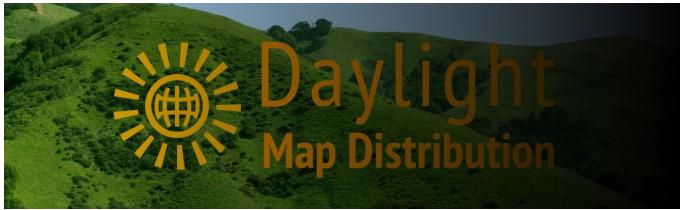


Goodbye Daylight, Hello Overture

Migrating from one QA'd OpenStreetMap Distribution to Another

Friday, January 24, 2025

Jennings Anderson, Software Engineer @ Meta



Mar 2020: v0 → 66 releases over 4 years → v1.58 (Nov 2024)

Daylight was a regularly released distribution of OSM data that came with a quality guarantee:

Meta's mapping team took regular snapshots of OSM and ran a number of data validation checks for broken polygons, vandalism, impossible intersections, etc., and fixed them.

Every couple of weeks, Meta published the resulting planet file as the next version of Daylight. This was the same data that powered their basemaps.

So what do I do if I still want QA'd OpenStreetMap Data?

What Now?

There is no longer a single ***daylight.osm.pbf***

(Note that Daylight was finding the upper limits of what was reasonable in a PBF when combining with additional datasets, such as Microsoft / Google ML Building footprints)

Today, we have



OVERTURE MAPS
FOUNDATION

Overture compiles data from a variety of open data sources, including **OSM**—AND has the same quality guarantees in place.



6 data themes

1. addresses
2. **base***
3. **buildings*** (~ 2.5B buildings: OSM + Esri + Microsoft + Google)
4. **divisions***
5. places
6. **transportation***

***themes include QA'd OpenStreetMap data.**



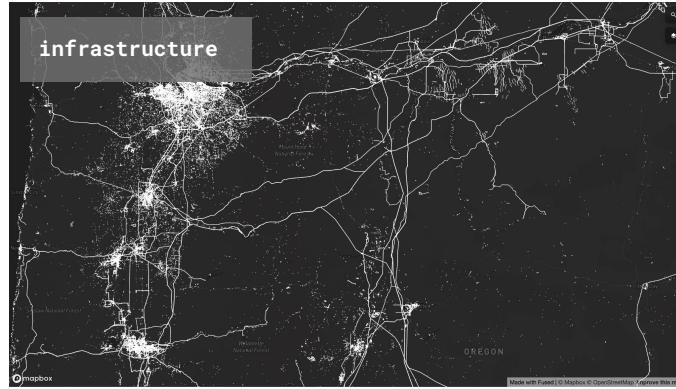
6 data themes

1. addresses
2. **base***
3. **buildings***
4. **divisions***
5. places
6. **transportation***

***themes include QA'd OpenStreetMap data.**

The Overture **Base** Theme

1. Infrastructure



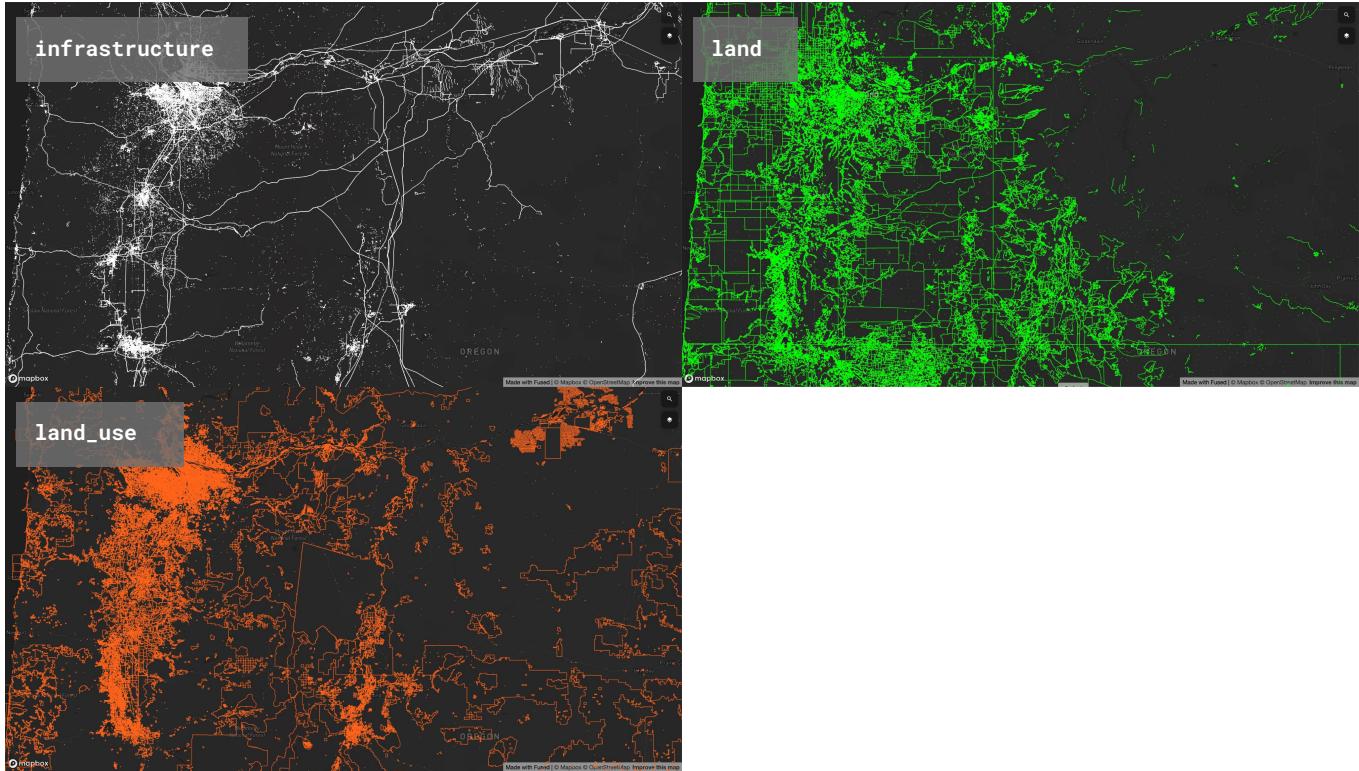
The Overture **Base** Theme

1. Infrastructure
2. land



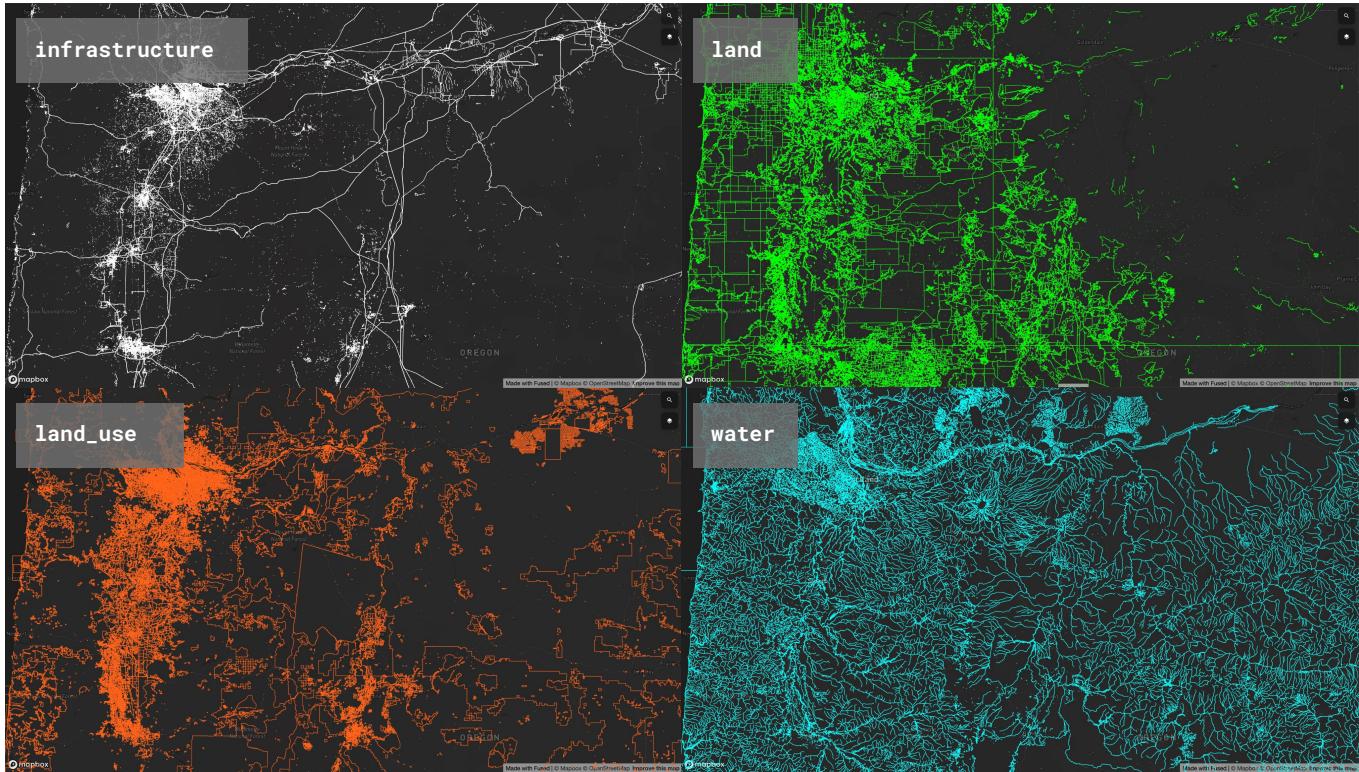
The Overture **Base** Theme

1. Infrastructure
2. land
3. land_use



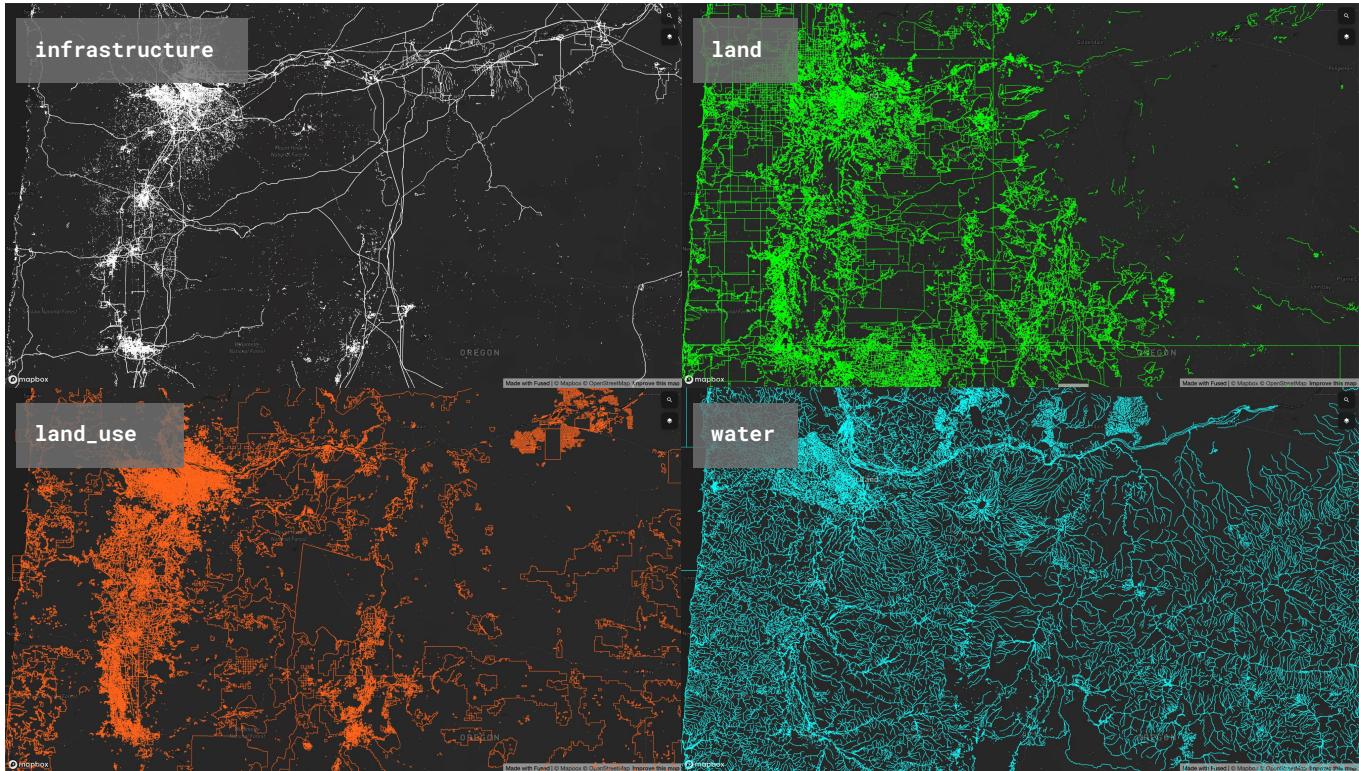
The Overture **Base** Theme

1. infrastructure
2. land
3. land_use
4. water

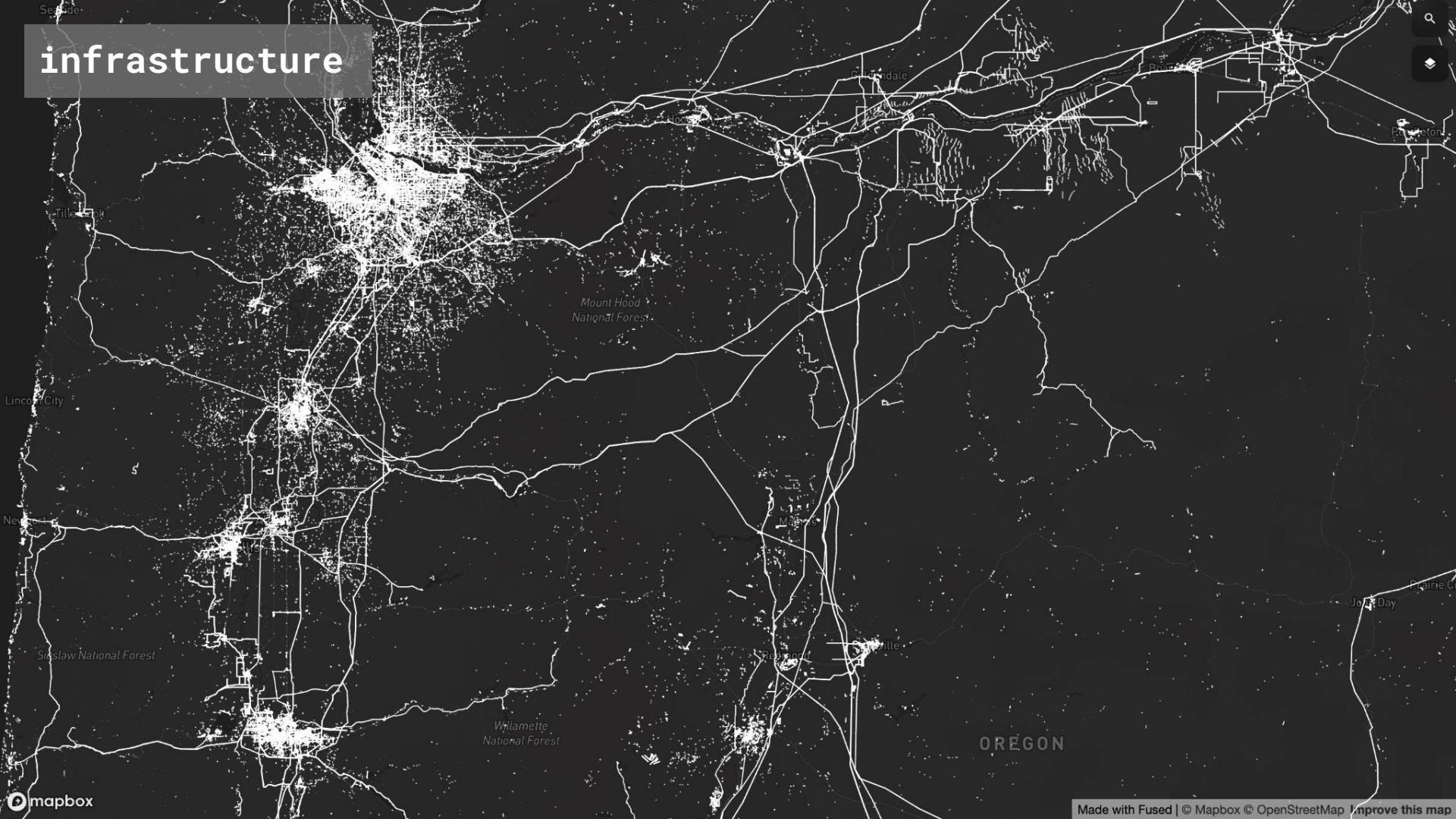


The Overture **Base** Theme

1. infrastructure
2. land
3. land_use
4. water
5. bathymetry
6. land_cover



infrastructure

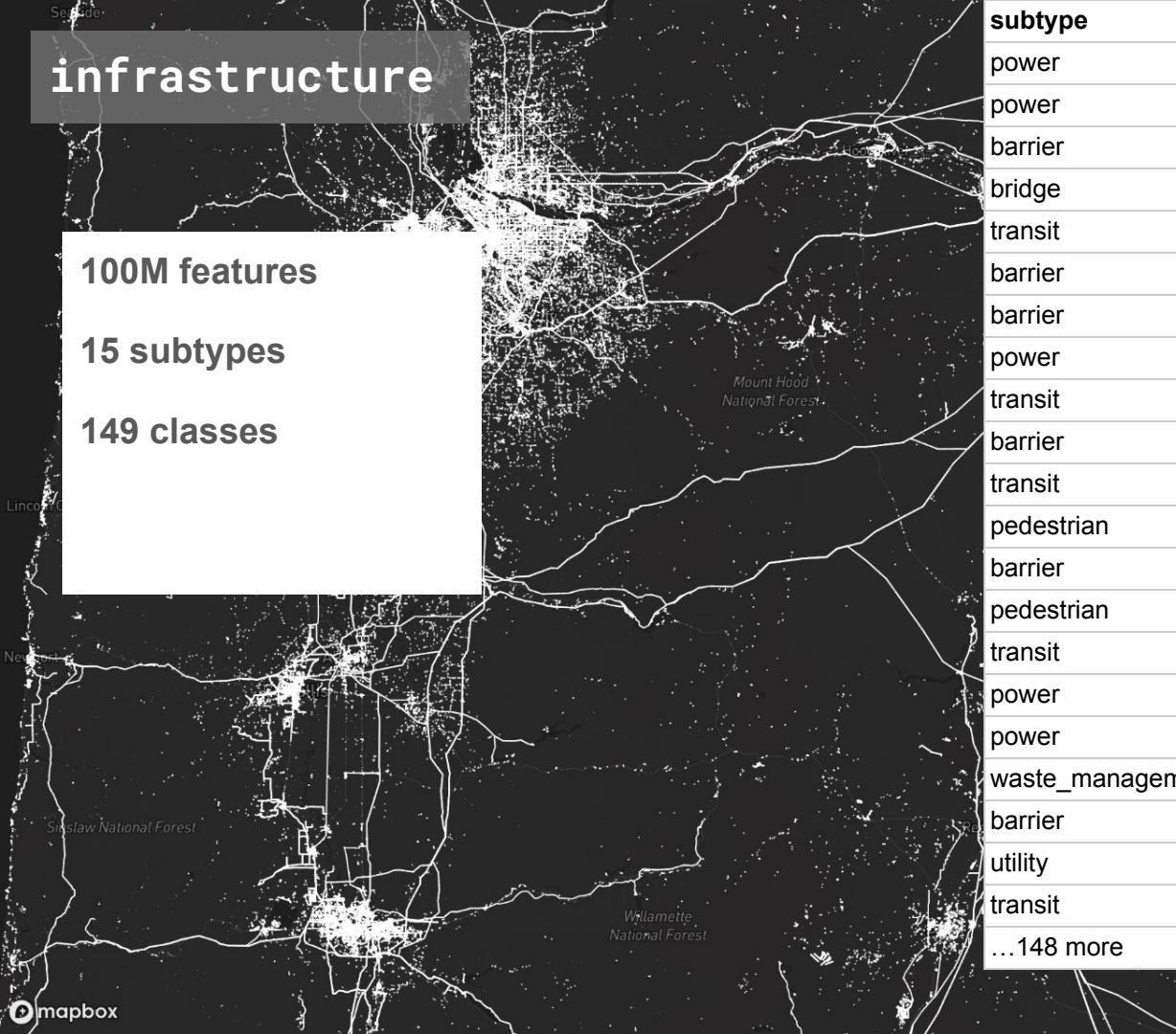


infrastructure

100M features

15 subtypes

149 classes



subtype	class	features
power	power_tower	16,749,835
power	power_pole	15,228,275
barrier	fence	7,846,924
bridge	bridge	6,664,298
transit	parking	5,728,020
barrier	gate	5,432,815
barrier	wall	5,045,885
power	generator	3,793,111
transit	bus_stop	3,714,931
barrier	hedge	3,492,388
transit	parking_space	3,140,407
pedestrian	bench	2,628,185
barrier	kerb	1,510,289
pedestrian	information	1,239,715
transit	stop_position	1,166,417
power	minor_line	1,044,782
power	power_line	1,002,257
waste_management	waste_basket	911,044
barrier	bollard	730,077
utility	storage_tank	725,419
transit	bicycle_parking	724,346
...148 more		...

infrastructure

subtype	transit
class	parking
names	
source_tags	access=customers amenity=parking fee=no parking=surface
sources	dataset=OpenStreetMap record_id=w39418397@3 update_time=2023-11-17



infrastructure

subtype	transit
class	parking
names	
source_tags	access=customers amenity=parking fee=no parking=surface
sources	dataset=OpenStreetMap record_id=w39418397@3 update_time=2023-11-1

The screenshot shows the OpenStreetMap website interface. The top navigation bar includes the OpenStreetMap logo, a search bar, and user account information for Jennings Anderson. Below the header, there's a search bar with 'Where is this?' and a magnifying glass icon, followed by a 'Search' button and a 'Map' button.

The main content area displays a map of a parking area. A red outline highlights a specific parking lot labeled 'Franks-A-Lot'. Other features visible on the map include 'Boca', 'Fonda Rosa', 'Whole Foods Market', 'Starbucks', 'Papa Dee', 'Quest Center for Integrative Health', and several street names like 'Northeast Couch Street', 'Northeast 29th Avenue', and 'East Burnside Street'. There are also building footprints with numbers like 103, 2903, 2907, and 28.

Way: 39418397

Version #5

Specify whether parking requires a fee

Edited 4 months ago by MiaVM
Changeset #156751272

Tags

access	customers
amenity	parking
fee	no
parking	surface

Nodes

► 11 nodes

[Download XML](#)

« Version #1 · View History · Version #5 »

Seattle



land

Goldendale

Burdman

Pendleton

Lincoln City

Newton

Okanogan National Forest

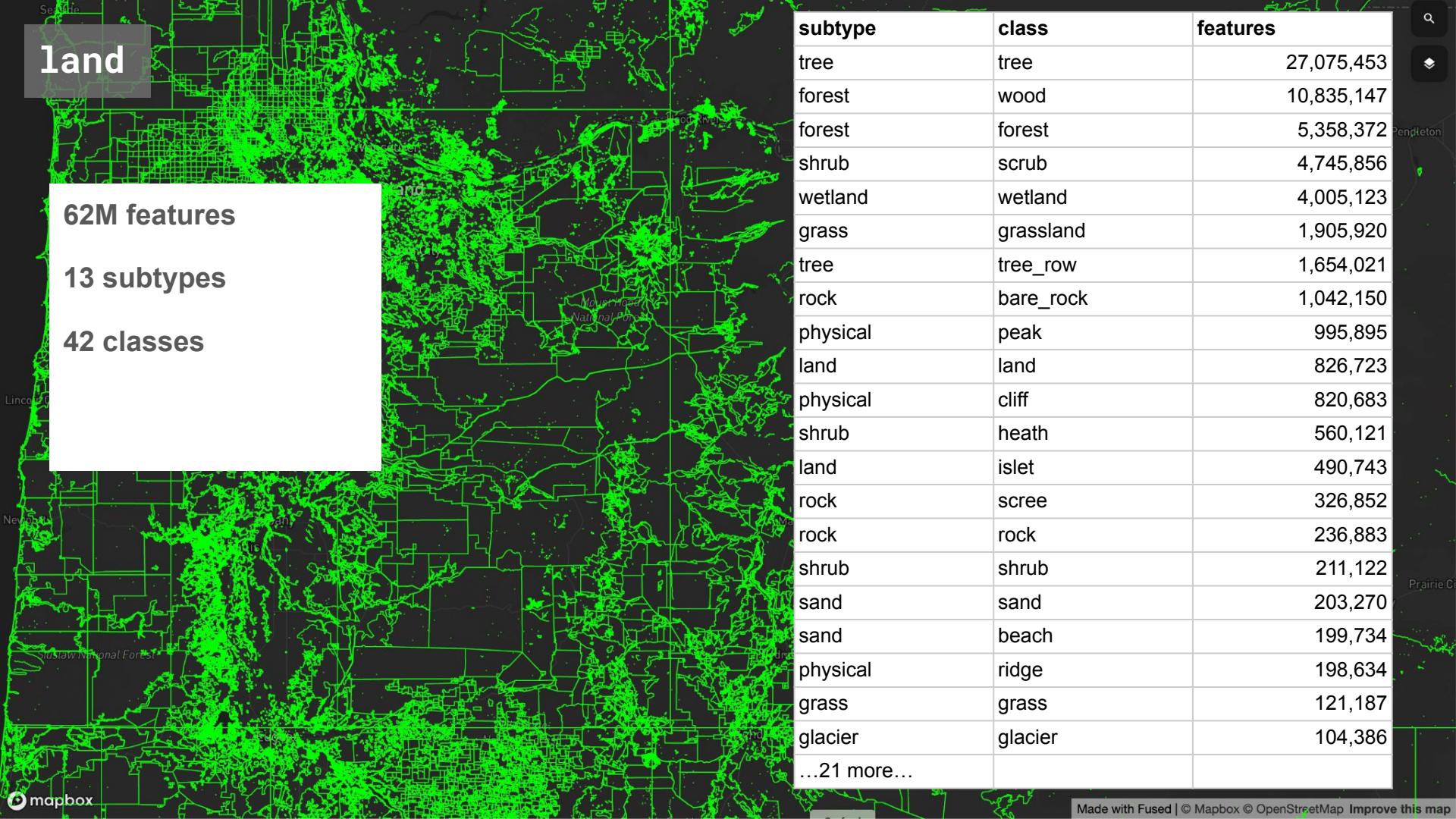
MOUNT
HOOD
National Park

McMinnville

Roseburg

OREGON

Prairie City
John Day



land

subtype	forest
class	wood
source_tags	natural=wood type=multipolygon
sources	dataset=OpenStreetMap record_id=r8679820@1 update_time=2018-09-11



Seattle

land

subtype	forest
class	wood
source_tags	natural=wood type=multipolygon
sources	dataset=OpenStreetMap record_id=r8679820@1 update_time=2018-09-11

Way: Columbia River (163917830) | OpenStreetMap

Relation: 8679820 | OpenStreetMap

More Jennings Anderson

Relation: 8679820

Version #1

Added woods & scrub to Sandy River Delta

Edited over 6 years ago by Adam Schneider
Changeset #62496579

Tags

natural	wood
type	multipolygon

Members

▼ 2 members

- Way 624784283 as outer
- Way 624784279 as inner

[Download XML](#)

[View History](#)

Columbia River Gorge National Scenic Area

Sandy River

Sundial Island

Mount Hood National Forest

Sandy River Delta

North Troutdale

Portland-TROUTDALE Airport

I-84 US 30

500 m 2000 ft

© OpenStreetMap contributors • Make a Donation. Website and API terms

Made with Fused | © Mapbox © OpenStreetMap Improve this map

Seaside

1

land_use

Goldendale

Pentafon

Lincoln

Mount Hood
National Forest

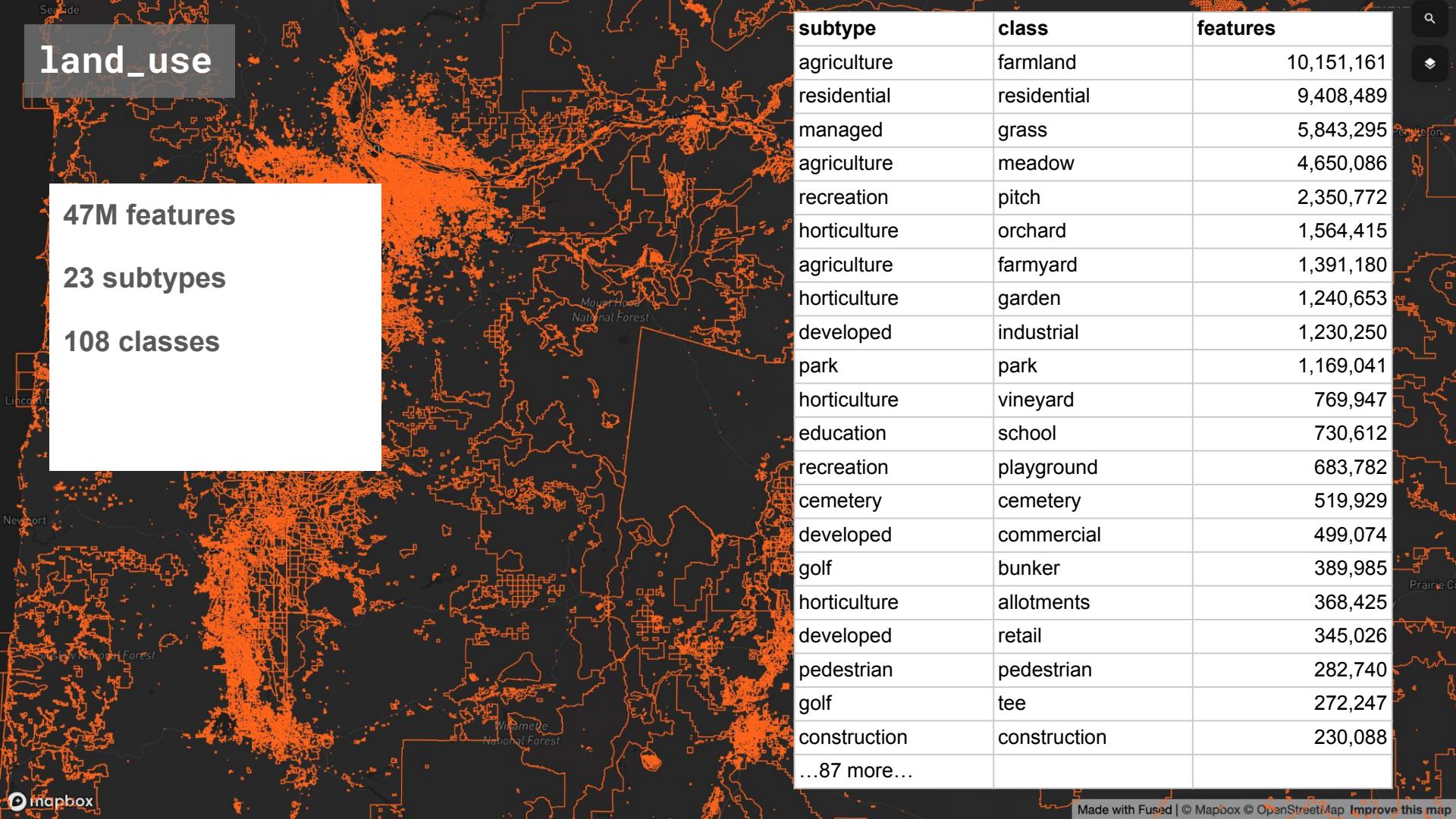
Newsoft

卷之三

Prairie C

OREGON



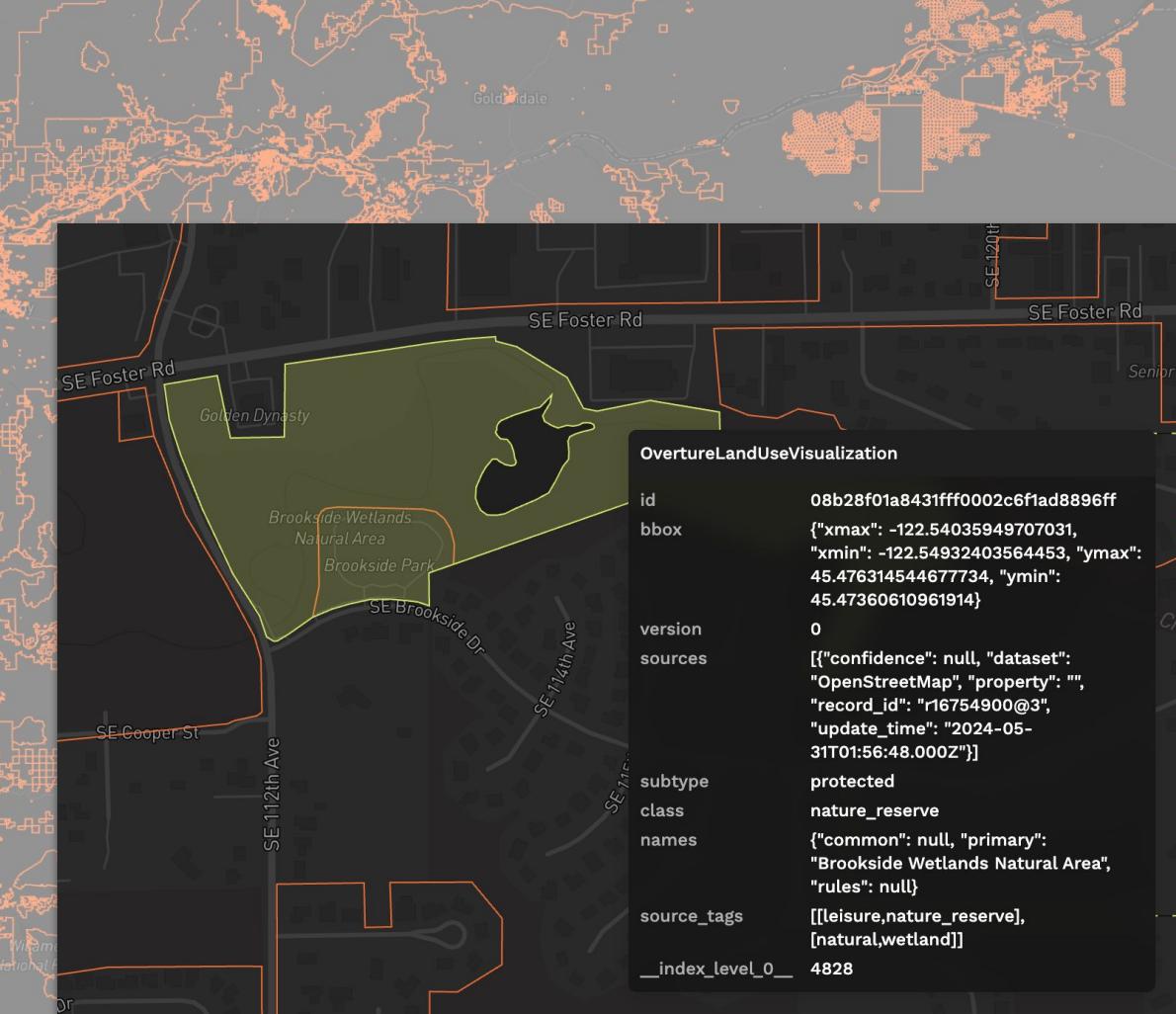


Seaside



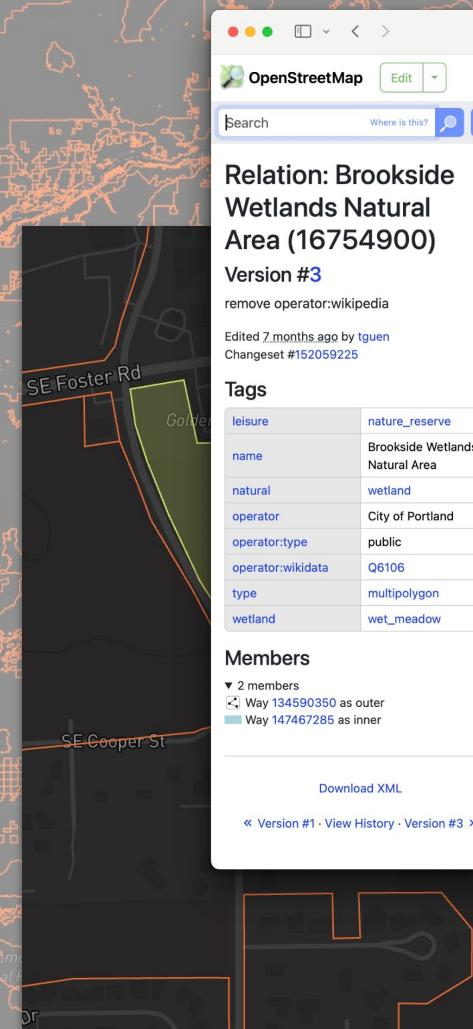
land_use

subtype	protected
class	nature_reserve
names - primary	Brookside Wetlands Natural Area
source_tags	leisure=nature_reserve natural=wetland
sources	dataset=OpenStreetMap record_id=r16754900@3 update_time=2024-05-31



land_use

subtype	protected
class	nature_reserve
names - primary	Brookside Wetlands Natural Area
source_tags	leisure=nature_reserve natural=wetland
sources	dataset=OpenStreetMap record_id=r16754900@3 update_time=2024-05-31



OpenStreetMap [Edit](#)

Search Where is this? [Help](#)

History Export GPS Traces User Diaries Communities Copyright Help About [Jennings Anderson](#)

Relation: Brookside Wetlands Natural Area (16754900)

Version #3

remove operator:wikipedia

Edited 7 months ago by [tguen](#)
Changeset #15205925

Tags

leisure	nature_reserve
name	Brookside Wetlands Natural Area
natural	wetland
operator	City of Portland
operator:type	public
operator:wikidata	Q6106
type	multipolygon
wetland	wet_meadow

Members

▼ 2 members

- Way 134590350 as outer
- Way 147467285 as inner

[Download XML](#)

[« Version #1 - View History](#) · [Version #3 »](#)

Brookside Wetlands Natural Area,
"rules": null}
[[leisure,nature_reserve],
[natural,wetland]]
_index_level_0_ 4828

Made with Fused | © Mapbox © OpenStreetMap Improve this map

Search

water



Oregon

Linco

New

Wash

OREGON

Pacific

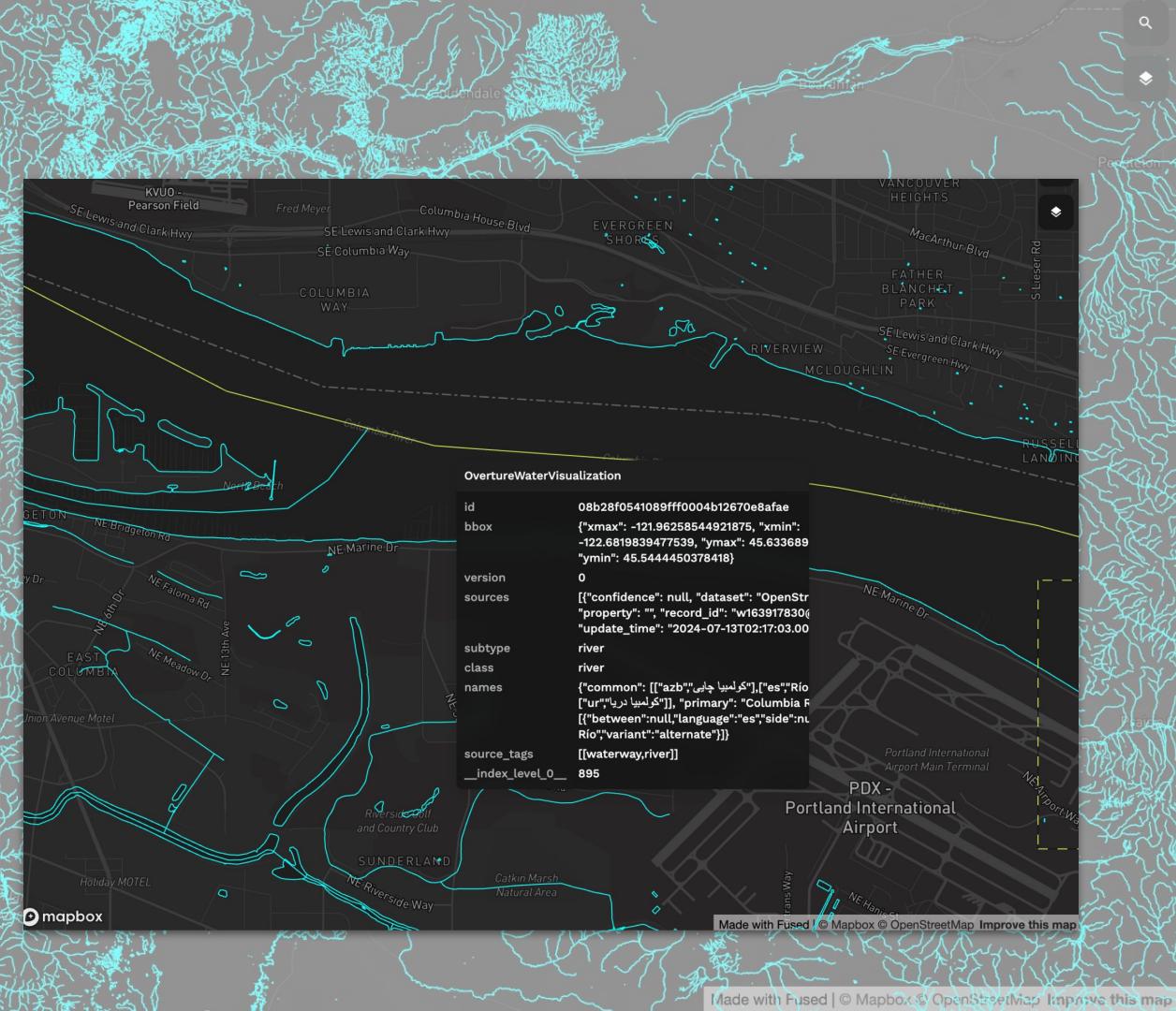




Sel

water

subtype	river
class	river
Names - primary	Columbia River
- common	azb = كولومبيا چاين es = Río Columbia ur = کولمبیا دریا"
- rules	[{language=es variant=alternate value=Gran Río }]
source_tags	waterway=river
sources	dataset=OpenStreetMap record_id=w163917830@14 update_time=2024-07-13



water

subtype	river
class	river
Names	
- primary	Columbia River
- common	azb = كولومبيا چاپي es = Río Columbia ur = "کولمبيا دريا"
- rules	[{language=es variant=alternate value=Gran Rio }]
source_tags	waterway=river
sources	dataset=OpenStreetMap record_id=w163917830@14 update_time=2024-07-13



Way: Columbia River (163917830) | OpenStreetMap

Relation: Brookside Wetlands Natural Area (16754900) | OpenStreetMap

OpenStreetMap [Edit](#)

Way: Columbia River (163917830)

Version #16

Improve accuracy of Wahkeena Lake and Benson Lake boundaries, connected waterways, various paths around Benson State Park

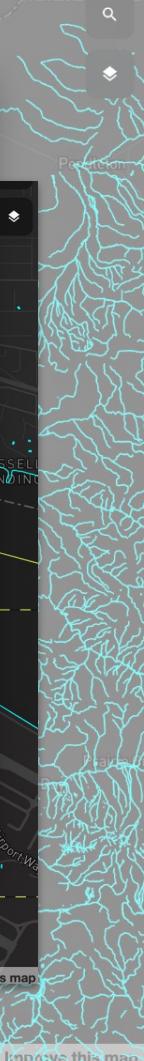
Edited about 1 month ago by Moonfire
Changeset #159795295

Tags

alt_name:es	Gran Rio
boat	yes
deep_draft	no
name	Columbia River
name:azb	كولومبيا ریو
name:es	Río Columbia
name:nez	xuyelp
name:ur	کولمبیا دریا
ship	yes
source:deep_draft	National Transportation Atlas Database 2011
waterway	river

Part of

Made with Fused © Mapbox © OpenStreetMap Improve this map



Accessing the Data?

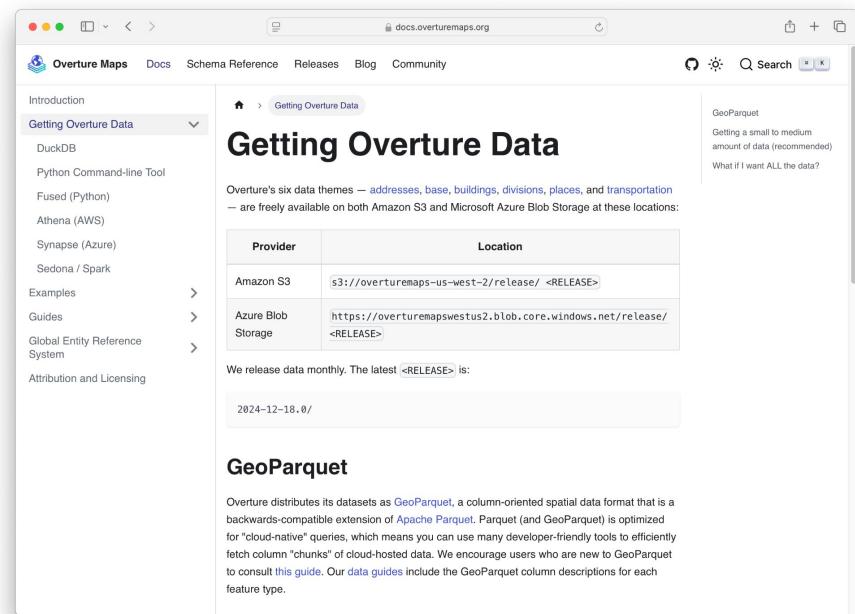
Overture data is published as GeoParquet on both S3 and Azure.

docs.overturemaps.org/release/latest/

The data is also available on source.coop, wherobots, and more.

Visit docs.overturemaps.org/getting-data to learn more about accessing the data.

Additionally, planetiler supports parquet input.
Build a global PMTiles archive from Overture Data:
docs.overturemaps.org/examples/overture-tiles/



The screenshot shows a web browser displaying the docs.overturemaps.org website. The page title is "Getting Overture Data". The left sidebar has a "Getting Overture Data" dropdown menu expanded, showing options like DuckDB, Python Command-line Tool, Fused (Python), Athena (AWS), Synapse (Azure), Sedona / Spark, Examples, Guides, Global Entity Reference System, and Attribution and Licensing. The main content area has a heading "Getting Overture Data" and a paragraph explaining that Overture's six data themes (addresses, base, buildings, divisions, places, and transportation) are freely available on both Amazon S3 and Microsoft Azure Blob Storage. It provides links for "Amazon S3" and "Azure Blob Storage". Below this, a note says "We release data monthly. The latest <RELEASE> is: 2024-12-18.0/". On the right side, there is a "GeoParquet" section with a detailed description of what it is and how it is used.

Getting Overture Data

Overture's six data themes — [addresses](#), [base](#), [buildings](#), [divisions](#), [places](#), and [transportation](#) — are freely available on both Amazon S3 and Microsoft Azure Blob Storage at these locations:

Provider	Location
Amazon S3	s3://overturemaps-us-west-2/release/ <RELEASE>
Azure Blob Storage	<a href="https://overturemapswestus2.blob.core.windows.net/release/<RELEASE>">https://overturemapswestus2.blob.core.windows.net/release/<RELEASE>

We release data monthly. The latest <RELEASE> is: 2024-12-18.0/

GeoParquet

Overture distributes its datasets as [GeoParquet](#), a column-oriented spatial data format that is a backwards-compatible extension of [Apache Parquet](#). Parquet (and GeoParquet) is optimized for “cloud-native” queries, which means you can use many developer-friendly tools to efficiently fetch column “chunks” of cloud-hosted data. We encourage users who are new to GeoParquet to consult [this guide](#). Our [data guides](#) include the GeoParquet column descriptions for each feature type.

OverPass Turbo (OSM)

```
/*
This has been generated by the
overpass-turbo wizard.
*/
```

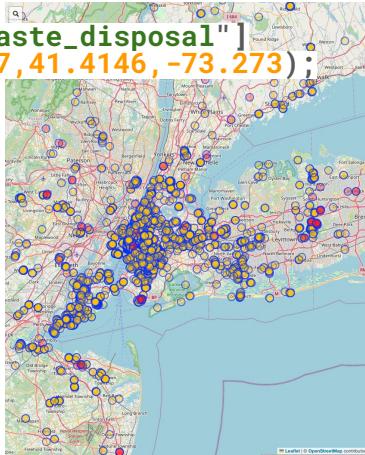
```
[out:json][timeout:25];
// gather results
```

```
(

nwr["amenity"="waste_basket"]
(40.122,-74.377,41.4146,-73.273);

nwr["amenity"="waste_disposal"]
(40.122,-74.377,41.4146,-73.273);

);
// print results
out geom;
```



DuckDB (Overture)

SELECT

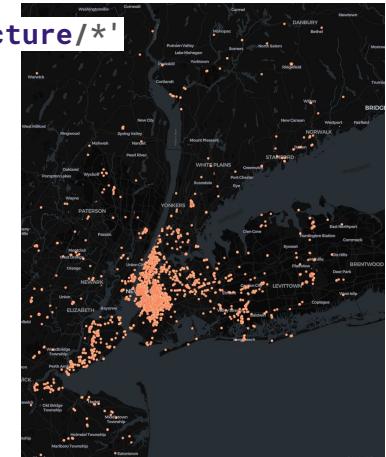
```
id,
subtype,
class,
source_tags,
sources[1].record_id AS osm_id,
geometry
```

FROM

```
read_parquet('s3://overturemaps-us-west-2/
              release/2025-01-22.0/
              theme=base/type=infrastructure/*'
            )
```

WHERE

```
subtype = 'waste_management'
AND bbox.xmin > -74.377
AND bbox.ymin > 40.122
AND bbox.xmax < -73.273
AND bbox.ymax < 41.414;
```



Are we missing a specific type of feature?

The base theme is not *all* of OSM...

... and it never will be.

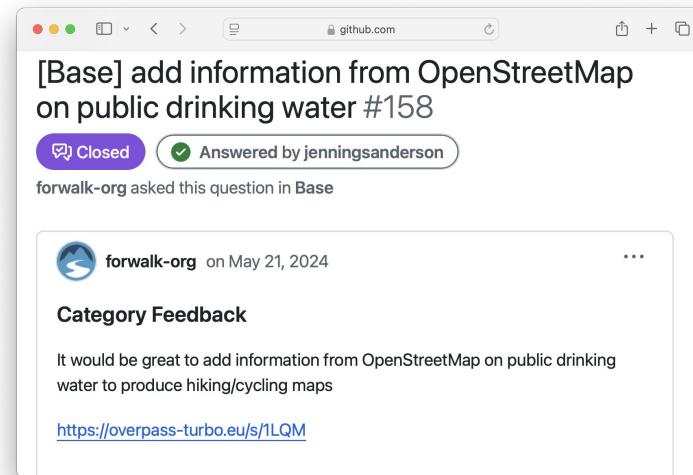
OSM is a map of *anything and everything*
made by a *community of mappers*

Overture distributes open map data (including OSM) that has been optimized for developers and map makers.

Base theme requires:

1. Explicit feature selection from OSM.
2. Well-defined mapping of OSM tags to Overture attributes.

We want your input!



docs.overturemaps.org

Overture Maps Docs Schema Reference Releases Blog Community

Introduction

Getting Overture Data

Examples

Guides

Addresses

Base

Buildings

Divisions

Places

Transportation

Global Entity Reference System

Attribution and Licensing

- **building**: The most basic form of a building feature. The geometry is expected to be the most outer footprint—roofprint if traced from satellite/aerial imagery—of a building. Buildings have a boolean attribute `has_parts` that describe whether there are any associated building parts.
- **building_part**: A single part of a building. Building parts may share the same properties as buildings. A building part is associated with a parent building via a `building_id`.

Overview

Data sources

Understanding the Parquet Files

Schema reference

Data sources

OpenStreetMap | Esri Community Maps | Microsoft ML Buildings | Google Open Buildings

Above is a visualization of San Diego, USA where multiple datasets are conflated. Notice how Esri and OSM buildings appear in big blocks while the Google and Microsoft buildings appear to mix together. This is a product of our conflation process that prioritizes community contributed data first and then "fills in" the rest of the map with the best ML data available.

Thank You!

Get in touch:

#overture channel on OSMUS slack

github.com/overturemaps

docs.overturemaps.org