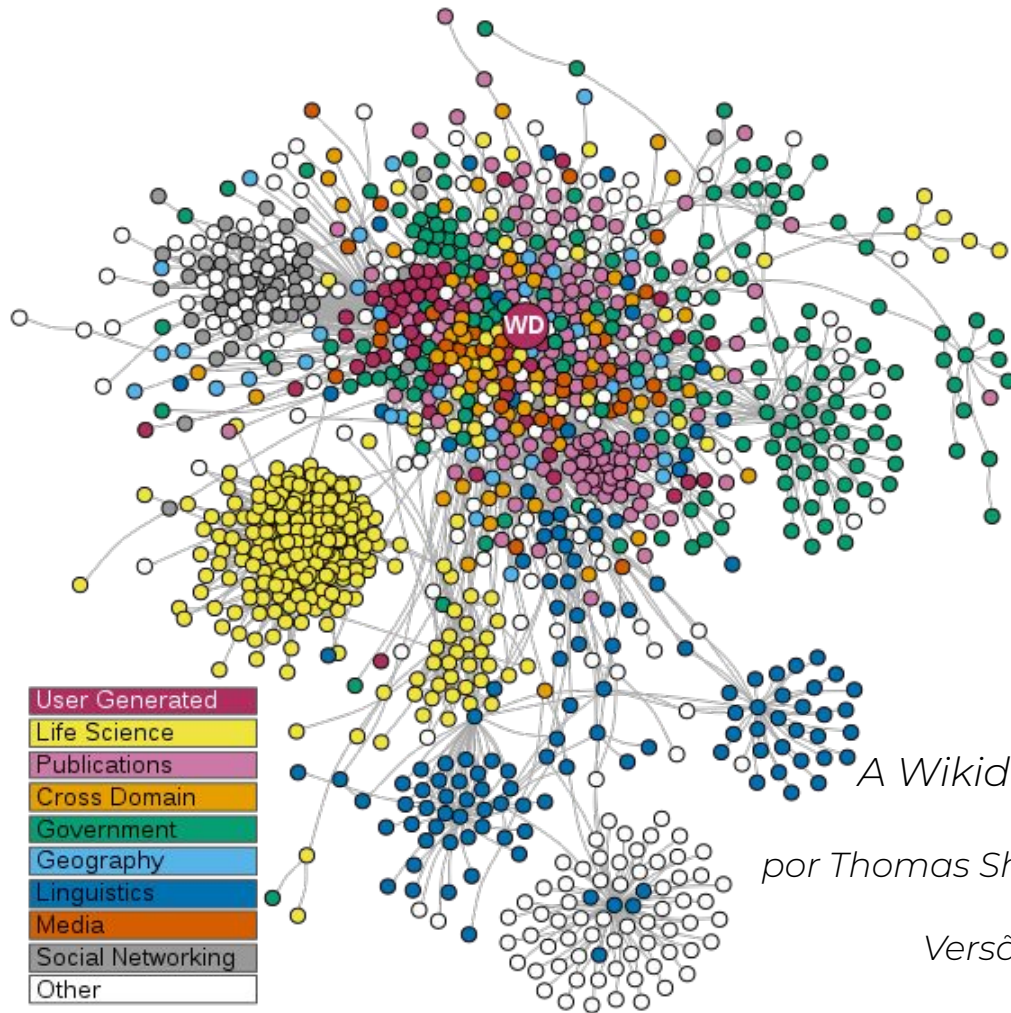


# Redundância no depósito de dados



# Multiplicando o alcance dos dados





A Wikidata na nuvem de dados abertos e interligados (2020),  
por Thomas Shafee [CC BY], [via Wikimedia Commons](#)

Versão de 2022, via [Linked Open Data Cloud](#)





# Dados interligados e abertos (LOD)

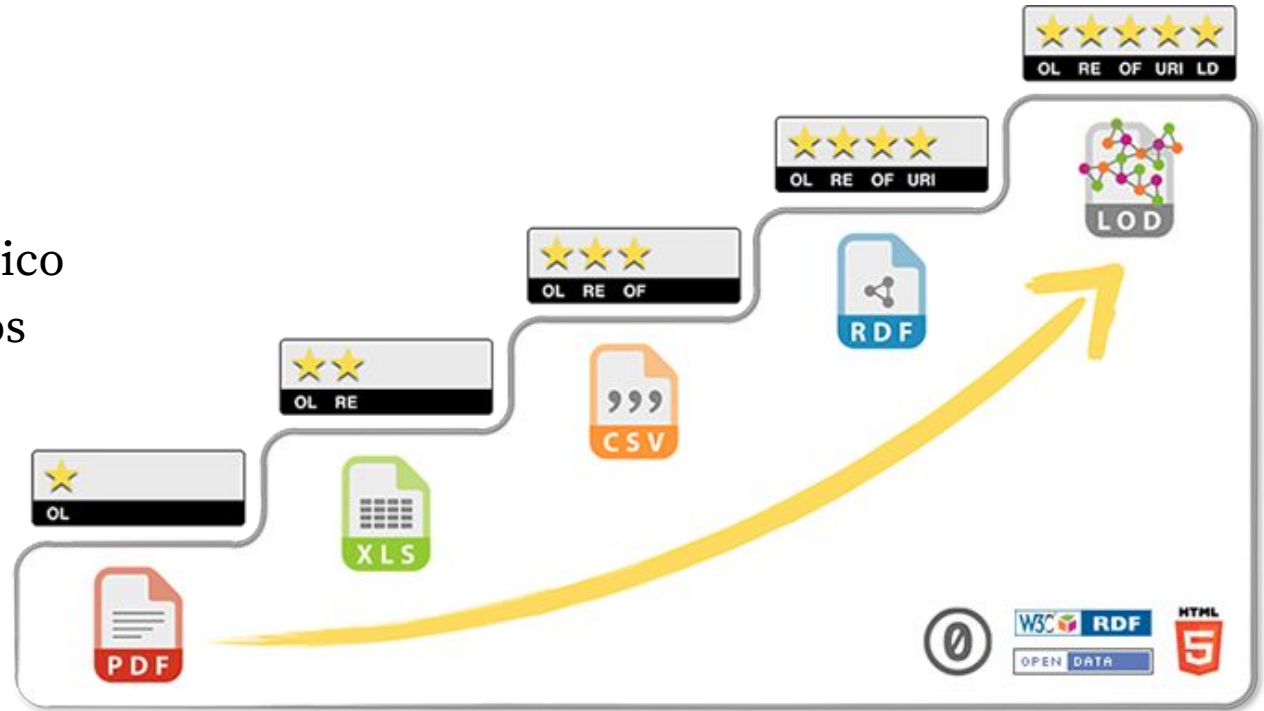
OL - licença aberta

RE - formato editável

OF - formato aberto

URI - identificador único

LD - dados interligados



Michael Hausenblas e James G. Kim, sistema de graduação de abertura de dados desenvolvido por Tim Berners-Lee. [CC0], [via Wikimedia Commons](#)

# Licenças abertas



Mais informação sobre licenças de dados de investigação: [How do I license my research data? \(via OpenAIRE\)](#)

# Dados abertos e interoperáveis

**ESTRUTURA  
DOS DADOS**

**TIPO DE  
DADOS**

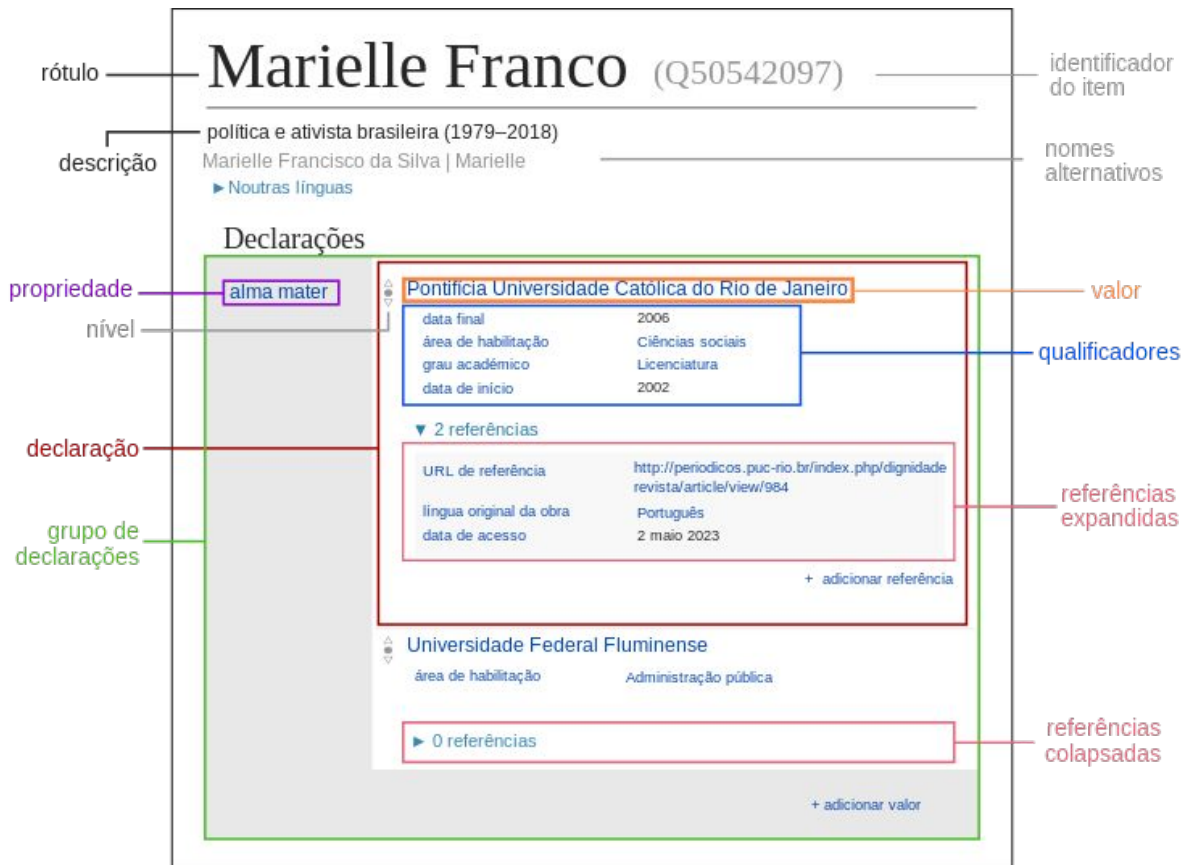
**FORMATOS**

Mais informação sobre a importância dos formatos na preservação de dados de investigação: [Data formats for preservation \(via OpenAIRE\)](#)

# Identificadores únicos (e persistentes)

- De publicações científicas (DOI, ISBN)
- Pessoais do investigador (ORCID, Ciência Vitae, ISNI)
- **De bases de dados (DOI)**
- **De registos dentro de bases de dados online (URI)**

# Estrutura de uma entidade Wikidata



**Declaração = Entidade +  
Propriedade + Valor**

# Faculdade de Ciências Sociais e Humanas (NOVA FCSH) ([Q5429075](#))

- > instância de ([P31](#))
- > parte de ([P361](#))
- > faculdade ([Q180958](#))
- > instituição de educação pública ([Q23002037](#))
- > Universidade Nova de Lisboa ([Q1979891](#))

# Grafos de conhecimento

```
1 #defaultView:Graph
2 SELECT ?item1 ?image ?item1Label ?item2 ?image2 ?item2Label ?edgeLabel
3 VALUES ?item1 {wd:Q1112288 wd:Q94994266 wd:Q102078233 wd:Q2745620}
4 VALUES ?item2 {wd:Q1112288 wd:Q94994266 wd:Q102078233 wd:Q2745620}
5 ?item1 ?prop ?item2.
6 ?edge ?dummy ?prop ; rdf:type wikibase:Property
7 OPTIONAL {?item1 wdt:P18 ?image}
8 OPTIONAL {?item2 wdt:P18 ?image2}
9 SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGU
10 }
```



*Exemplo de grafo de conhecimento extraído do Wikidata através de SPARQL*  
*Ligações entre as entidades Columbano Bordalo Pinheiro(Q1112288), Maria Augusta Bordalo Pinheiro (Q94994266), A luva cinzenta (Q102078233) e o Museu Nacional de Arte Contemporânea do Chiado (Q2745620)*



# Possibilidades de integração de dados no Wikidata

- Carregamento da base de dados ligados e estruturados para o Wikidata (possibilidade de ligação a outras bases de dados; [Mix 'n' Match](#))
- Identificadores de referência de [entidades no Wikidata](#)
- Melhoria dos registos [meta-bibliográficos de cada projeto](#) (incluindo equipas e publicações) no Wikidata
- Incorporação dos sites dos projectos em certas propriedades (ex. P195 para coleções, P953 para URL de texto ou obra integral, P217 para número de inventário)
- Incorporação de dados estruturados na descrição de objetos digitais

# Coleções e ligações para arquivos externos


```
1 SELECT DISTINCT ?revista ?revistaLabel ?VIAF ?ISSN ?catBNP ?BND ?textointegral ?sitelink
2 WHERE
3 {
4   ?revista wdt:P31/wdt:P279* wd:Q41298 ; #revista
5           wdt:P195 wd:Q112612514 . #na coleção do RIC
6   OPTIONAL { ?revista wdt:P214 ?VIAF . } # ID VIAF
7   OPTIONAL { ?revista wdt:P236 ?ISSN . } # ID ISSN
8   OPTIONAL { ?revista wdt:P6373 ?catBNP . } # ID catBNP
9   OPTIONAL { ?revista wdt:P5691 ?BND . } # ID BND
10  OPTIONAL { ?revista wdt:P953 ?textointegral . } # texto disponível em
11  OPTIONAL { ?sitelink schema:about ?revista . ?sitelink schema:inLanguage "pt" } # se tiver artigos wiki (sitelinks) em Português ("pt"), listar
12  SERVICE wikibase:label { bd:serviceParam wikibase:language "pt". }
13 }
14
15 ORDER BY ?revista
```

**Query:** Revistas na coleção do RIC, com ligação para o texto integral

# Bases de dados como identificador externo de referência








Property talk:P11073


## Documentação

<b>ROSSIO ID</b>	<a href="#">Iniciar uma consulta</a> - <a href="#">Utilização atual</a> - <a href="#">Estatísticas por classe</a> - <a href="#">Comprimento da cadeia de caracteres</a> - <a href="#">identificador para uma entidade nos Vocabulários da Infraestrutura ROSSIO (lugares, agentes, períodos ou tesouro)</a>
<b>Item associado</b>	Infraestrutura ROSSIO (Q113720295)
<b>Tipo de dados</b>	Identificador externo
<b>Domínio</b>	ser humano (Q5), grupo de humanos (Q16334295), organização (Q43229), conceito (Q151885), posição geográfica (Q2221906) ou intervalo de tempo (Q186081)
<b>Valores permitidos</b>	(lugares agentes periodos tesouro)\c_([0-9a-fA-F]){8}
<b>Exemplo</b>	<a href="#">Teatro Nacional D. Maria II (Q116125)</a> → <a href="#">agentes/c_1674913a</a> (RDF) <a href="#">↗</a> <a href="#">Repositório da Universidade Nova de Lisboa (Q113733025)</a> → <a href="#">agentes/c_c37b6dc5</a> (RDF) <a href="#">↗</a> <a href="#">ciências sociais (Q34749)</a> → <a href="#">tesouro/c_d7329929</a> (RDF) <a href="#">↗</a> <a href="#">humanidades (Q80083)</a> → <a href="#">tesouro/c_6949eed2</a> (RDF) <a href="#">↗</a> <a href="#">Lisboa (Q597)</a> → <a href="#">lugares/c_19d11643</a> (RDF) <a href="#">↗</a>
<b>Origem</b>	<a href="http://vocabs.rossio.fcsh.unl.pt/pt/">http://vocabs.rossio.fcsh.unl.pt/pt/</a> <a href="#">↗</a>
<b>Formato da URL</b>	<a href="http://vocabs.rossio.fcsh.unl.pt/\$1">http://vocabs.rossio.fcsh.unl.pt/\$1</a>
<b>Relativo a país</b>	 Portugal (Q45) (See 33 others)

[Vocabulários controlados da Infraestrutura ROSSIO](#) são um identificador externo na Wikidata

# Descrição de objetos digitais em coleções

P-Cua IV-3ª Gav. 44 (22) 	
<b>Title</b>	P-Cua IV-3ª Gav. 44 (22)
<b>Object type</b>	manuscript fragment / notated music / Mozarabic chant 
<b>Genre</b>	Liber mysticus 
<b>Description</b>	<b>English:</b> Temporale in December: end of Vespers for Saturday before the 3rd Sunday in Advent, and beginning of this Sunday's early morning hour (corresponding to the Lauds of Roman rite). - Folio recto.
<b>Language</b>	Latin 
<b>Publication date</b>	circa 1000 
<b>Medium</b>	parchment 
<b>Collection</b>	<a href="#">Portuguese Early Music Database</a> 
<b>Source/Photographer</b>	<a href="#">Portuguese Early Music Database</a> 



A [Portuguese Early Music Database](#) usa o Wikidata para a descrição de manuscritos e os respetivos objetos digitais.

# Wikidata:WikiProject sum of all paintings/Collection/Biblioteca Nacional de Portugal

< Wikidata:WikiProject sum of all paintings | Collection

instância de (P31) -> pintura (Q3305213) & localização (P276) -> Biblioteca Nacional de Portugal (Q245966)

This list is periodically updated by a bot. Manual changes to the list will be removed on the next update!

[Manually update list](#)

WDQS | PetScan | TABernacle | Find images | Mudanças recentes | Query: `SELECT ?item WHERE { ?item wdt:P276 wd:Q245966 . ?item wdt:P31 wd:Q3305213 . } LIMIT 2000`

image	label	description	creator	inception	inventory number	genre	main subject	depicts	described at URL	BND ID
	Father Manuel Bernardes	painted portait of a Portuguese priest	Portuguese painter	18th century	Inv. 14492	portrait	Manuel Bernardes	Manuel Bernardes	<a href="http://catalogo.bnportugal.pt/ipac20/ipac.jsp?profile=bn&amp;uri=full=3100024~!1743787-10">http://catalogo.bnportugal.pt/ipac20/ipac.jsp?profile=bn&amp;uri=full=3100024~!1743787-10</a>	22954
	Afonso Bourbon e Meneses	1938 painting of Afonso Bourbon e Meneses	Helena de Meneses	1938	Inv. 17440	portrait	Afonso Augusto Bourbon e Menezes	Afonso Augusto Bourbon e Menezes	<a href="http://catalogo.bnportugal.pt/ipac20/ipac.jsp?profile=bn&amp;uri=full=3100024~!1750064-10">http://catalogo.bnportugal.pt/ipac20/ipac.jsp?profile=bn&amp;uri=full=3100024~!1750064-10</a>	23859
	D. Afonso Henriques	painting of D. Afonso Henriques	No/unknown value	16th century	Inv. 10932	portrait	Afonso I of Portugal	Afonso I of Portugal	<a href="http://catalogo.bnportugal.pt/ipac20/ipac.jsp?profile=bn&amp;uri=full=3100024~!1744408-10">http://catalogo.bnportugal.pt/ipac20/ipac.jsp?profile=bn&amp;uri=full=3100024~!1744408-10</a>	22974
	Padre José Vaz	painted portrait at bust level of Portuguese priest from Congregação do Oratório de São Filipe de Neri	Portuguese painter	18th century	Inv. 14496	portrait		Joseph Vaz	<a href="http://catalogo.bnportugal.pt/ipac20/ipac.jsp?profile=bn&amp;uri=full=3100024~!1743752-10">http://catalogo.bnportugal.pt/ipac20/ipac.jsp?profile=bn&amp;uri=full=3100024~!1743752-10</a>	22952
	Irmão Pedro de Alpoim	painted portrait at bust level of Portuguese priest from Congregação do Oratório de São Filipe de Neri	Portuguese painter	18th century	Inv. 14497	portrait		priest	<a href="http://catalogo.bnportugal.pt/ipac20/ipac.jsp?profile=bn&amp;uri=full=3100024~!1743781-10">http://catalogo.bnportugal.pt/ipac20/ipac.jsp?profile=bn&amp;uri=full=3100024~!1743781-10</a>	22953

# Conciliação de identificadores

```
1 SELECT DISTINCT ?researchunit ?researchunitLabel ?VIAF ?LibCon ?BNP
2 WHERE
3 {
4   ?researchunit wdt:P31/wdt:P279* wd:Q31855 ; #research unit
5                 wdt:P361 wd:Q5429075 . #parte da FCSH
6   OPTIONAL { ?researchunit wdt:P214 ?VIAF . } # ID VIAF
7   OPTIONAL { ?researchunit wdt:P244 ?LibCon . } # ID LibCon
8   OPTIONAL { ?researchunit wdt:P1005 ?BNP . } # ID BNP
9   SERVICE wikibase:label { bd:serviceParam wikibase:language "pt". }
10 }
11
12 ORDER BY ?researchunit
```

**Query:** Unidades de investigação da NOVA FCSH e respetivos identificadores VIAF, BNP e Biblioteca do Congresso

## Vantagens

- Sem limites de armazenamento
- Solução integrada com Princípios FAIR
- Integração dos dados na sociedade civil
- Interligação com fontes e bases de dados externas
- Conciliação de identificadores, vocabulários e ontologias distintos
- Variedade de formatos e pesquisas
- Carregamento em lote possível

## Limitações

- Propriedades podem estar desajustadas das necessidades de investigação
- Atribuição de autoria apenas mediante uso de certas propriedades
- Carregamento e acesso de dados de grande escala sujeito a disponibilidade dos servidores

# Fluxo de trabalho para carregamento de conjuntos de dados no Wikidata

PREPARE and tidy source data and images	RECONCILE with Wikimedia modeling and coverage	INGEST data and media		
<ul style="list-style-type: none"> <li>• Wikidata Data Donation</li> <li>• Creative Commons</li> <li>• Flickr <a href="#">image hosting</a></li> <li>• Google Spreadsheets <a href="#">, Wikipedia and Wikidata tools for Google Spreadsheets</a></li> </ul>	<ul style="list-style-type: none"> <li>• OpenRefine <a href="#"></a></li> <li>• Wikidata Mix'n'match</li> <li>• Wikidata Query <a href="#"></a></li> <li>• Wikidata Graph Builder <a href="#"></a></li> <li>• SQID</li> <li>• Petscan <a href="#"></a></li> <li>• PAWS <a href="#"></a></li> <li>• Mediawiki API</li> </ul>	<ul style="list-style-type: none"> <li>• QuickStatements</li> <li>• OpenRefine <a href="#"></a></li> <li>• Wikidata-CLI <a href="#"></a></li> <li>• Pattypan</li> <li>• url2commons</li> <li>• PyWikibot / upload.py</li> </ul>		
		ANALYZE, correct, and enrich	RE-USE content intra-wiki and externally	REPORT and measure
		<ul style="list-style-type: none"> <li>• Wikidata Query <a href="#"></a></li> <li>• TABernacle</li> <li>• Listeria</li> <li>• Maintenance query scripts</li> <li>• Wikidata Distributed Game</li> <li>• Cat-a-lot</li> <li>• VisualFileChange.js</li> <li>• SPARQL RC</li> </ul>	<ul style="list-style-type: none"> <li>• Wikidata infobox / Infobox templates</li> <li>• Creator templates</li> <li>• Template:Artwork, Template:Book, etc</li> <li>• Template:Label, Template:WrapWD, etc</li> <li>• Listeria / Template:Wikidata_list</li> </ul>	<ul style="list-style-type: none"> <li>• GLAMorgan</li> <li>• BaGLAMa</li> <li>• GLAM Wiki Dashboard <a href="#"></a></li> </ul>

Detalhes em: [Wikidata:Linked open data workflow](#)



# Wikidata: acervo e fonte de informação

## Artigos em falta sobre mulheres de Portugal

Esta é uma lista de itens no Wikidata sobre mulheres portuguesas que ainda não possuem verbetes na Wikipédia lusófona. Não é preciso editar a tabela após a criação do verbete, ela é atualizada periodicamente pelo bot [?](#).



Esta lista é gerada a partir dos dados do Wikidata e é atualizada periodicamente por um robô??.  
Edições feitas no conteúdo da lista serão removidas na próxima atualização!

Atualizar a lista automaticamente agora? | SPARQL? | Pesquisar imagens?

name	imagem	description	ocupação	data de nascimento	data de morte	local de nascimento	local de morte	wikidata item	site links
<a href="#">Hanna Debora Marques</a>				No/unknown value	1638			<a href="#">Q54251324</a>	0
<a href="#">Etelvina Silva</a>		esperantista portuguesa	<a href="#">esperantista</a>	1855				<a href="#">Q19723779</a>	1
<a href="#">Maria Teresa de Bragança</a>		Person (1881–1945) ♀; child of <a href="#">Miguel Januário de Bragança</a> , <a href="#">Isabel de Thurn e Taxis</a> ; spouse of <a href="#">Karl Ludwig Prinz von Thurn und Taxis</a>		1881-01-26	1945-01-17	<a href="#">Sopron</a>	<a href="#">Stainach</a>	<a href="#">Q4281713</a>	6
<a href="#">Amélia Bacelar</a>			<a href="#">aracnólogo</a> <a href="#">entomologista</a> <a href="#">zoólogo</a>	1890 1893	1976			<a href="#">Q56854926</a>	1
<a href="#">Maria Henriques Oswald</a>		escritora e tradutora portuguesa	<a href="#">tradutor</a> <a href="#">escritor</a>	1893	1988	<a href="#">Porto</a>		<a href="#">Q15431940</a>	1
<a href="#">Margarida Trincão</a>		jornalista portuguesa	<a href="#">jornalista</a>	19th century		<a href="#">Torres Novas</a>		<a href="#">Q25244346</a>	0
<a href="#">Ester Conceição Pereira de Sousa</a>			<a href="#">botânico</a>	1907-11-27	1994-01-16			<a href="#">Q21391064</a>	1
<a href="#">Maria Adelaide Rosado Pinto</a>		professora, musicóloga e poetisa setúbalense	<a href="#">professor</a> <a href="#">musicólogo</a> <a href="#">escritor</a> <a href="#">poeta</a>	1913-06-13	1997-09-22	<a href="#">Setúbal</a>	<a href="#">Setúbal</a>	<a href="#">Q85308041</a>	0

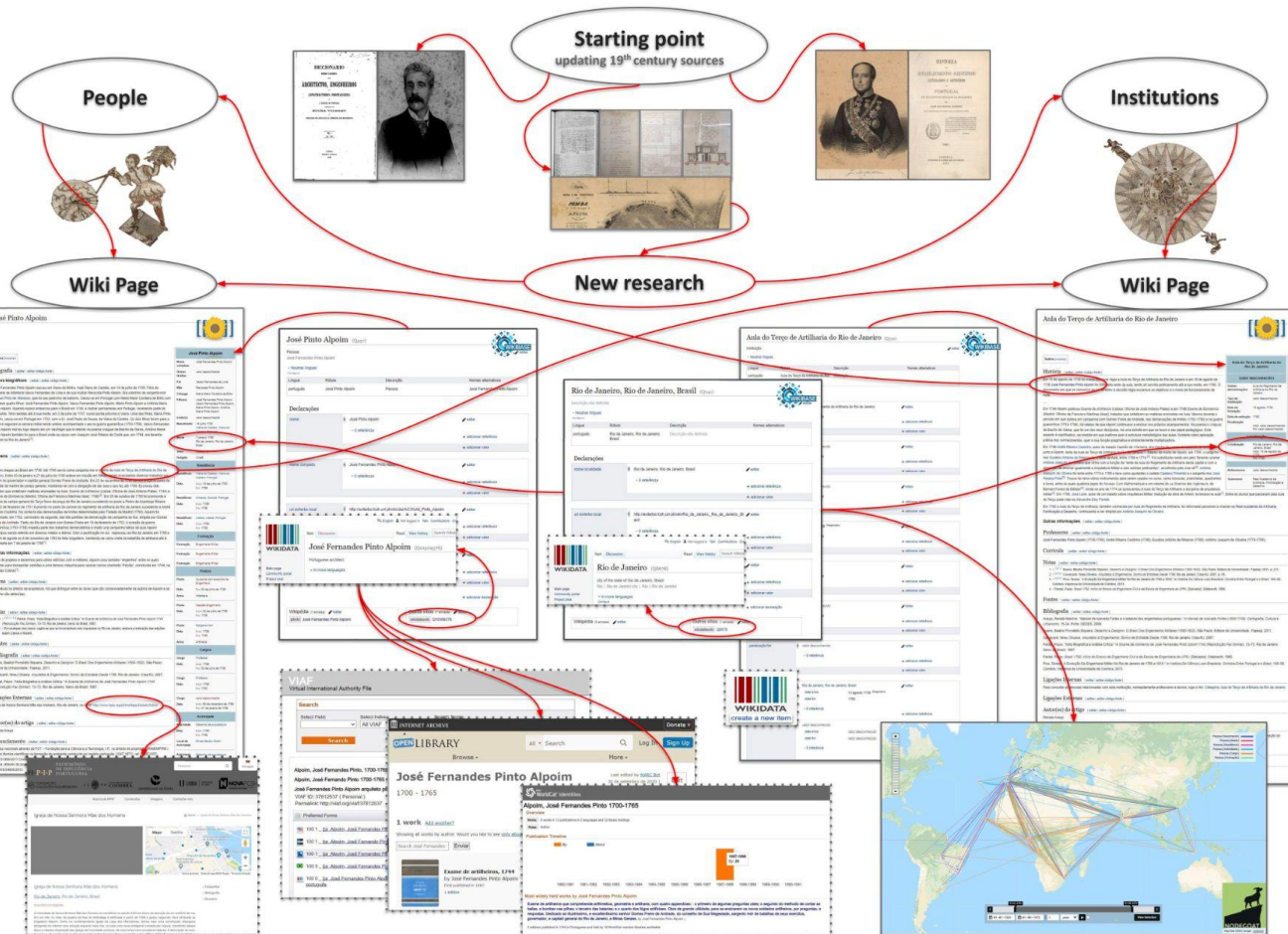
# eViterbo

## Enhancing Linked Research

eViterbo, is a Wiki format platform developed by the on-going research project TechNetEMPIRE.

eViterbo builds upon biographical data of building experts (people) and institutions across the former Portuguese empire and uses mediawiki software, wikibase, with future data insertion in wikidata.

This poster presents its general aim and structure, underlying linked open data possibilities.



# Em suma:

- Utilização do Wikidata como complemento à partilha em repositórios científicos/institucionais
- A importância dos planos de gestão de dados: ajudam a definir o escopo dos dados
- Licenças abertas, formatos abertos, identificadores persistentes
- Definição de prioridades de partilha junto da sociedade civil

# Questões?

# Obrigada

Rute Correia  
Wikimedista em Residência,  
WMPT / NOVA FCSH

[rute.correia@wikimedia.pt](mailto:rute.correia@wikimedia.pt)

Página da [residência na Wikipédia](#)

