

Wikidata Lab 39

[[User:Mike Peel]]

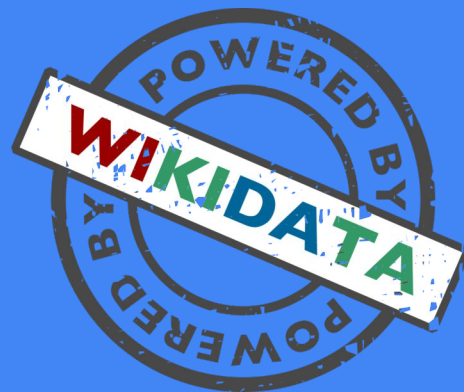
Online & NeuroMat,
São Paulo, Brazil

21 November 2023

Supported by



Structuring the Wikimedia Ecosystem



Overview

- 1) Interwiki links through Wikidata
- 2) Matching new Wikipedia articles with Wikidata
- 3) Importing basic information about new articles
- 4) Matching Commons categories with Wikidata
- 5) Expanding to cover other sibling projects
- 6) Conclusions
- 7) Afternoon tasks

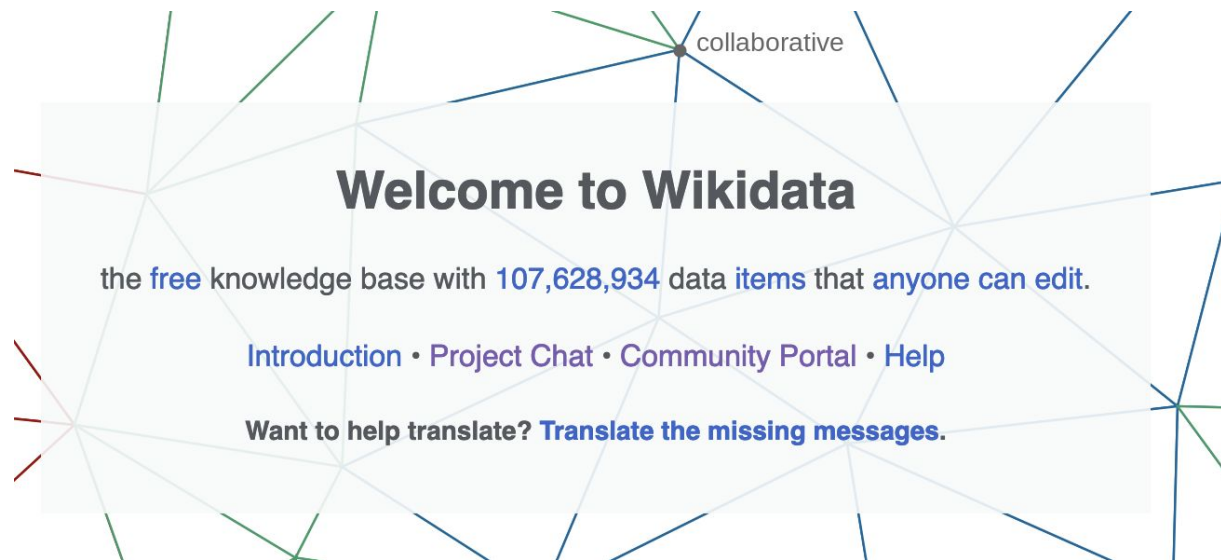
The Wikimedia Ecosystem

- Wikipedia is the most well-known Wikimedia project. It has siblings.
- Knowledge exists in many ways, so we need many projects. Not everything should become an encyclopedic entry.
- Wikimedia projects are sustained by a non-proprietary content licence, a collaborative technical infrastructure and a community.



Wikidata is an open collaborative knowledge base

It is a fast-growing Wikimedia project, one of the newest, started in 2012 - over 2 billion edits!



(# of items will already be out-of-date!)

Wikidata Lab (Q61791944) ... QIDs

Item [Discussion](#)

series of workshops for capacity building on Wikidata editing in Brazil

▼ In more languages

Language	Label	Description	Also known as
English	Wikidata Lab	series of workshops for capacity building on Wikidata editing in Brazil	
Spanish	Wikidata Lab	serie de formación para la capacitación en edición de Wikidata en Brasil	
French	Wikidata Lab	No description defined	
Portuguese	Wikidata Lab	série de treinamentos para capacitação em edição da Wikidata no Brasil	

All entered languages

Statements

instance of	<div>recurring event ...</div> <div>workshop ...</div>	<div>0 references</div> <div>0 references</div>
subclass of	<div>occurrence ...</div>	<div>0 references</div>

logo image




Logo Wikidata Lab.svg

512 × 128; 238 KB

0 references

image



0 references

Labels

Linked values

Media

Wikipedia (1 entry) [move]

pt [Categoria:!Wikidata Labs](#) [ref]

Wikibooks (0 entries)

Wikinews (0 entries)

Wikiquote (0 entries)

Wikisource (0 entries)

Wikiversity (0 entries)

Wikivoyage (0 entries)

Wiktionary (0 entries)

Multilingual sites (2 entries) [move]

commons [Category:Wikidata Labs](#)

wikidata [Wikidata:Wikidata Labs](#)

Interwikis

Interwiki links through Wikidata - simpler

- From the beginning, Wikidata has greatly improved interwiki linking
- Used to have to link to every language from every language!
- NxN problem, had to be bot-managed, many edits (esp. for small wikis)
- Now, link languages to Wikidata
- → N+1 problem
- Also link sibling projects
- Significantly simpler and scalable!

(example: Q1513315)

Wikipedia (20 entries) [move]

ar	مرصد القطب الجنوبي	[ref]
cy	Telesgop Pegwn y De	[ref]
de	South Pole Telescope	[ref]
en	South Pole Telescope	[ref]
eo	Suda Polusa Teleskopo	[ref]
es	Telescopio Polo Sur	[ref]
fa	تلسکوپ قطب جنوب	[ref]
fr	South Pole Telescope	[ref]
id	Teleskop South Pole	[ref]
it	South Pole Telescope	[ref]
ja	南極点望遠鏡	[ref]
lb	South Pole Telescope	[ref]
lt	Pietų ašigalio teleskopas	[ref]
no	Sørpolteleskopet	[ref]
pl	South Pole Telescope	[ref]
pt	Telescópio do Polo Sul	[ref]
ru	Южный полярный телескоп	[ref]
si	උතුරු ටෙලිස්කොප්	[ref]
uk	Південний полярний телескоп	[ref]
zh	南極望遠鏡	[ref]

Wikibooks (0 entries)

Wikinews (0 entries)

Wikiquote (0 entries)

Wikisource (0 entries)

Wikiversity (0 entries)

Wikivoyage (0 entries)

Wiktionary (0 entries)

Multilingual sites (1 entry) [move]

commons Category:South Pole Telescope

Interwiki links through Wikidata - but complications

- Still some complications...
- Only one sitelink per language project per item
- Commons galleries are in main item, category gets separate item, then 'Topic's main category' (P910) & 'Category's main topic' (P301) properties to link between them (unless there is no gallery, then the category link goes in main item)
- Also links between lists and category items, 'list related to category' (P1753) and 'category related to list' (P1754)
- What happens when Wikipedia articles combine topics?
- Sitelinks now allowed to redirects, which helps!

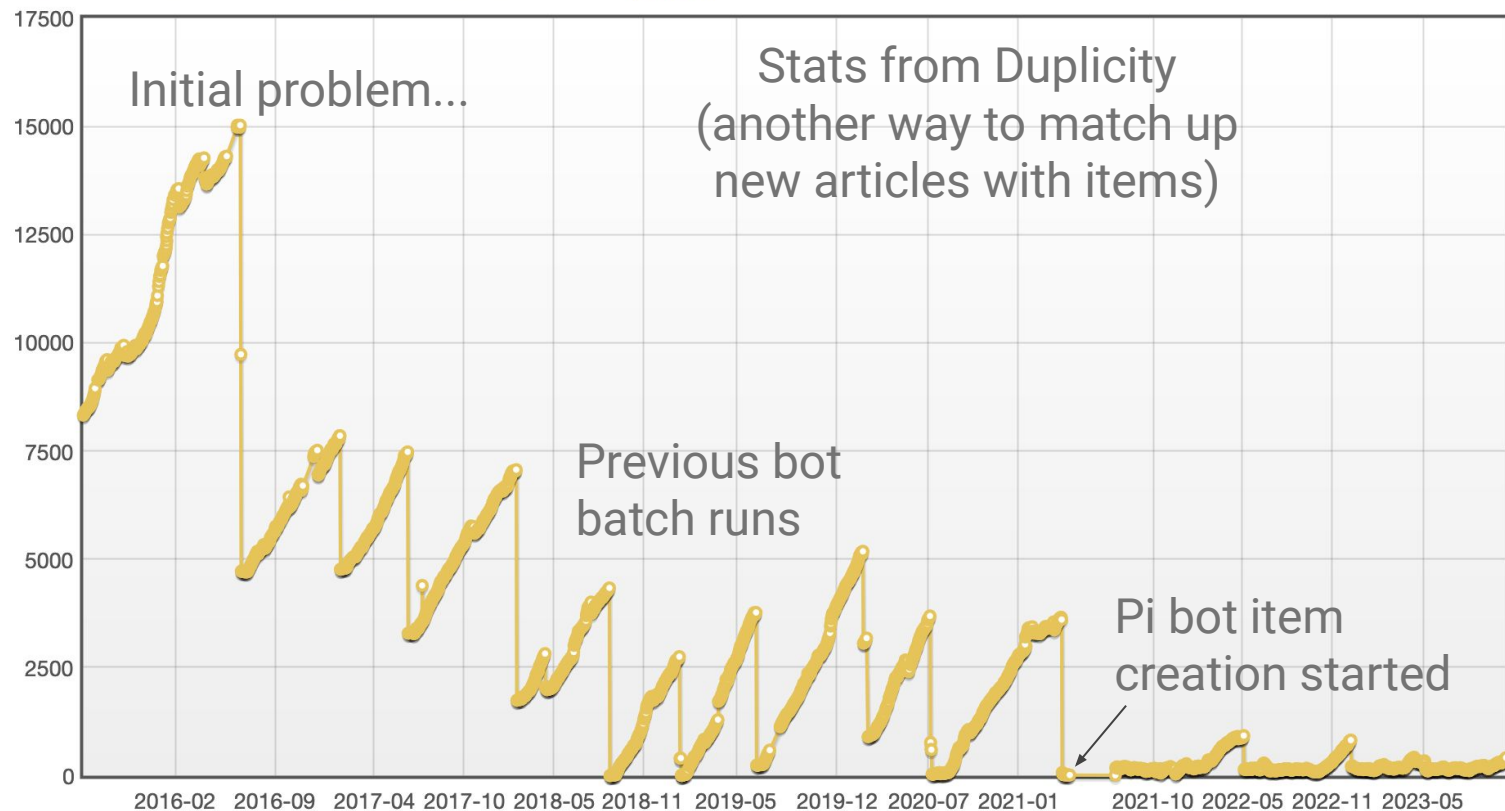
Matching new Wikipedia articles with Wikidata

- When a new article is created, it needs linking with Wikidata
- Ideally, article creators match up their new article to Wikidata! (automatic when using the article translation tool, manual otherwise)
- Pi bot runs an automated process for matches/new items
 - Only looks at mainspace/categories/places that should be connected to Wikidata (not everything is in scope)
 - Searches for matches, loads results into a game - only creates a new item once all these have been manually checked
 - Waits 14 days before creating new item, also avoiding recently edited items, to give time for manual matches...
- Currently live on a few projects - potential roll-out to the rest? Community consensus needed, or proven to work now?

Matching new Wikipedia articles with Wikidata

Number of articles without Wikidata item on [ptwiki](#)

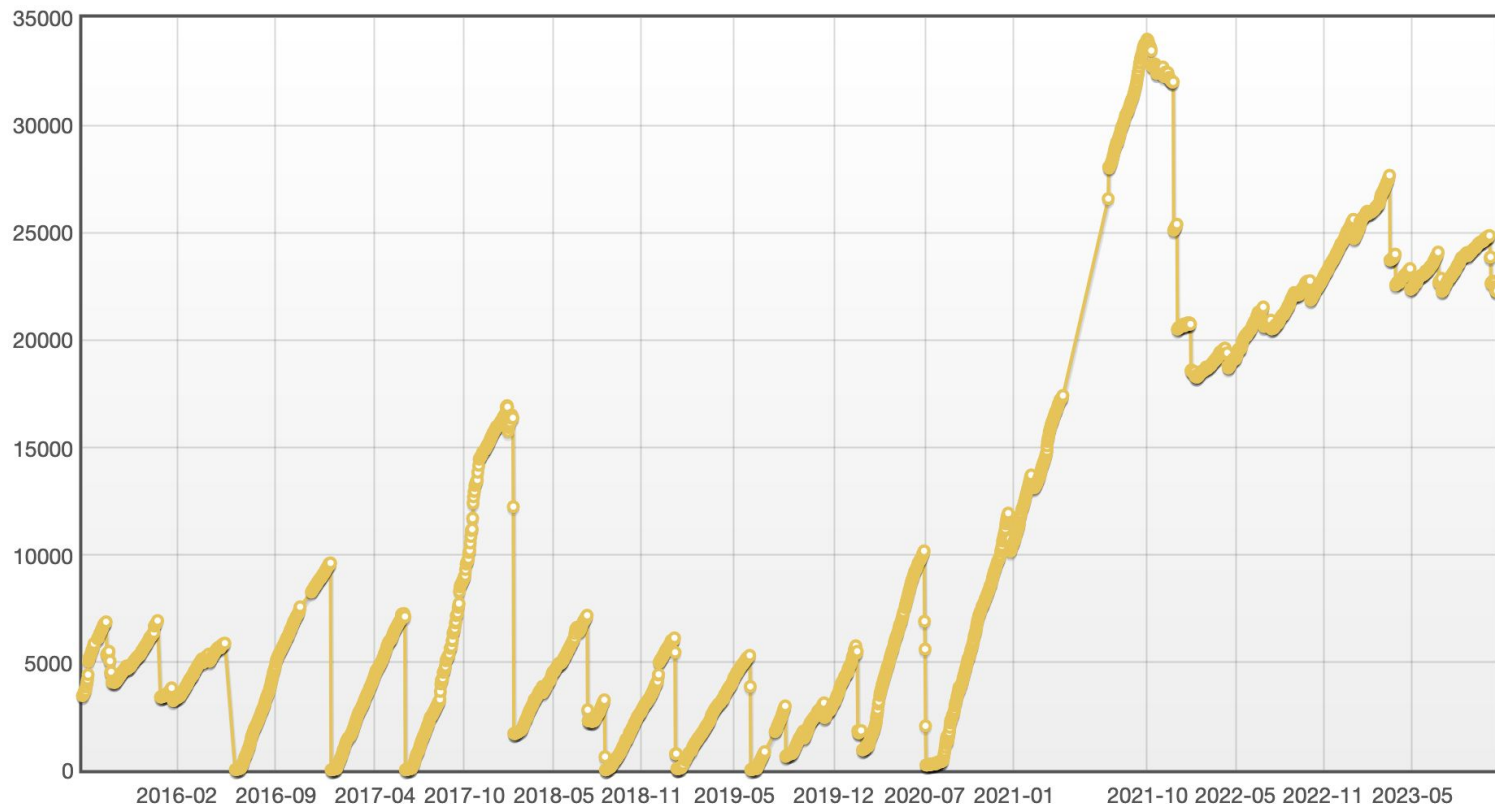
<https://wikidata-todo.toolforge.org/duplicity/>



Matching new Wikipedia articles with Wikidata

Number of articles without Wikidata item on [zhwiki](#)

<https://wikidata-todo.toolforge.org/duplicity/>



Matching new Wikipedia articles with Wikidata

- <https://wikidata-game.toolforge.org/distributed/#game=82>

New Wikipedia article and category matches

Match new Wikipedia articles and categories with Wikidata items, and add the sitelink to Wikidata.

Pi bot is thinking about creating new items for these articles, but first it wants your help to match them to existing items.

If the match is right, please add the link to Wikidata using "Match". If it is clearly wrong, select "No". If you are not sure, press "Skip".

Bug reports and feedback should be sent to User:Mike_Peel.

Entry type :

Any

English

Simple

Português

Deutsch

Español

Français

Italiano

Nederlands

Polski

Svenska

Esperanto

Jose Ronaldo Medeiros [Q64849742]

Auto

Civil servant (*1965) ♂

Ronaldo Medeiros [ptwiki]

José Ronaldo Medeiros (Rio de Janeiro, 17 de fevereiro de 1965), é um servidor público e político brasileiro, filiado ao PT. Exerceu o mandato de deputado estadual entre 2011 à 2019, quando foi eleito suplente em 2018, voltando para o cargo em 2021 ao assumir a vaga de Marcelo Beltrão que se elegera prefeito de Coruripe.

Nas eleições de 2022, foi reeleito deputado estadual por AL.

[Google Translate](#)

Referências</

Match

Skip

No

Matching new Wikipedia articles with Wikidata

- <https://wikidata-game.toolforge.org/distributed/#game=82>

K-9 and Company [Q2604192]
К-9 и компания El K9 e Votikans K-9 i компанія К-9 и компанија 機械狗K-9與同伴 K-9與公司

Auto | de | en | es | fr | hu | pl | ru | simple | uk | vo

1981 television pilot and Doctor Who spin-offs by Fiachra Trench, John Black, and Terence Dudley



K-9 and Company [ptwiki]

K-9 and Company é um episódio piloto para televisão, feito com a proposta de um *spin-off* da série britânica de ficção científica Doctor Who. A série apresenta a personagem Sarah Jane Smith (Elisabeth Sladen), uma jornalista investigativa apresentada na 11ª temporada clássica de Doctor Who, e seu cachorro robótico K-9 (dublado por John Leeson). O único episódio, "*A Girl's Best Friend*", foi transmitido pela BBC One como um especial de natal em 28 de dezembro de 1981, porém não foi desenvolvido em uma série devido a divergências entre os executivos da BBC. Foi o primeiro *spin-off* televisionado de Doctor Who.

O episódio se passa no vilarejo fictício de Moreton Harwood em 1981, onde Sarah e K-9 investigam o desaparecimento de Brendan Richards (Ian Sears), o cuidador da tia de Sarah, Lavinia (Mary Wimbush), ao mesmo tempo em que aparecem rumores sobre a prática de bruxaria na região.

Sob uma nova equipe de produção, o conceito geral da série foi

Match Skip No

Importing basic information about new articles

- It's more helpful to have basic information in a Wikidata item than empty
- Based on article title, category usage, etc.
- Biographies are handled differently - immediate matches and new items created, based on name and birth dates (unless there are ambiguities)
 - This includes gender identification (open question: is this a good idea?)

```
new_item.addClaim(claim, summary='List item')
elif 'lista de' in page.title()[0:10].lower():
    # input('Is list - OK?')
    claim = pywikibot.Claim(repo, 'P31')
    claim.setTarget(pywikibot.ItemPage(repo, 'Q13406463')) # List item
    new_item.addClaim(claim, summary='List item')
elif pageIsBiography(page, lang_prefix):
```

Importing basic information about new articles

- Pi bot also runs tasks to bulk-import content from Wikipedias:
 - Short descriptions from enwiki
 - Coordinates from enwiki, commons
 - Birth/death dates (although this can import vandalism...)
- Could do more?
 - Authority control
 - Infobox parameters? (but tricky to match with Wikidata items)
 - ideas...?

Matching Commons categories with Wikidata

- Commons links before Wikidata were mostly Wikipedia templates
- Wikidata changed this dramatically - can now sitelink to Commons!
- Big effort to do this over the last 5 years, driven by Wikidata Infobox
- Wikidata now has nearly 5 million sitelinks to Commons!
- 'Commons category' P373 also exists, but is not so useful
 - Was set up as a temporary measure until sitelinks existed...
 - Can have multiple values, link to same category from multiple items
 - But is one-directional, can find the Commons category from Wikidata, but not Wikidata from the Commons category
 - (deletion attempts haven't succeeded yet - open discussion point?)

Wikidata Infobox on Wikimedia Commons

NeuroMat [\[Collapse\]](#)

research institute



☒ image ☐ logo image
☐ Show all

[Upload media](#)

 [Wikimedia](#)

Instance of	research institute
Location	NeuroMat building , São Paulo , Brazil
Headquarters location	São Paulo
Inception	2013

[official website](#)



300 m [Wikimedia maps](#) | [Map data](#) | ...
23° 33′ 31.14″ S, 46° 43′ 55.45″ W

Authority file [\[Collapse\]](#)

 [Q18477654](#)

[Reasonator](#) | [Scholia](#) | [Wikidocumentaries](#)
• [PetScan](#) | [statistics](#) | [WikiMap](#)
[Locator tool](#) | [KML file](#) | [WikiShootMe](#)
[OpenStreetMap](#) | [Search depicted](#)

- Started in 2018, bot-deployed in 4.7 million Commons categories!
- Small infobox to add context for category contents, without hiding category contents
- Needs sitelink to connect to Wikidata
- Using Wikidata information only (no local definitions) - completely multilingual code: `{{Wikidata Infobox}}`
- Images on Wikidata shown alongside Commons category - are best images selected? (+ night image etc...?)
- (Ideally used across all projects, same for important data like birthdays etc.)
- Authority control at the bottom (none for NeuroMat - please add some!)
- Continuous development with community input and suggestions

Matching Commons categories with Wikidata

- A lot of links are bot-added, e.g., when someone adds {{Commons category}} to enwiki, or a P373 value to Wikidata.
- Various users batch-add links, e.g., via Wiki Loves Monuments process
- Also bulk-creating, e.g., categories for ships (IMO #) and ship names
- Many have been manually added when categories are created etc.
- A Wikidata Game exists to review possible matches
 - Fed by Python scripts that, e.g., search for category name matches
 - So far has added 46k links, 67k possible matches currently in game
 - (can easily add more when the backlog is lower!)

Matching Commons categories with Wikidata

<https://wikidata-game.toolforge.org/distributed/#game=44>

Commons category matches

Match Commons categories with Wikidata items, and add the commons sitelink to Wikidata.

These matches look plausible. But are they really? Please help us to reject the bad ones by clicking "No" - and if you are sure that it is right, add the link to Wikidata using "Match". If you are not sure, press "Skip".

Bug reports and feedback should be sent to commons:User:Mike Peel.

Entry type : Any

Galeria Kaufhof Pokal Challenge [Q5518439]

Auto | en | nl

Golf tournament

Category:Galeria Kaufhof (Cottbus) [commonswiki]

[Google](#)
[Translate](#)

Categories that this category is in:

Category:Shopping malls in Cottbus - Category:Kaufhof | Cottbus - Category:Karl-Liebknecht-Straße (Cottbus) - Category:Stadtpromenade (Cottbus) - Category:Konsument-Warenhaus - Category:Harry Müller - Category:Buildings in Cottbus-Mitte

Match

Skip

No

Matching Commons categories with Wikidata

- Could bulk-create new items, except:
 - Might cause duplicates (could minimise using the Game)
 - Notability issues, particularly for humans
 - 'Category combines' categories aren't liked by Wikidata people
- Open question: galleries vs. categories - do the former still make sense?
 - There are a lot of unmaintained categories - delete/redirect?
- We now have Structured Data on Commons - will this replace categories?
- There is also the P373 data duplication issue, with a backlog of mismatches (also in Wikipedia articles via templates)

Matching Commons categories with Wikidata

Category:IMO 8511835

Category [Discussion](#)

[Read](#) [Edit](#) [View history](#) [☆](#)

[Categories](#) [\(+\)](#): [Ships by IMO number](#) [\(-\)](#) [\(±\)](#) [\(↓\)](#) [\(↑\)](#) | [Ships built in 1989](#) [\(-\)](#) [\(±\)](#) [\(↓\)](#) [\(↑\)](#) | [\(+\)](#)
[Non-topical/index: Uses of Wikidata Infobox](#) | [Uses of Wikidata Infobox missing Commons Category statement](#)

From Wikimedia Commons, the free media repository

Ship [\[edit \]](#)

- Type: Training ship
- Built by Gdanska Lenina, Gdansk, Poland
- Yard No: b810/03
- Launch Date: 10.6.88
- Date of completion: 3.89
- Length overall: 108.6
- LPP: 79.4
- Beam: 14.0
- Tons: 2284
- DWT: 840
- Sails: 2.770 m2
- Main engine: 2x Sulzer 1.000 Pk
- Speed under power: 9 kn

History [\[edit \]](#)

- 1989 Named: **Khersones** for the Kerch State Maritime Technological University (KSMTU) Flag: Ukraine
- Look up ["IMO 8511835"](#) [↗](#)

Subcategories

This category has only the following subcategory.

K

► [Khersones \(ship, 1989\)](#) (37 F)

Khersones [\[Collapse\]](#)

full-rigged ship



[Upload media](#) [↗](#)

[Wikipedia](#)

Instance of	full-rigged ship
Call sign	ENFI
Has use	training vessel
Manufacturer	Gdańsk Shipyard
Yard number	b810/03
Shipping port	Kerch
Country of registry	Russia
Service entry	1989
Gross tonnage	2,284
Beam	14 m
Draft	6.6 m
Length	108.6 m

Expanding to cover other sibling projects

- Ultimate aim: all Wikimedia projects should have all sitelinks on Wikidata, how to get there?
- Currently looking at en.wikiquote as an example - can find matches and create new items OK - but what to do with possible matches?
 - It seems there is a bug in the Wikidata game - it works OK for Wikipedias and Commons, but not other projects - can anyone help?
- Wikisource - more complicated, since Wikidata has separate items for book editions, which may not match up with Wikipedia articles
 - Some language Wikisources are well integrated, others not yet
- Wikispecies - still needed? Multiple taxon trees on different projects

Conclusions

- Wikipedias: mostly linked, but new articles are tricky
 - Can expand to cover more languages
 - Can improve with importing statements into new items
- Commons: about half way, still a lot to do here
 - P373 still a tricky issue
- Other sibling projects need a lot of work
 - Also related, BOT currently has 'Siblings Projects Task Force' looking at how to handle new projects, and close/merge/split existing projects if appropriate - more news from that soon.

Afternoon tasks

- Play the Wikidata games!
- Identify patterns in new articles that Pi bot can use
- Other ways of selecting information from Wikipedia articles to import?
- Additional languages that the bot can work on?
- Ideas for new games / approaches to this problem
- Write some code (game/bot script/etc.?)
 - Pi bot code all available at <https://github.com/mpeel/wikicode/>
- Create new Wikidata items for Commons categories
- Event main page: [https://w.wiki/85J\\$](https://w.wiki/85J$)
- Make sure to register for edit tracking: <https://t.ly/rFhnt>