



Hi, my name is Minh, and you're going to learn about integrating Wikidata with OpenStreetMap.



OSM is the wiki world map. Like Wikimedia's projects, OSM is ambitious in its scope and radical in its openness to volunteer contributions.



© Shane Adams, CC BY 2.0

If you dive into raw OSM data for the first time, your first reaction is likely to be... colorful. The data is amazingly uneven.



A mess, even. But the most important thing is that we're collecting it.

operator

Company, corporation, person or any other entity who is directly in charge of the current operation of a map object

[Overview](#)
[Values](#)
[Combinations](#)
[Similar](#)
[Map](#)
[Wiki](#)
[Projects](#)

Values used with this key

[←](#)
[▶](#)
Page 1 of 4
 [↻](#)
[JSON](#)
 Displaying 1 to 15 of 55 items
 [postal](#)

Value	Count	Wiki	Description
United_States_Postal_Service	1 838 0.29%	-	
US_Postal_Service	451 0.07%	-	
U.S._Postal_Service	163 0.03%	-	
United_State_Postal_Service	15 0.00%	-	
Unites_States_Postal_Service	8 0.00%	-	
Disneyland/Anaheim_Postal	7 0.00%	-	
United_States_Postal_Office	6 0.00%	-	
U.S._Postal_Service	6 0.00%	-	
United_States_Postal_System	5 0.00%	-	
Postal_Annex	4 0.00%	-	
US_Postal_Office	4 0.00%	-	
United_Postal_Service	4 0.00%	-	
United_States_Postal_Service_(USI	4 0.00%	-	
Walt_Disney_World/Orlando_Postal	3 0.00%	-	
Disneyland/Anahime_Postal	3 0.00%	-	









For example, here are all the different ways people have referred to the U.S. Postal Service in OSM. It's this messy because OSM is built on a folksonomy of tags, kind of like MediaWiki categories.

761	<pre>"amenity": "bank" "brand": "BB&T" "brand:wikidata": "Q95984154" "name": "BB&T" "official_name": "Branch Banking and Trust Company"</pre>	BB&T "former Bank Brand" Q95984154 http://www.bbt.com/ 	
161	<pre>"amenity": "bank" "brand": "BBVA Compass" "brand:wikidata": "Q4835088" "brand:wikipedia": "en:BBVA Compass" "name": "BBVA Compass" "official_name": "BBVA Compass Bancshares"</pre>	BBVA USA "American bank holding company" Q4835088 https://www.bbvausa.com/ 	
< 50	<pre>"amenity": "bank" "brand": "BECU" "brand:wikidata": "Q4835409" "brand:wikipedia": "en:BECU" "name": "BECU"</pre>	BECU "Credit union headquartered in Tukwila, Washington, USA" Q4835409 https://www.becu.org/ 	
202	<pre>"amenity": "bank" "brand": "BMO Harris Bank" "brand:wikidata": "Q4835981" "brand:wikipedia": "en:BMO Harris Bank" "name": "BMO Harris Bank"</pre>	BMO Harris Bank "US bank" Q4835981 https://www.bmoharris.com/ 	

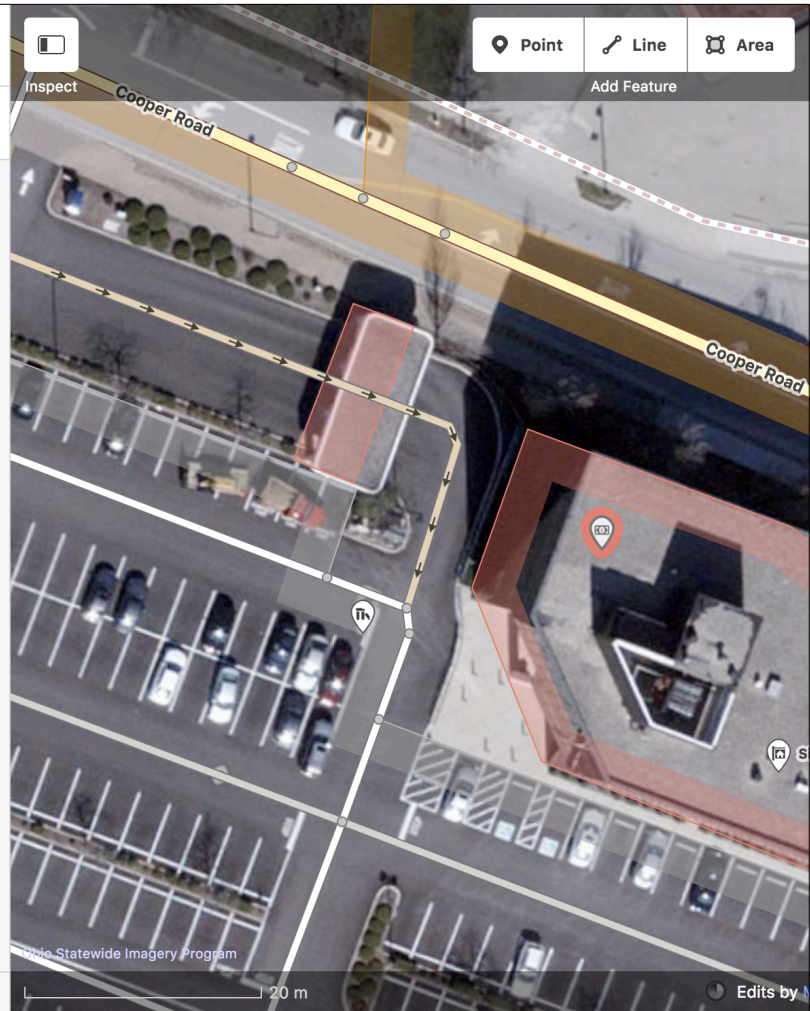
Each preset in NSI comes with a set of tags, including machine-readable Wikidata QIDs.

42 results for u.s. >

u.s.

	U.S. Bank ATM	i
	U.S. Bank Bank	i
	U.S. Cellular Mobile Phone Store	i
	U.S. Bureau of Reclamation Power Generator	i
	Bus Stop	i
	Drugstore	i
	Fuel Shop	i
	Guidepost	i

[View on openstreetmap.org](#)



Cooper Road

Cooper Road

Statewide Imagery Program

20 m

Edits by


When you map something in OSM, these presets make it easy for you to use the standardized raw tags.

Language	Label	Description	Also known as
English	BBVA USA	American bank holding company	BBVA Compass Bancshares, Inc. BBVA Compass
Spanish	BBVA Estados Unidos	holding bancario	Compass Bancshares
Vietnamese	No label defined	No description defined	
Arabic	No label defined		شركة أمريكية
South Azerbaijani	بهره‌برداري کاسپس	No description defined	
Czech	No label defined	americká firma	
German	BBVA USA	No description defined	
Persian	بهره‌برداري کاسپس	No description defined	
French	BBVA USA	No description defined	
Italian	BBVA USA	No description defined	
Dutch	BBVA Compass	bedrijf uit Verenigde Staten van Amerika	
Chinese	BBVA罗盘公司	银行控股公司	
Simplified Chinese	BBVA罗盘公司	No description defined	
Traditional Chinese	BBVA羅盤公司	銀行控股公司	

[Fewer languages](#)

Statements

Instance of business edit
 - 0 references + add reference
 + add value

Logo image edit

 BBVA,2019.png
 512 × 153; 3 KB
 - 0 references + add reference
 + add value

Industry Finance edit
 + 1 reference + add value

Inception 1964 edit
 + 1 reference + add value

founded by Henry B. Brock, Jr. edit

ISO 9362 SWIFT/BIC code CPASUS4400X edit
 + 1 reference + add value

Facebook ID bbvaintheusa edit
 + 0 references + add reference
 + add value

Instagram username bbva_usa edit
 + 0 references + add reference
 + add value

LinkedIn company ID bbva-usa edit
 + 0 references + add reference
 + add value

Microsoft Academic ID 883594155 edit
 + 0 references + add reference
 + add value

OSM Name Suggestion Index Identifier bbvacompass-es2e2d edit
 + 0 references + add reference
 + add value

PermID 4296525156 edit
 + 0 references + add reference
5000051177 edit
 + 0 references + add reference
 + add value

Twitter username bbva_usa edit
 Twitter user numeric ID 16349341

On Wikidata, the brand has statements about many aspects of the brand that would never be considered germane to OSM.



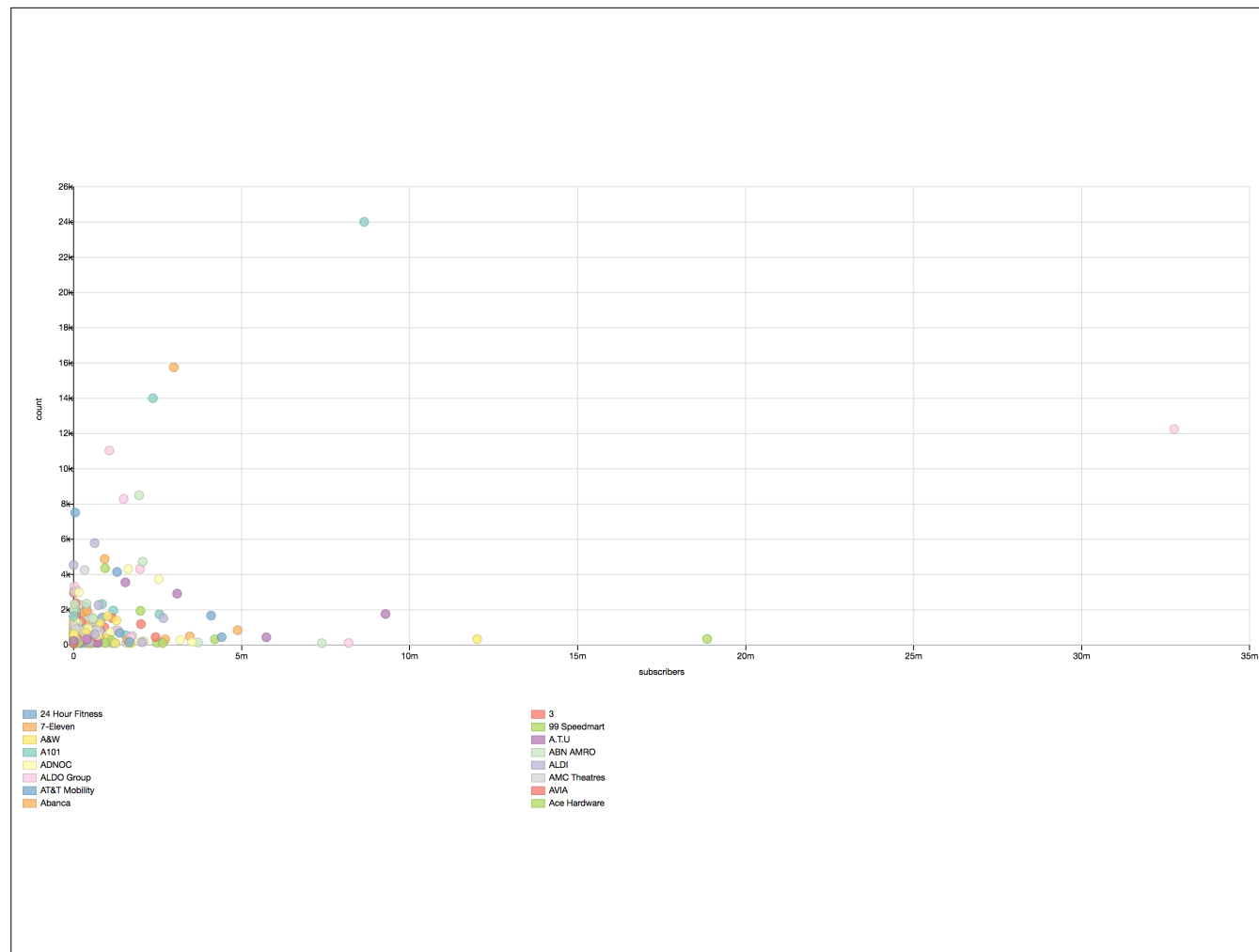
© Washuotaku, CC BY-SA 4.0

What's the point of tagging OSM features with Wikidata QIDs? It's all about being able to analyze OSM data with federated queries, aka crosswalks.

The screenshot shows the OSM Sophox web interface. At the top, there is a header with the logo, the text "OSM Sophox", and three menu items: "Examples", "Help", and "Tools". Below the header is a text editor area containing a SPARQL query. The query is as follows:

```
1 #defaultView:ScatterChart
2 SELECT (SUM(?subscribers) AS ?subscribers) (SAMPLE(?count) AS ?count) (SAMPLE(?name) AS ?name) WHERE
3   hint:Query hint:optimizer "None" .
4
5 {
6   # Query OpenStreetMap for
7   SELECT ?wd (COUNT(*) AS ?count) WHERE {
8     # Commercial brands
9     ?osm osmt:brand:wikidata ?wd.
10  }
11  GROUP BY ?wd
12  # With at least 100 locations
13  HAVING(?count >= 100)
14 }
15
16 # Query Wikidata for
17 SERVICE <https://query.wikidata.org/sparql> {
18   # The brand's Twitter user name
19   ?wd p:P2002 [ps:P2002 ?twitter; pq:P3744 ?subscribers].
20
21   # Get the brand name
22   OPTIONAL {
23     ?wd rdfs:label ?name.
24     FILTER(LANG(?name) = "en")
25   }
26 }
27 }
28 GROUP BY ?wd
29 ORDER BY DESC(?count)
```

This is Sophox, a fork of the Wikidata Query Service that Yuri Astrakhan set up. You use the SPARQL query language to query both OSM and Wikidata at the same time.



The query you just saw plots brands by the number of Twitter subscribers along the X axis and the number of store locations mapped in OSM along the Y axis. You can see Starbucks standing out on Twitter and McDonald's standing out on OSM.

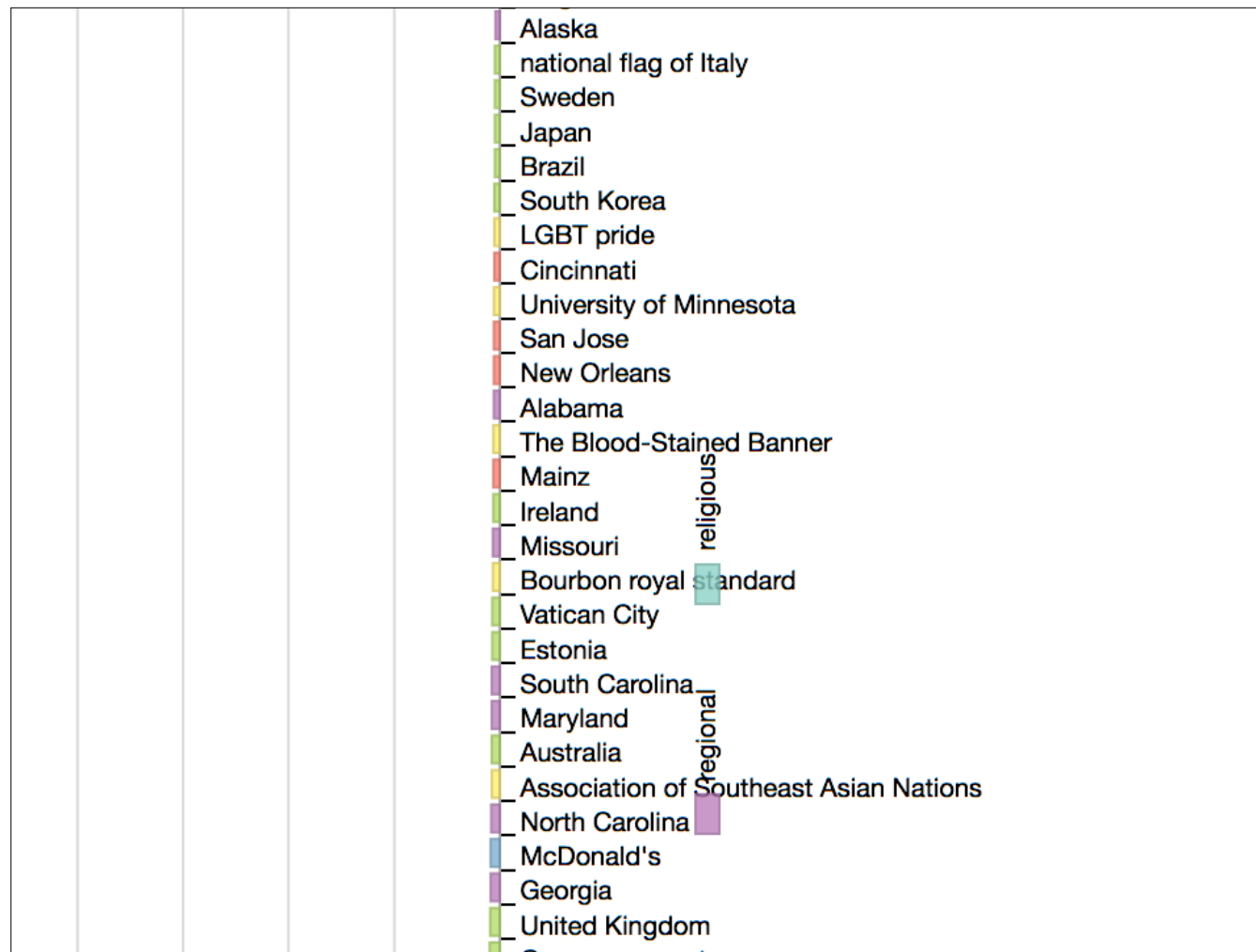


Here's a different view of the same data: the logos of those brands.

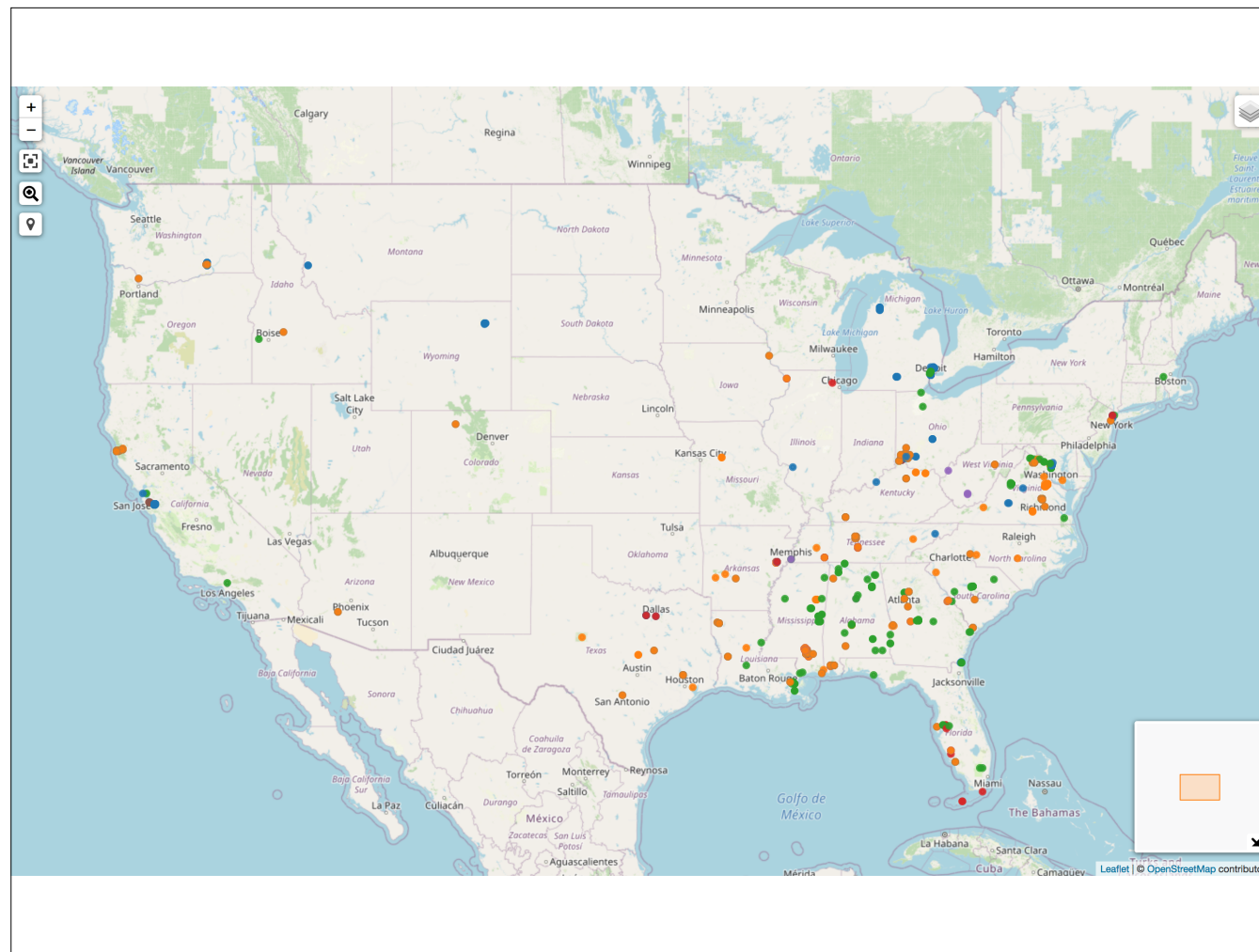


© Oregon Department of Transportation, CC BY 2.0

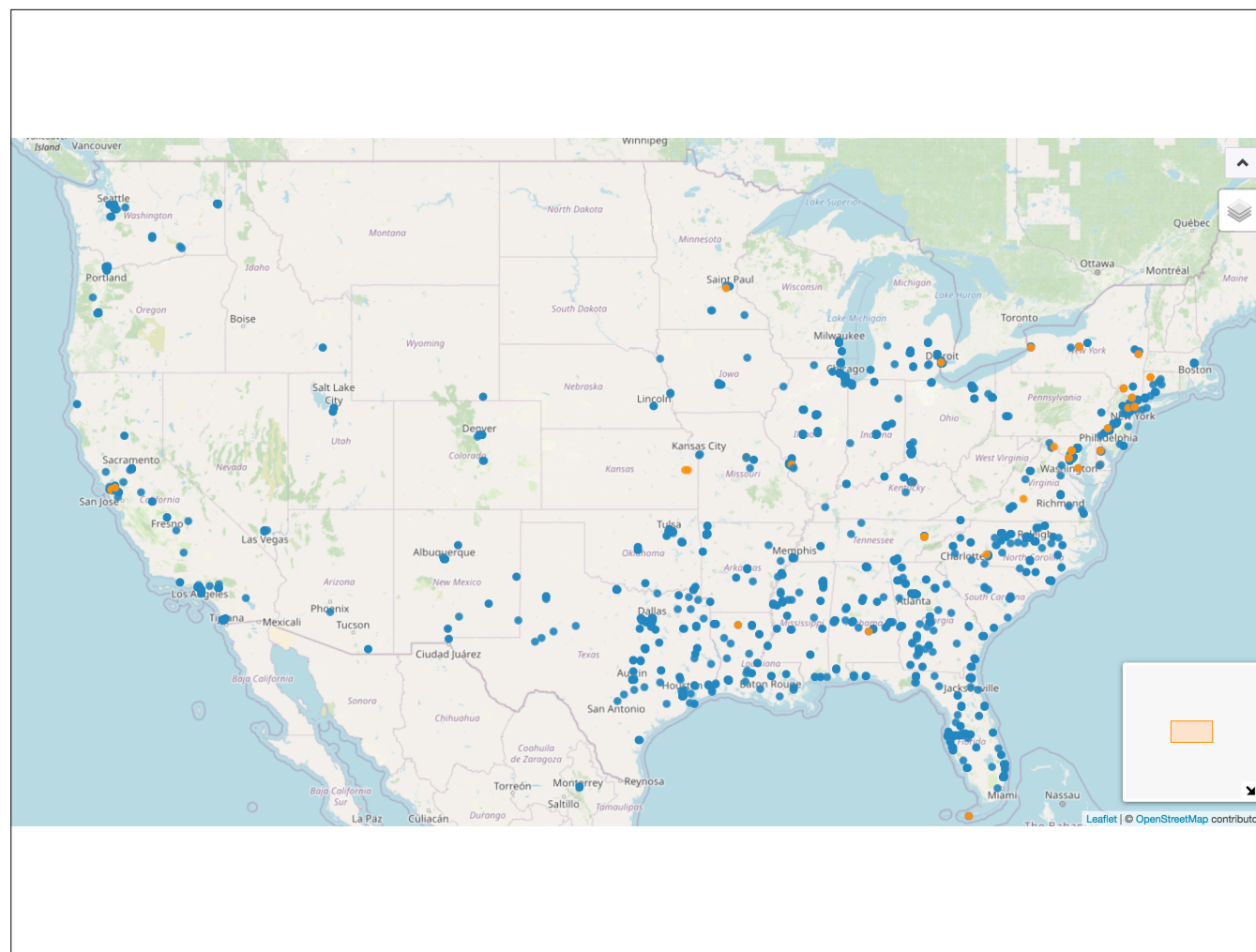
But federated queries are about much more than branding.



And there's a long tail of obscure flags, like the Bourbon royal standard, and common ones we haven't mapped enough of yet, like LGBT pride flags and McDonald's flags.



These queries aren't just for our visual enjoyment. Here's a map of memorials and other namesakes of notable Confederates, slave traders, segregationists, etc. The memorials and their locations come from OSM but the namesakes' identities come from Wikidata. Not all of these things are even notable enough for Wikidata – small-town streets, shops, and even named rooms inside buildings.



On the positive side, here's a map of memorials and namesakes of Black human rights activists and abolitionists.

The screenshot shows the GitHub interface for the repository 'Sophox / sophox'. At the top, there are navigation links for 'Why GitHub?', 'Team', 'Enterprise', 'Explore', 'Marketplace', and 'Pricing'. A search bar and 'Sign in' / 'Sign up' buttons are also present. Below the repository name, there are buttons for 'Notifications', 'Star' (23), and 'Fork' (5). The main navigation bar includes 'Code', 'Issues' (16), 'Pull requests' (1), 'Actions', 'Security', and 'Insights'. A search bar contains the text 'is:issue is:open'. To the right of the search bar are filters for 'Labels' (7) and 'Milestones' (0), along with a 'New issue' button. The main content area displays a list of 16 open issues, each with a title, a description, and a comment icon. The issues are:

- OSM data is stale, last updated atleast 22 days ago (1 comment)
- Nodes are not listed in results, when they are part of relation (2 comments)
- Many subjects are missing (2 comments)
- OpenHistoricalMap
- Tree view only maps Wikidata entities to tree nodes (1 comment)
- GeoJSON export (2 comments)
- Feature/Data Request: changeset (2 comments)
- Most examples are uncategorized in Examples panel (1 comment)
- Support GeoSPARQL functions (1 comment)
- Return full geometries
- Reenable tabular data (2 comments)
- Enable gzip (and brotli) compression

Unfortunately, Sophox isn't in great shape these days. In the past few months, it has developed some serious issues including missing data that impacts the queries I just showed you. The tool is running on Google Cloud but not very well. We're looking for a suitable host for it.

Keep Going

- OpenStreetMap
openstreetmap.org
- Name Suggestion Index
nsi.guide
- Sophox
sophox.org
- SPARQL examples
wiki.osm.org/wiki/SPARQL_examples
- [#wikimedia](#) [#sophox](#) on OSMUS Slack

© [Infrogmation](#), CC BY 2.0

Here are links to the sites I showcased in this talk, along with a wiki page containing all the Sophox queries and more. Thank you!