

#WikiForHumanRights 2023

Right to a Healthy Environment

A Brief (or not so brief) introduction
to Petscan



**We will be
recording
today!**

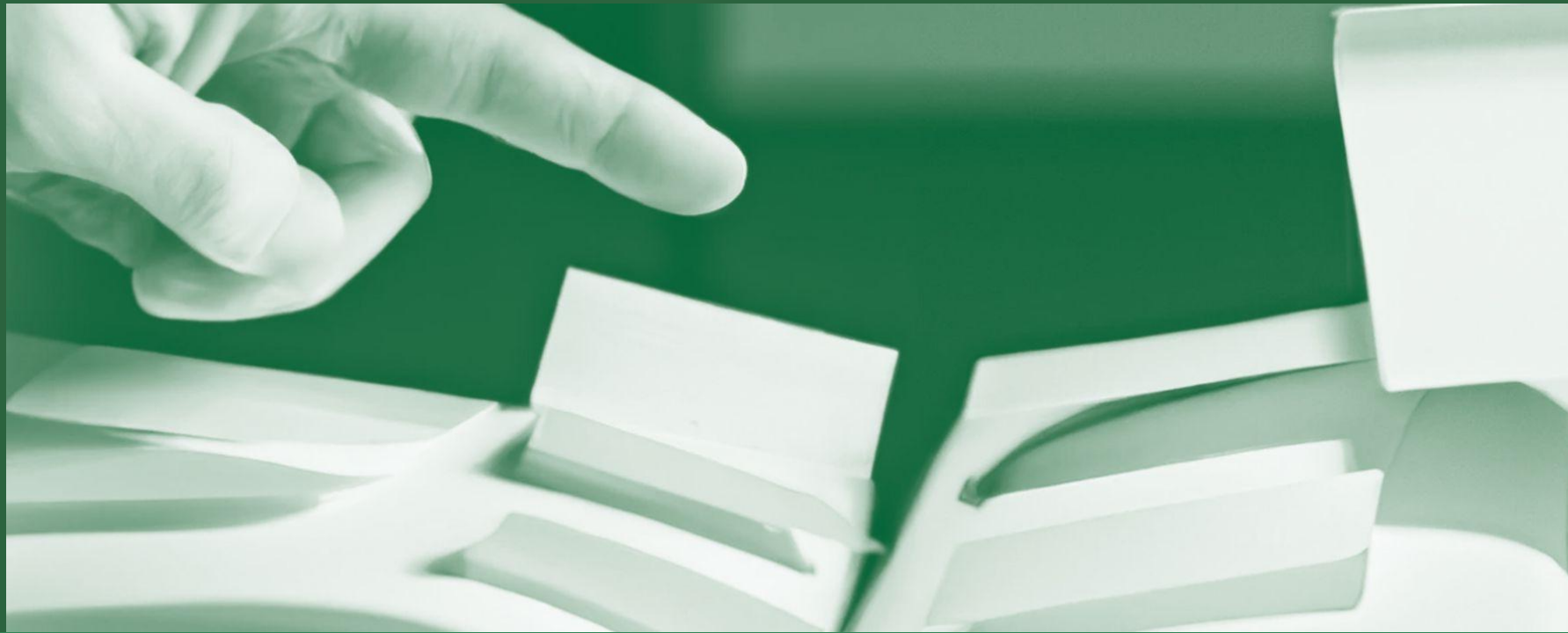
We plan to record this presentation today.

If you say something aloud or are on camera, it will like be in the recording.

If you share something in the chat, that will not be shared in the recording

Today's **agenda**

- Reminder of why you might use Petscan
- Orientation to Simple Petscan
- Using it with other tools
- Orientation to Petscan with Wikidata



What is Petscan?



What is it exactly?

Petscan is a Magnus Manske tool that

- mixes different kinds of data to create topical lists
- succeeded several tools focused only on Categories
- Adds features like filtering with Wikidata, editing Wikidata, and editing Wikimedia Commons

Why it's useful for organizers:

- Allows for an auto-updating list based on the criteria you set (if something changes on the wikis, it stays up to date).
- Helps take a massive body of knowledge and focus it

In the chat...

- Share examples of when you have seen Petscan in use?
- Why was it useful in that context?

You
probably
**want to use
Petscan
when:**

- You expect the topic list that you are building will change
- Other tools don't meet your needs (see last week)
- You need to edit Wikidata, Commons or use other tools w/ Pagepile



Item

Discussion

PetScan (Q23665536)

Using PetScan simply...

Categories Page properties Templates&links Other sources Wikidata Output

Language en Commons Wikispecies Wikidata

Project wikipedia

Depth 0 (depth 0 means no subcategories; appending "|" and a number to a category in the fields below will override it for that category)

Categories

Combination Intersection Union

Negative categories

Do it!

Let's look at the first three tabs



Categories Page properties Templates&links Other sources Wikidata Output

Language en Commons Wikispecies Wikidata

Project wikipedia

Depth 0 (depth 0 means no subcategories; appending "|" and a number to a category in the fields below will override it for that category)

Categories

Combination Intersection Union

Negative categories

Do it!

Pay attention to depth

The first tab is focused on categories and is most intuitive



▶ Climate change by country and region (2 C, 7 P)

*

▶ Climate change-related lists (19 P)

A

▶ Climate change assessment and attribution (1 C, 42 P)

E

▼ Climate change and the environment (3 C, 46 P)

▼ Algal blooms (1 C, 19 P)

▶ Red tide (3 P)

▶ Effects of climate change (3 C, 94 P)

▶ Land surface effects on climate (1 C, 4 P)

F

▶ Climate change feedbacks (1 C, 18 P)

G

▼ Greenhouse gases (4 C, 67 P)

▼ Carbon dioxide (3 C, 43 P)

▶ Applications of carbon dioxide (1 C, 8 P)

▶ Capnophiles (4 P)

▶ Carbon capture and storage (1 C, 31 P)

▶ Emissions reduction (7 C, 96 P)

▶ Methane (3 C, 50 P)

▶ Lakes of Titan (moon) (24 P)

▶ Natural gas (14 C, 57 P)

▶ Rocket engines using methane propellant (16 P)

M

▼ Climate change mitigation (6 C, 41 P)

▶ Carbon dioxide removal (7 C, 17 P)

▶ Climate change mitigation researchers (11 P)

▶ Emissions reduction (7 C, 96 P)

▶ Climate engineering (1 C, 29 P)

▶ Reforestation (2 C, 55 P)

▼ Sustainable energy (13 C, 110 P)

▶ Algaculture (1 C, 23 P)

▶ Alternative energy economy (2 C, 7 P)

▶ Compressed air power (10 P)

▶ Energy conservation (9 C, 164 P)

▶ Energy democracy (1 C, 1 P)

▶ Documentary films about alternative energy (16 P)

▶ Energy (4 P)

▶ Energy efficiency (3 C, 33 P)

▶ Energy harvesting (3 C, 19 P)

▶ Fuel cells (4 C, 81 P)

▼ Alternative fuels (5 C, 16 P)

▶ Biofuels (8 C, 71 P)

▶ Hydrogen (8 C, 48 P)

▶ Liquefied petroleum gas (11 P)

▶ Synthetic fuels (4 C, 24 P)

▶ Transit authorities with alternative-fuel vehicles (4 P)

▶ Renewable energy (19 C, 1221 P)

▶ Water fuel (1 C, 6 P)

S

▼ Climate change and society (16 C, 63 P)

▶ Climate activists (5 C, 222 P)

▶ Climate change adaptation (1 C, 30 P)

▶ Climate change and agriculture (2 C, 33 P)

▶ Climate change in art (2 C, 9 P)

▶ Climate change mass media (3 C, 21 P)

▶ Climate communication (1 C, 30 P)

▶ Climate change conferences (1 C, 24 P)

▶ Climate change controversies (2 C, 24 P)

▶ Economics and climate change (3 C, 41 P)

▶ Climate change in fiction (3 C, 26 P)

▶ Climate justice (1 C, 14 P)

▶ Climate change organizations (5 C, 79 P)

▶ Politics of climate change (2 C, 46 P)

▶ Climate change protests (20 P)

▶ Religious action on climate change (9 P)

▶ Climate change treaties (1 C, 6 P)

Σ

▶ Climate change stubs (129 P)

For example, I probably want most of the articles in these categories, but... several subcategories introduce problems...





Categories Page properties Templates&links Other sources Wikidata Output

Language

en

Commons Wikispecies Wikidata

Project

wikipedia

Depth

3

(depth 0 means no subcategories; appending "1" and a number to a category in the fields below will override it for that category)

Categories

Climate change

Combination

Intersection Union

Negative categories

Methane
Alternative fuels

Do it!

Link to a pre-filled form for *the query you just ran with* and *without* auto-run. PSID is **23934397**. Find images for Wikidata.

2936 results

The list is not bad...
<https://petscan.wmflabs.org/?psid=23934397>





Categories

Page properties

Templates&links

Other sources

Wikidata

Output

Language

en

Commons

Wikispecies

Wikidata

Project

wikipedia

Depth

3

(depth 0 means no subcategories; appending "1" and a number to a category in the fields below will override it for that category)

Categories

Climate change

Combination

 Intersection Union

Negative categories

Methane
Alternative fuels

Do it!

Link to a pre-filled form for *the query you just ran with* and *without* auto-run. PSID is **23934397**. Find images for Wikidata.

2936 results

But let's try filtering it...



Categories Page properties Templates&links Other sources Wikidata Output

Namespaces en.wikipedia User Wikipedia File MediaWiki Template Help Category Portal Draft

Talk User talk Wikipedia talk File talk MediaWiki talk Template talk Help talk Category talk Portal talk Draft talk

TimedText Module Gadget Gadget definition

TimedText talk Module talk Gadget talk Gadget definition talk

Size

Larger or equal (bytes) Smaller or equal (bytes)

<= % of first revision eg 120 for pages that are <=120% of first revision

Links

Minimum links # Maximum links #

Last edit

Before (or exactly) YYYYMMDDHHMMSS (shorter OK) After (or exactly) YYYYMMDDHHMMSS (shorter OK)

Max age hours since current UTC time Only pages created during the above time window (overrides "last revision")

Redirects Either Yes No **Soft redirects** Either Yes No **Disambiguation** Either Yes No

Last edit: Bots Either Yes No **Anons** Either Yes No **Flagged** Either Yes No

Lead image Either Yes, free or non-free Yes, free only Yes, non-free only No

ORES Probability range — Prediction Either Yes No

[Do it!](#) Link to a pre-filled form for *the query you just ran with* and *without* auto-run. PSID is [23934397](#). [Find images for Wikidata](#).

Bytes is a proxy for complexity and length of article -- lots of bytes implies more references

ORES uses other signals like references, other page features

Two important filters for organizers of writing events, others useful in other contexts....



Categories Page properties **Templates&link** Other sources Wikidata Output

Templates

Has all of these templates:

Use talk pages instead

Has any of these templates:

Use talk pages instead

Has none of these templates:

Use talk pages instead

Linked from

All of these pages:

Any of these pages:

None of these pages:

Links to

All of these pages:

Any of these pages:

None of these pages:

[Do it!](#)

Link to a pre-filled form for *the query you just ran* with and without auto-run. PSID is [23934397](#). [Find images for Wikidata](#).

You should have a theory about what kinds of topics should be covered in the text.

For example, if I think all topics that talk about agriculture or climate smart agriculture should talk about deforestation....

Let's filter this query Templates/Links page
<https://petscan.wmflabs.org/?psid=23934397>



Let me show you how I built another query about crops

https://petscan.wmflabs.org/?p_sid=23229381

Why did I want to build the list?

- Food choices are important for individual climate action
- Climate impacts are profound on food

So I needed to identify:

- Articles about food vulnerable to climate change (all of it)
- Remove any of the non-agriculture articles
- Make it easy to figure out what articles already talk about climate change

Want to clarify anything?

Now is the time to ask questions or share Comments about the first section of the presentation.

Outputs from Petscan

Item [Discussion](#) [Read](#) [View](#)

PetScan (Q23665536)

tool for querying Wikimedia databases and editing Wikidata
#petscan | Petscan

▼ In more languages
[Configure](#)

Language	Label	Description
English	PetScan	tool for querying Wikimedia databases and editing Wikidata
Spanish	PetScan	herramienta para Wikidata
Latin	No label defined	No description de

Categories Page properties Templates&links Other sources Wikidata Output

Format HTML CSV TSV Wiki JSON PagePile KML Plain text

JSON output CatScan QuickIntersection Sparse (prefixed page titles instead of objects) Pretty (not machine-readable)

Sort not at all (outputs by page ID = upload date for files) by title by namespace, title
 by size by date by incoming links (ns0) by file size by upload date by sitelinks randomly default sort

Sort order ascending descending

Files Extended data for files File usage data Main namespace only Generate thumbnails in Wiki output

Page metadata Image Coordinates Default sort Disambiguation

Regexp filter Only show titles/labels matching the regular expression

Search filter A search query that will be used as a filter on every single result.
Expensive, please use common sense

Redlinks Show redlinks Remove redlinks linked from templates Show only redlinks to main (article) namespace

Limit

[Do it!](#) [Link to a pre-filled form for the query you just ran with and without auto-run. PSID is 23934397. Find images for Wikidata.](#)

The 8 types of outputs are super useful, let me show you what they look like....



File 47791: 2,936 pages on enwiki. Other formats: [Plain text](#), [wikitext](#), [JSON](#), [JSON \(with metadata\)](#)

Use this pagepile in

1–100 | [101–200](#) | [End](#)

1. [Albedo](#)
2. [Arbor Day](#)
3. [Attribution of recent climate change](#)
4. [Geography of Botswana](#)
5. [Blade Runner](#)
6. [Benjamin D. Santer](#)
7. [Carbon dioxide](#)
8. [Cycling](#)
9. [Carbon sink](#)
10. [Convention on Long-Range Transboundary Air Pollution](#)
11. [El Niño](#)
12. [Greenhouse effect](#)
13. [Global warming potential](#)
14. [Global Climate Coalition](#)
15. [Global warming controversy](#)
16. [Intergovernmental Panel on Climate Change](#)
17. [Kyoto Protocol](#)
18. [Geography of Morocco](#)
19. [Naomi Klein](#)
20. [Ozone](#)
21. [The Three Stigmata of Palmer Eldritch](#)
22. [Polar bear](#)
23. [Piezoelectricity](#)
24. [Renewable energy](#)

For example, we can take the list to Massviews and learn which articles have the most pageviews ([link](#))

Or can take it too tools like WDFIST to [enrich information on Wikidata](#)

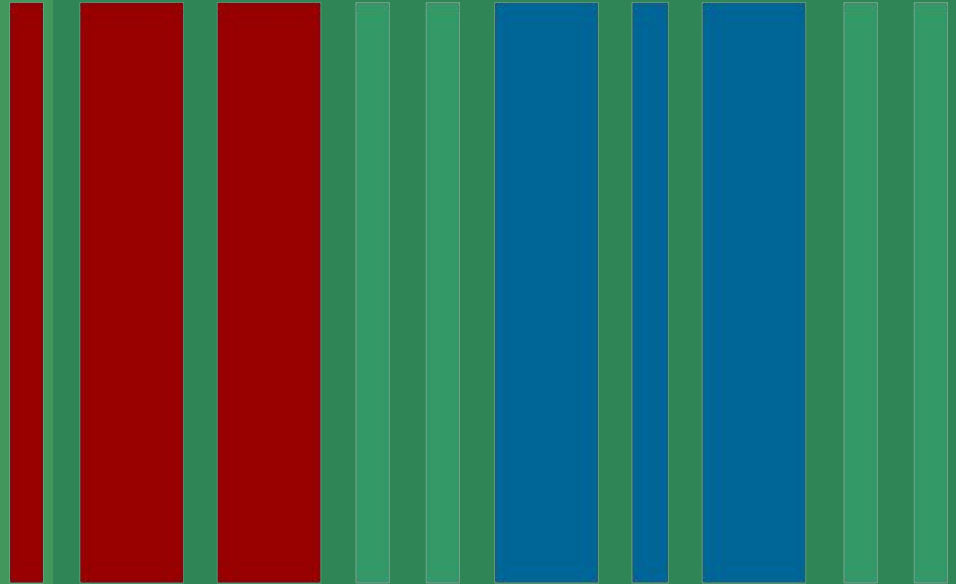
Or if you convert the output to add Wikidata, edits via [Tabernacle](#)

Particularly pagepile, which makes the list portable....

https://pagepile.toolforge.org/api.php?action=get_data&id=47791



Using Petscan w/ Wikidata or more complex actions



Categories Page properties Templates&links **Other sources** **Wikidata** Output

Language Commons Wikispecies Wikidata

Project

Depth (depth 0 means no subcategories; appending "|" and a number to a category in the fields below will override it for that category)

Categories

Combination Intersection Union

Negative categories

We skipped two tabs... because they are mostly for Wikidata and Commons



Categories Page properties Templates&links Other sources **Wikidata** Output

Pages with items Do not add Wikidata items Add items, where available Only pages with item
 Only pages without item (enables Wikidata item creation mode)

Label language Defaults to interface language

Uses items/props Q123,P456,... Any All None
 Has no statements Has no sitelinks

Site links
*Note: The options below will be used as a generator, if and only if no other generator is used!
Also, using it as a generator can be quite slow and resource-intensive, please use common sense.*

Has all of these site links Has any of these site links Has none of these site links

Site names, one per row (e.g. enwiki, dewikisource)

Minimum number of site links Maximum number of site links

Labels etc.
Note: The options below will be used as a generator, if and only if no other generator is used!

Has all of these labels Has any of these labels Has none of these labels

One per line, SQL LIKE syntax (e.g., "Bob %")

Label Alias Desc. Label Alias Desc. Label Alias Desc.

This first widget allows you to add the Wikidata output to a query -- important for most Wikidata based tools.

This section allows you to find out which wikis are missing for translation

Let's look at the Wikidata tab first



SPARQL

Input SPARQL query here

Manual list

Enter one (namespace-prefixed) page title or Wikidata item per line here

Wiki e.g. enwiki, dewikisource, enwik

PagePile

PagePile ID

Search

Search query wiki (e.g. enwiki; Max results 500

Wikidata

enwiki,dewiki,.. Any sitelink

This source provides some large-scale access to Wikidata that SPARQL queries do not provide sufficiently

Namespaces

Keep namespace Change to page/topic Change to talk page

Subpages

Add subpages Filter subpages: Either Only subpages No subpages

Use wiki

Automatic From categories From PagePile From manual list Wikidata

e.g. enwiki, dewikisource, enwi

This first section allows you to grab results from other queries (i.e. pagepile, search, the query service, etc). In my experience, I use SPARQL most, but all have been helpful.

The Other Sources tab has two types of functions



Manual list

Enter one (namespace-prefixed) page title or Wikidata item per line here

Wiki e.g. enwiki, dewikisource, enwik

PagePile

PagePile ID

Search

Search query

wiki (e.g. enwiki)

Max results

500

Wikidata

enwiki,dewiki,...

Any sitelink

This source provides some large-scale access to Wikidata that SPARQL queries do not provide sufficiently

Namespaces

Keep namespace Change to page/topic Change to talk page

Subpages

Add subpages Filter subpages: Either Only subpages No subpages

Use wiki

Automatic From categories From PagePile From manual list Wikidata

e.g. enwiki, dewikisource, enwi

Combination

optional, e.g. "categories NOT sparql"; leave blank for "AND" all used sources

You can combine sources with a simple syntax. By default (leave field empty), all used sources are combined with AND.

You can use AND, OR, and NOT, as well as () to combine sources, as in categories NOT (sparql OR pagepile).

Valid sources are categories, sparql, manual, pagepile, search, and wikidata

Do it!

This section allows you to change the output from lists

Notice that this combination feature allows you to use machine logic to combine results. I find this feature breaks a lot of my queries...

The Other Sources tab has two types of functions



Locations in Nigeria that aren't stubs without climate change links

Scenario:

- I am running editathons in Nigeria about impacts of climate change
- I want editors to write about climate change in city pages while improving articles that already exist

Here are the results:

- Main Query:
<https://petscan.wmflabs.org/?psid=23935851>
- Minimum size of article 5000 bytes:
<https://petscan.wmflabs.org/?psid=23935901>
- [Pagepile](#)
- [Massviews of articles bigger than 5000 bytes](#)

Want me to clarify anything?

Now is the time to ask Questions or share Comments.

**We can use the rest of the time
to try the tool and give you
feedback/answer questions.**