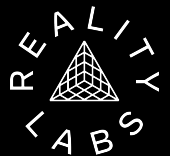


Daylight Earth Tables

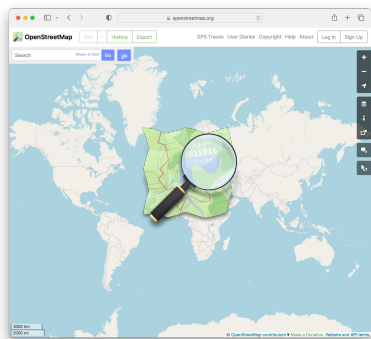
Jennings Anderson, Jonah Adkins, Jacob Wasserman



Maps at Meta

19/October/2022

∞ Meta



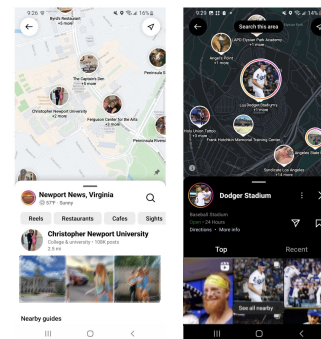
OpenStreetMap



Daylight Map Distribution



Earth Table



Maps in Meta Products

- Instagram
- Facebook Marketplace
- And more...

Overview

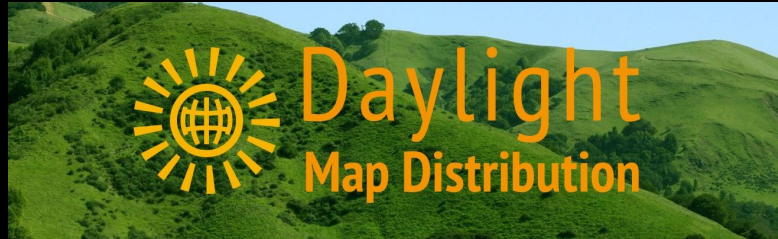
What is the Daylight Map Distribution?

What is the Earth Table?

Introducing: Daylight Earth Table

Working with the Daylight Earth Table

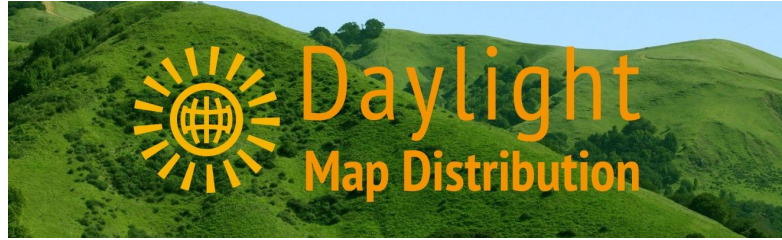
Daylight



A monthly distribution of OpenStreetMap that undergoes a series of quality control and vandalism checks to ensure a degree of map quality and integrity.



Timeline of Daylight OpenStreetMap Distribution



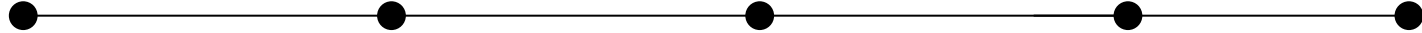
< 2019

2020

2021

2022

Today



State of the Map
2018 & 2019

March 2020

April 2021

January 2022

October 2022

Mobius Logical
Changesets
(LoChas)

Announcing
Daylight v0.1

Daylight v1.0
released

Daylight published
on AWS Registry of
Open Data as PBF
and *Analysis-Ready*
parquet files

Daylight v1.18

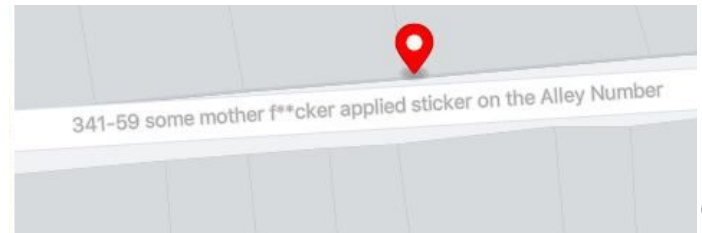
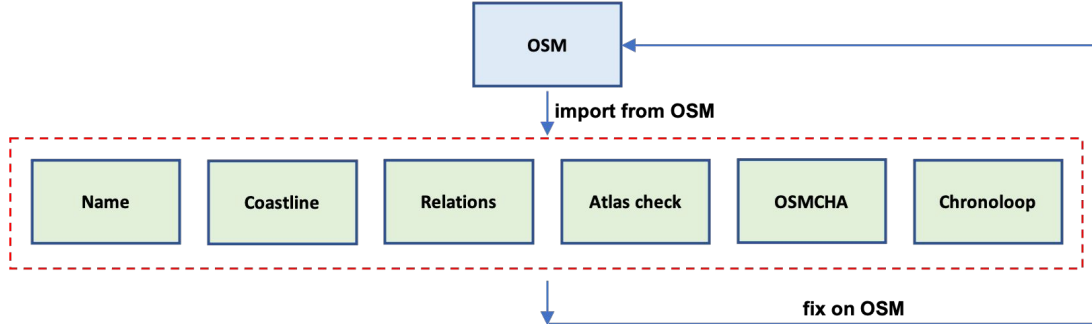
<https://daylightmap.org>

Daylight Process: Find-Fix-Import Loop

Find: Discover errors or other issues anywhere on the map

Fix: Submit fixes on live OSM, not in an internal database

Import: Apply fixes from OSM into the Daylight map



Daylight v1.18

525 M

Buildings

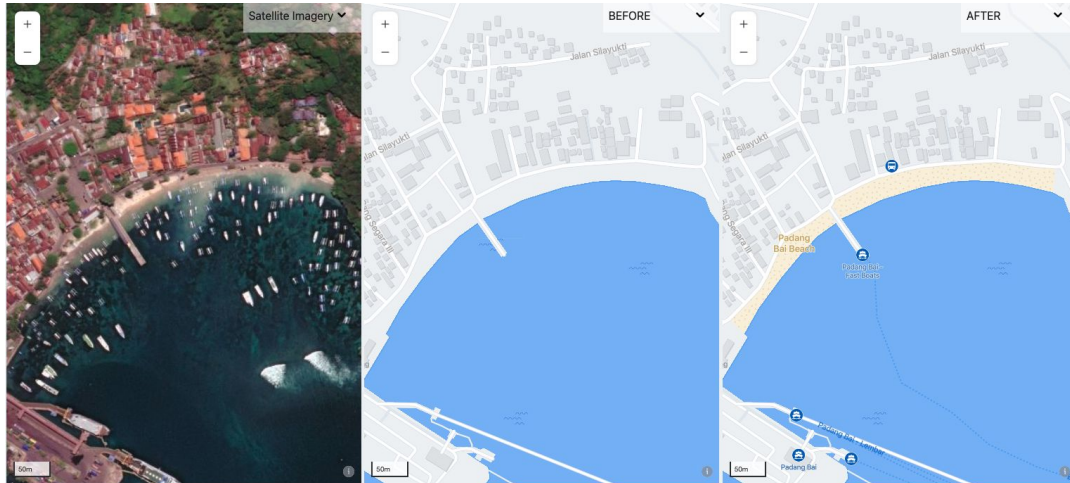
78 M

Kilometers of roads / paths

100%

OpenStreetMap Data

Think of Daylight as a snapshot of OSM where each feature might not be from the same snapshot.



Example: Repaired beach relation for Padang Bai Beach in Bali, Indonesia

Daylight v1.18

Changelog and summaries of fixes made to OSM available on daylightmap.org

⏏🔍

🔒 📄 🔗 daylightmap.org 🔄

v1.18 Summary

Most Notable 4

- 122 features were fixed from relation check.
- 172 features were fixed from coastline check.
- 4648 features were fixed from various atlas checks.
- 207 features were fixed from nlp check.

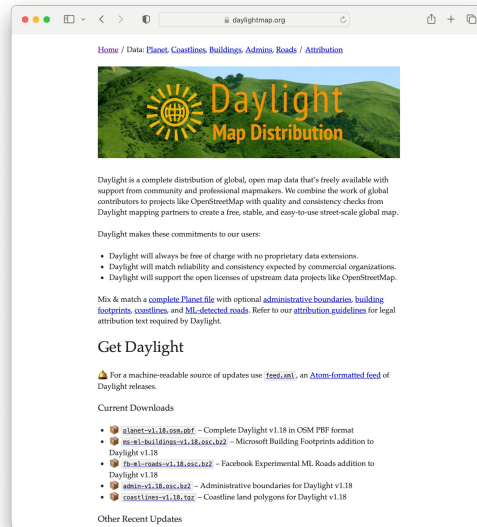
Fixed broken relation for [Kapchagay Reservoir](#), in Almaty Oblysy, Kazakhstan

Fixed a broken relation for [Área Natural Protegida La Auxiliadora](#) in [Departamento de Sonsonate](#), El Salvador

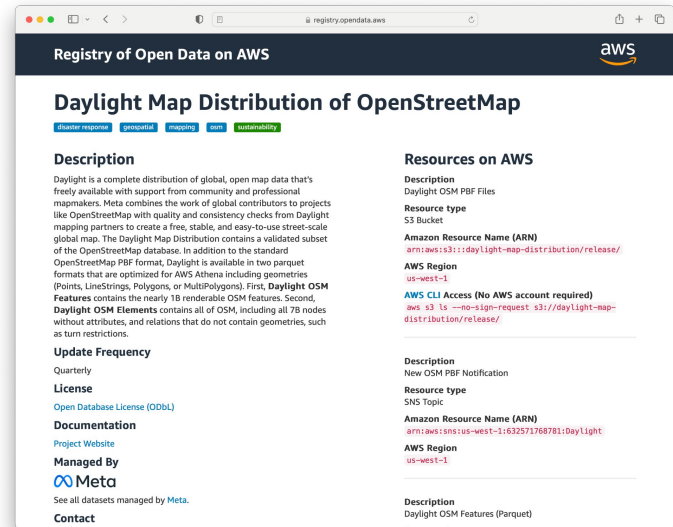
Fixed the [Área de la Biosfera de Transición del Bosque Mbaracayu](#), in Región Oriental, Paraguay nature reserve

Where can I find Daylight?

- OSM PBF Format
- Analysis-Ready Cloud-Optimized Parquet Files

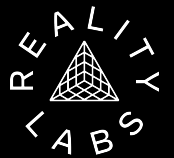


daylightmap.org



registry.opendata.aws/daylight-osm/

Earth Table



Where did the Earth Table come from?

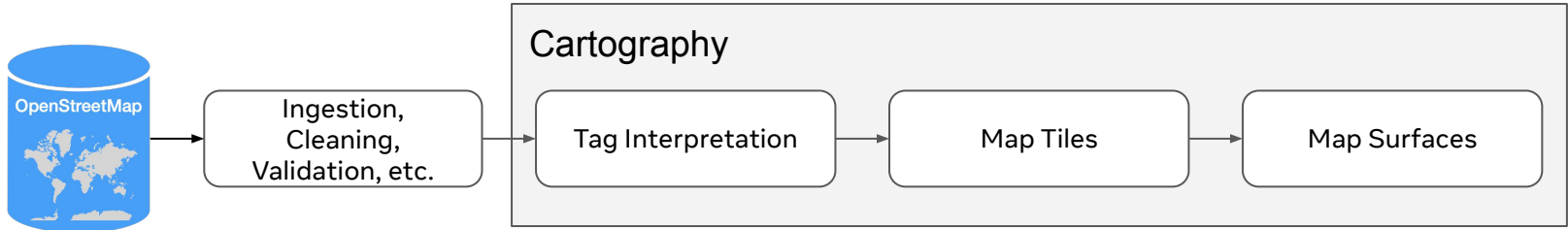
- Growing internal use cases to query the earth - “I need all the parks in the world to do x”
- Using basemaps is a cross-functional tool for projects across the org.
- Needs to be simple - tagging complexities of OSM are a high barrier to entry
- Needs to be in a single table, with easy to understand schema & hierarchy



Before:

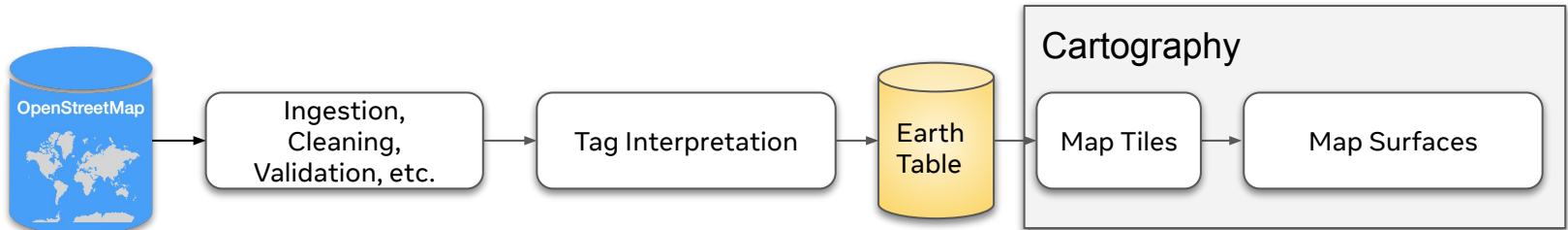
Cartography team did all of the OSM tag interpretation and created map tiles.

Needed to look inside tiles to find well-formatted, translated map data

**After:**

The earth table is created by interpreting OSM tags into a simplified 3-level ontological schema.

Anyone can access this table, especially cartography



It doesn't solve everything...

OpenStreetMap Edit History Export GPS Traces User Diaries Copyright

Search Where is this? Go

Way: 98224505

Version #4
set to footpath
Edited about 6 years ago by mjfoster83
Changeset #41836726

Tags

access	no
highway	footway
surface	unpaved

Nodes

5 nodes

- 1136317564
- 1136317632
- 1136317375
- 1136317525
- 1136317564

[Download XML - View History](#)

Way: 1058376267

Version #1

adding or improving sport objects #maproulette

Edited 5 months ago by conradoos
Changeset #120717647

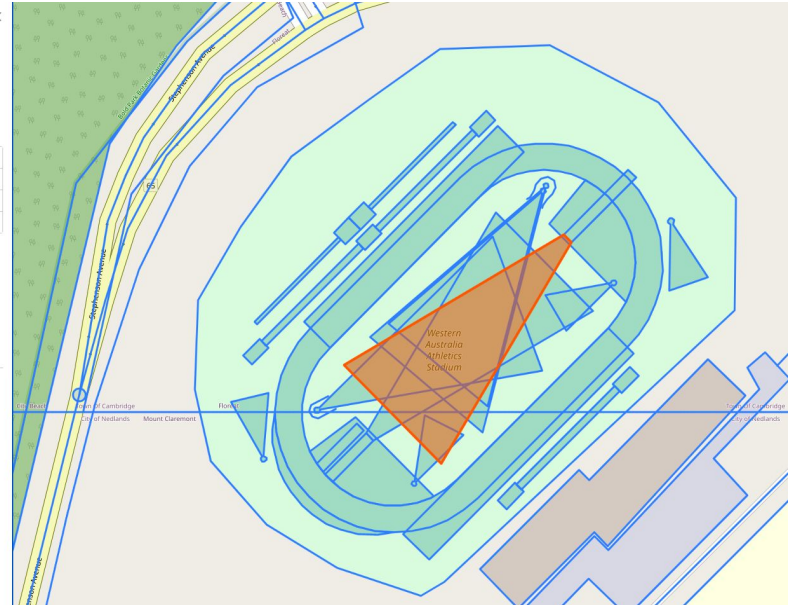
Tags

athletics	javelin_throw
leisure	pitch
sport	athletics
surface	grass

Nodes

- 5 nodes
- 9725338222 (part of way 1058376268)
 - 9725338221
 - 9725338220
 - 9725338223 (part of way 1058376268)
 - 9725338222 (part of way 1058376268)

[Download XML - View History](#)



Healthcare			
amenity	baby_hatch		A place where a baby can be, out of necessity, anonymously left to be safely cared for and perhaps adopted.
amenity	clinic		A medium-sized medical facility or health centre.
amenity	dentist		A dentist practice / surgery.
amenity	doctors		A doctor's practice / surgery.
amenity	hospital		A hospital providing in-patient medical treatment. Often used in conjunction with <code>emergency=*</code> to note whether the medical centre has emergency facilities (A&E (brit.) or ER (am.))
amenity	nursing_home		Discouraged tag for a home for disabled or elderly persons who need permanent care. Use <code>amenity=social_facility + social_facility=nursing_home</code> now.

Value	
<code>healthcare=alternative</code>	
<code>healthcare=audiologist</code>	
<code>healthcare=birthing_centre</code>	
<code>healthcare=blood_bank</code>	
<code>healthcare=blood_donation</code>	
<code>healthcare=centre</code>	
<code>healthcare=clinic</code>	
<code>healthcare=community_health_worker</code>	
<code>healthcare=counselling</code>	
<code>healthcare=dentist</code>	
<code>healthcare=dialysis</code>	
<code>healthcare=doctor</code>	
<code>healthcare=hospice</code>	
<code>healthcare=hospital</code>	
<code>healthcare=laboratory</code>	
<code>healthcare=midwife</code>	
<code>healthcare=nurse</code>	

Way: Riverside Regional Medical Center (298161573)

Version #12

Roads and Sidewalks

Edited almost 2 years ago by [pardthemonster](#)
Changeset #96888275

Tags

<code>addr:city</code>	Newport News
<code>addr:housenumber</code>	500
<code>addr:postcode</code>	23601
<code>addr:state</code>	VA
<code>addr:street</code>	J. Clyde Morris Boulevard
<code>amenity</code>	hospital
<code>beds</code>	450
<code>emergency</code>	yes
<code>healthcare</code>	hospital
<code>name</code>	Riverside Regional Medical Center
<code>opening_hours</code>	24/7
<code>operator</code>	Riverside
<code>phone</code>	+1 757 594 2000
<code>website</code>	https://www.riversideonline.com/trmc

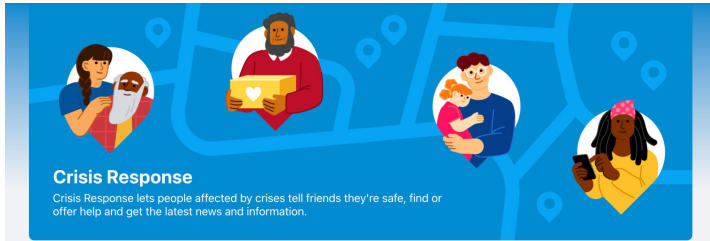


theme	class	subclass	metadata	wkt
poi	medical	hospital	<code>{"amenity":"hospital","is_area":true,"quadkey":"032010233003132"}</code>	POINT (-76.4
landuse	medical	hospital	<code>{"quadkey":"032010233003132","surface_area_sq_m":267697.27}</code>	POLYGON (

building	525,240,408
building_detail	2,232,081
infrastructure	11,226,443
land	40,831,400
landuse	39,397,340
placename	3,430,404
poi	38,777,950
road	216,890,017
transit	6,716,006
water	36,401,694

Cartographic Data Improvements:

- land theme includes processed global coastlines
- building height information is validated and normalized
- placename theme includes simplified classification -> all places grouped into 3 classes: urban, settlement, local
- lengths and areas calculated as attribute columns
- Bing Tile Quadkey applied to each feature
- Lots of boolean evaluations for easy filtering:
 - “Is_indoor”, “is_intermittent”, “is_bridge”, “is_area”, etc
- Buildings include “landuse” class they are within
- Building_detail includes building id they are part of



Recent crises

Happening around the world

The Flooding Across Eastern El Salvador
Media sources indicate that heavy rainfall from Tropical Storm Julia...

6 help offers
6 donations

[Learn more](#) [Donate](#)

The Flooding in Huehuetenango Department,...
Updated media sources report that at least 28 people have been killed in Guatemala, Honduras, and El...

1 help offer

[Learn more](#) [Follow](#)

The Building Collapse in Farash Khana, Delhi, India
Updated sources reported that an unspecified house has partially collapsed in the Farash Khana area, reportedly near Vainiki Mandir Temple. At least three people have...

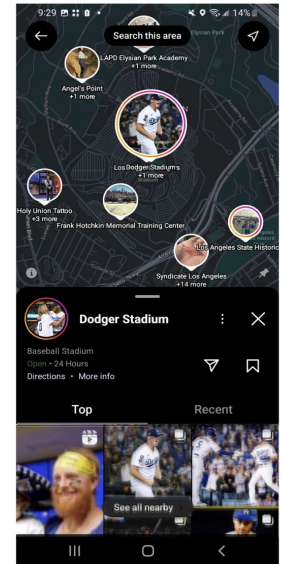
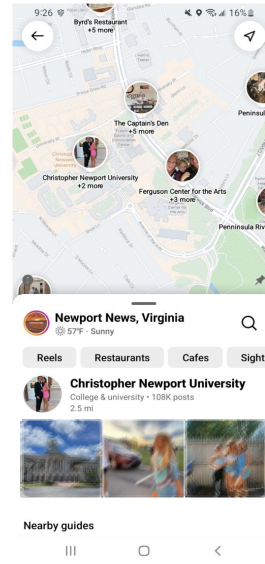
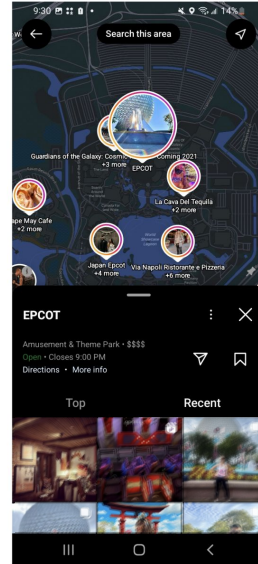
[Learn more](#) [Follow](#)

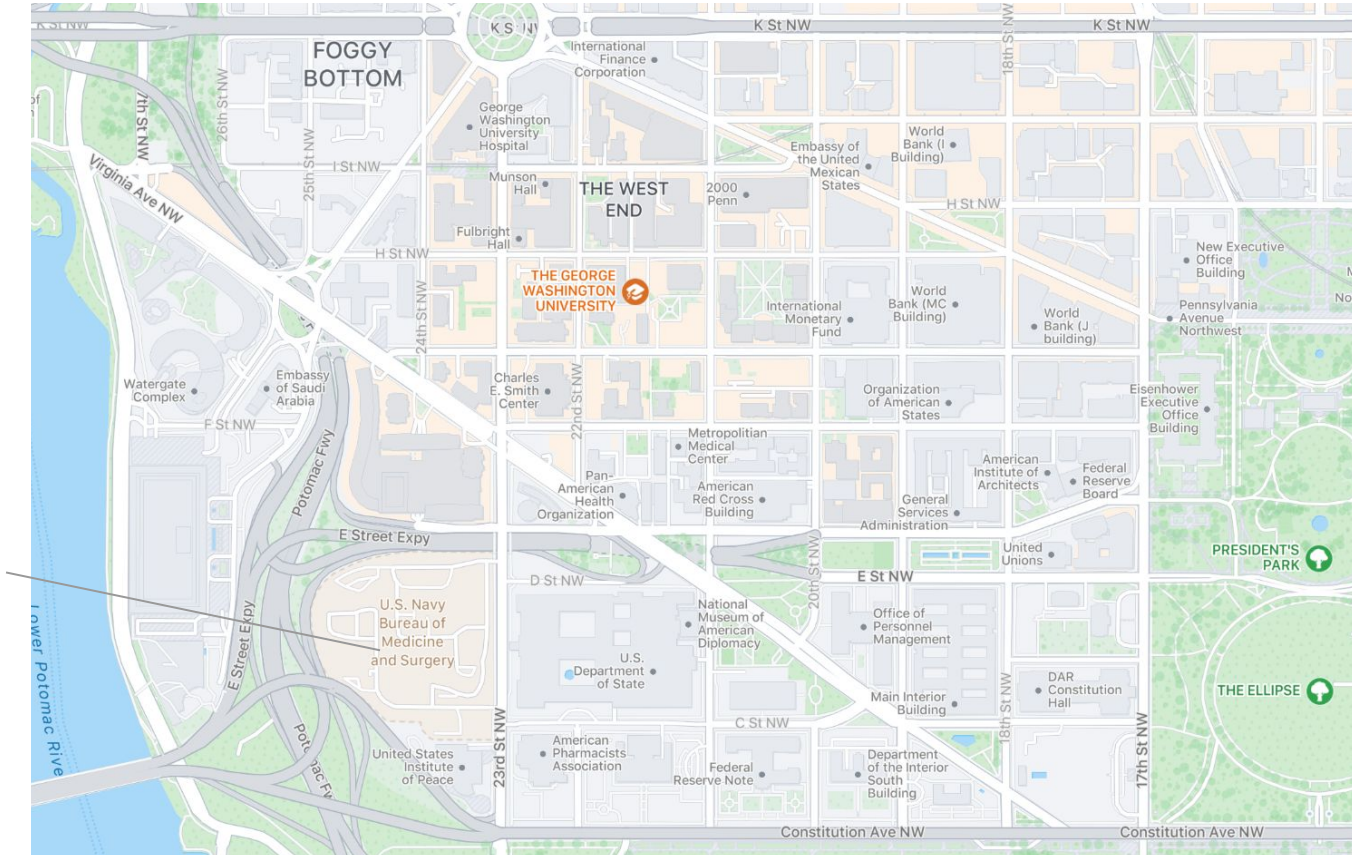
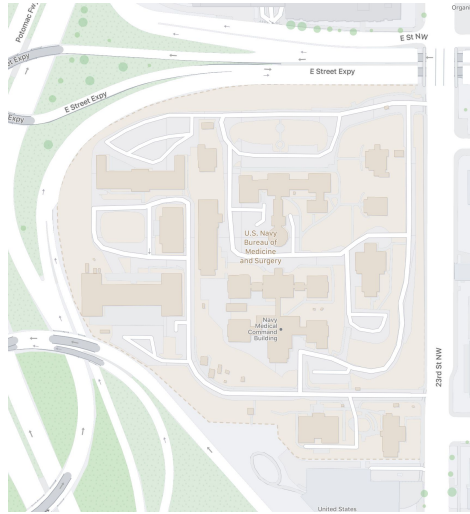
floodin across

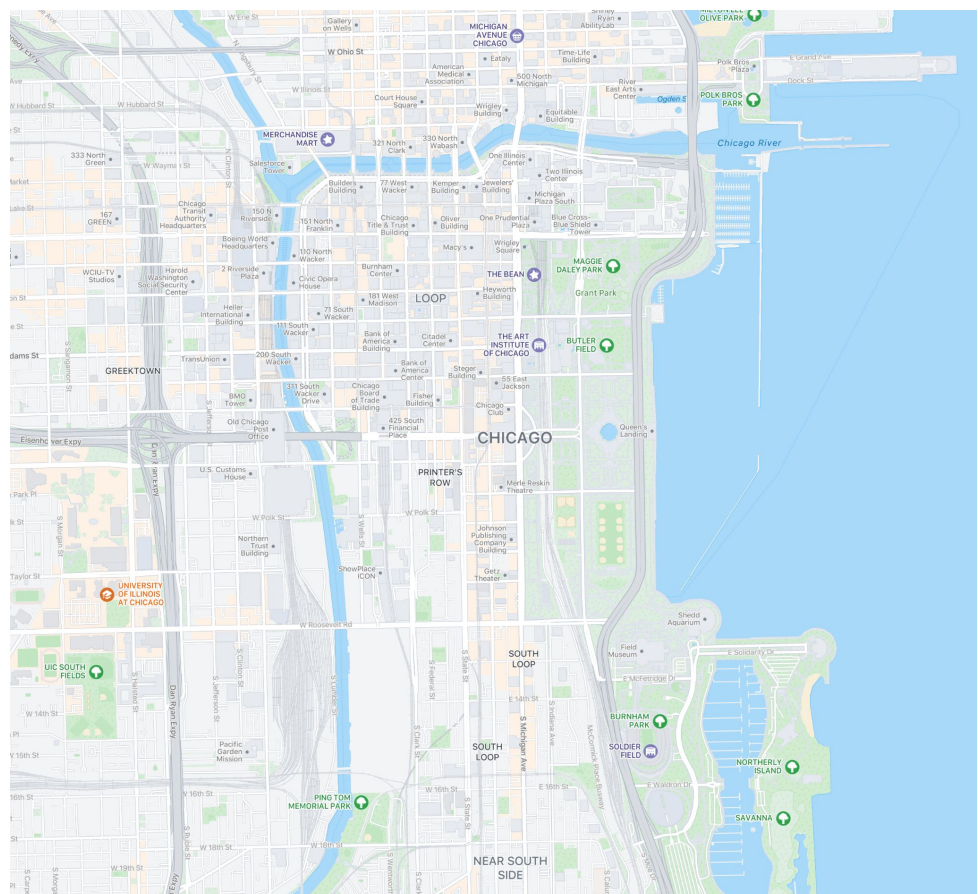
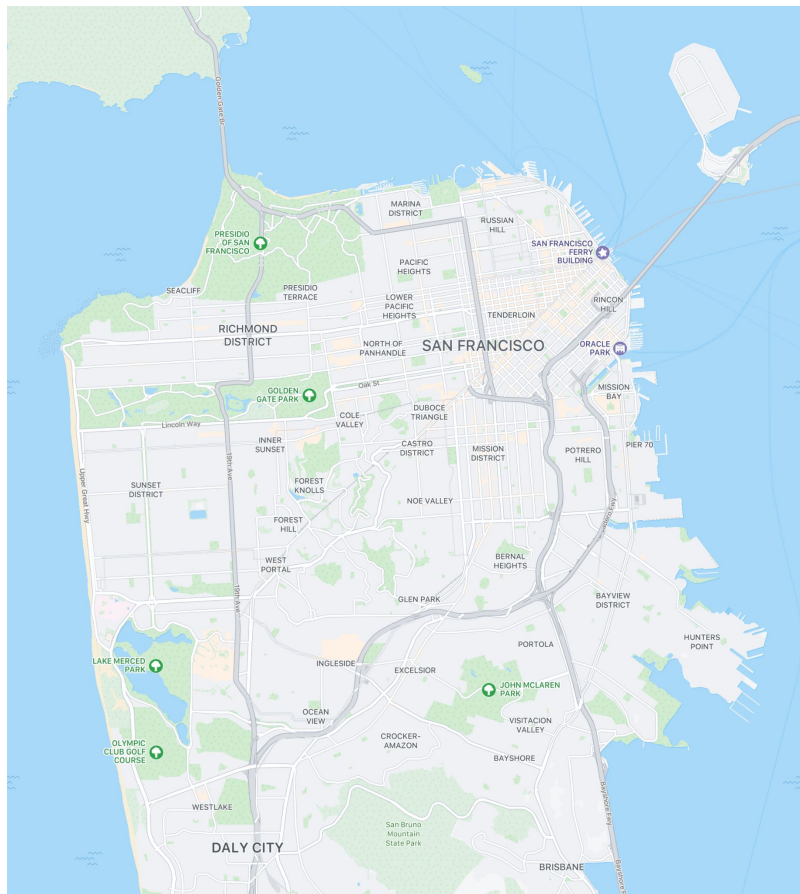
15

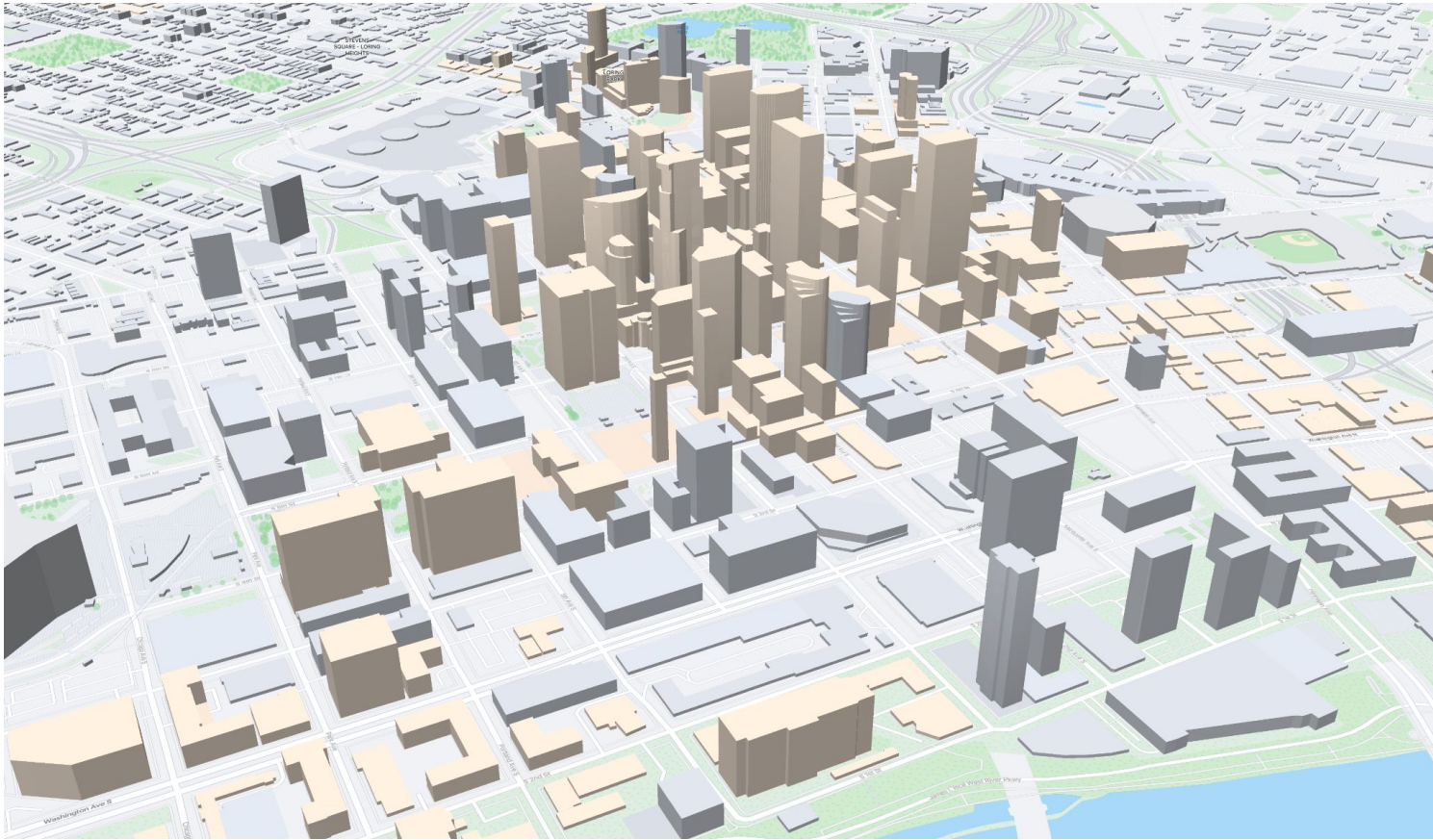
[Lea](#)

83%	Earth Table Data Source
16%	Internal Data Sources that are “Earthified”
1%	Curated Cartographic Map Features



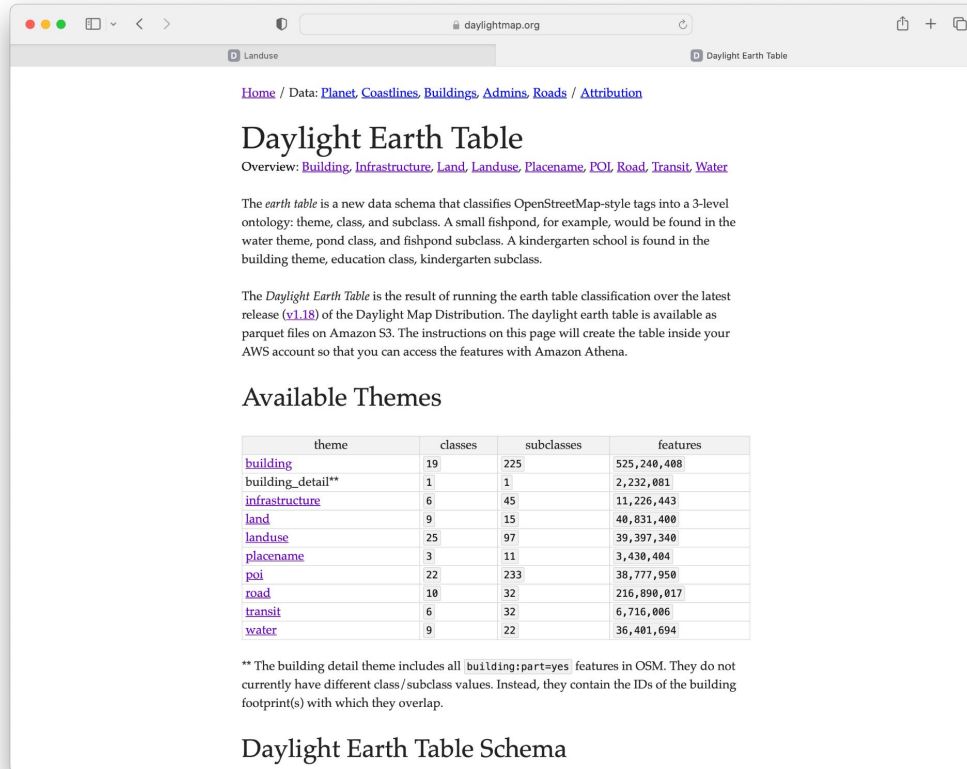






Accessing Daylight Earth Tables

s3://daylight-openstreetmap/earth



Home / Data: [Planet](#), [Coastlines](#), [Buildings](#), [Admins](#), [Roads](#) / [Attribution](#)

Daylight Earth Table

Overview: [Building](#), [Infrastructure](#), [Land](#), [Landuse](#), [Placename](#), [POI](#), [Road](#), [Transit](#), [Water](#)

The *earth table* is a new data schema that classifies OpenStreetMap-style tags into a 3-level ontology: theme, class, and subclass. A small fishpond, for example, would be found in the water theme, pond class, and fishpond subclass. A kindergarten school is found in the building theme, education class, kindergarten subclass.

The *Daylight Earth Table* is the result of running the earth table classification over the latest release ([v1.18](#)) of the Daylight Map Distribution. The daylight earth table is available as parquet files on Amazon S3. The instructions on this page will create the table inside your AWS account so that you can access the features with Amazon Athena.

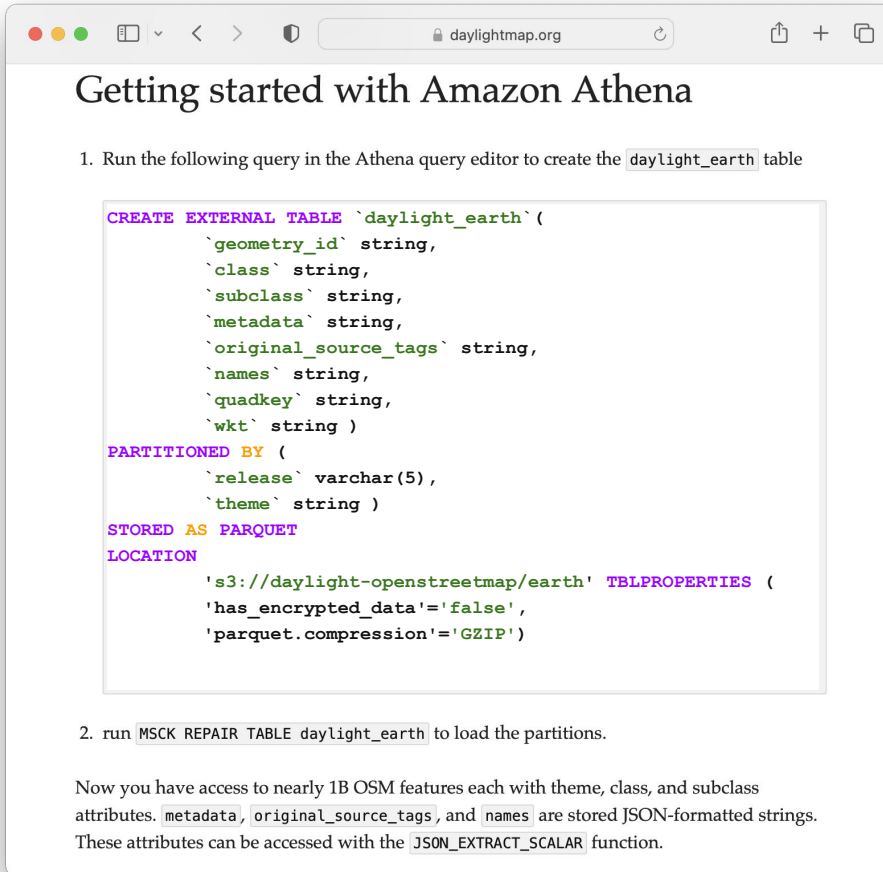
Available Themes

theme	classes	subclasses	features
building	19	225	525,240,408
building_detail**	1	1	2,232,081
infrastructure	6	45	11,226,443
land	9	15	40,831,400
landuse	25	97	39,397,340
placename	3	11	3,430,404
poi	22	233	38,777,950
road	10	32	216,890,017
transit	6	32	6,716,006
water	9	22	36,401,694

** The building detail theme includes all `building:part=yes` features in OSM. They do not currently have different class/subclass values. Instead, they contain the IDs of the building footprint(s) with which they overlap.

Daylight Earth Table Schema

Using Amazon Athena



Getting started with Amazon Athena

1. Run the following query in the Athena query editor to create the `daylight_earth` table

```
CREATE EXTERNAL TABLE `daylight_earth` (  
  `geometry_id` string,  
  `class` string,  
  `subclass` string,  
  `metadata` string,  
  `original_source_tags` string,  
  `names` string,  
  `quadkey` string,  
  `wkt` string )  
PARTITIONED BY (  
  `release` varchar(5),  
  `theme` string )  
STORED AS PARQUET  
LOCATION  
  's3://daylight-openstreetmap/earth' TBLPROPERTIES (  
    'has_encrypted_data'='false',  
    'parquet.compression'='GZIP')
```

2. run `MSCK REPAIR TABLE daylight_earth` to load the partitions.

Now you have access to nearly 1B OSM features each with theme, class, and subclass attributes. `metadata`, `original_source_tags`, and `names` are stored JSON-formatted strings. These attributes can be accessed with the `JSON_EXTRACT_SCALAR` function.

The screenshot shows the Amazon Athena Query Editor interface. The browser address bar indicates the URL is `us-west-2.console.aws.amazon.com`. The page title is "Amazon Athena > Query editor".

The interface includes a top navigation bar with "Services", a search bar, and a dropdown menu for "Oregon" and "Jennings Anderson". Below this, the "Amazon Athena" logo and "S3" are visible.

The main content area is divided into several sections:

- Editor:** Contains the SQL query for "Query 1".
- Data:** A sidebar on the left with sections for "Data source" (set to "AwsDataCatalog"), "Database" (set to "daylight"), and "Tables and views" (showing 8 tables and 0 views).
- Query Editor:** Displays the following SQL code:


```

1 CREATE EXTERNAL TABLE `daylight_earth`(
2   `geometry_id` string,
3   `class` string,
4   `subClass` string,
5   `metadata` string,
6   `original_source_tags` string,
7   `names` string,
8   `quadkey` string,
9   `wkt` string)
10 PARTITIONED BY (
11   `release` varchar(5),
12   `theme` string)
13 STORED AS PARQUET
14 LOCATION
15   's3://daylight-openstreetmap/earth'
16 TBLPROPERTIES (
17   'has_encrypted_data'='false',
18   'parquet.compression'='GZIP')
            
```
- Execution Controls:** Below the query, there are buttons for "Run again", "Explain", "Cancel", "Save", "Clear", and "Create".
- Query Results:** Shows a status of "Completed" with the following metrics: "Time in queue: 115 ms", "Run time: 440 ms", and "Data scanned: -". Below this, it states "Query successful."

The footer of the page contains a "Feedback" link, a note about language selection, and copyright information: "© 2022, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences".

The screenshot shows the AWS Athena console interface. The browser address bar indicates the URL is `us-west-2.console.aws.amazon.com`. The console header shows the user is logged in as Jennings Anderson in the Oregon region. The main content area displays a SQL query for 'Query 5' with the following code:

```

1 select theme,
2   count(geometry_id)
3   from daylight_earth
4   group by theme
5   order by count(geometry_id) DESC
    
```

Below the query editor, there are buttons for 'Run again', 'Explain', 'Cancel', 'Save', 'Clear', and 'Create'. The 'Query results' tab is active, showing a status of 'Completed' with a 'Time in queue: 108 ms', 'Run time: 4.453 sec', and 'Data scanned: 5.04 GB'. There are 'Copy' and 'Download results' buttons. The results are displayed in a table with 10 columns and 7 rows of data:

#	theme	_col1
1	building	525240408
2	road	216890017
3	land	40831400
4	landuse	39397340
5	poi	38777950
6	water	36401694
7	infrastructure	11226443

The footer of the console includes a 'Feedback' link, a language selection notice, and copyright information for Amazon Web Services, Inc. (© 2022), along with links for 'Privacy', 'Terms', and 'Cookie preferences'.

Query 25 × | Query 2 × | ✔ Query 3 × | Query 4 × | Query 5 × | Query 6 × | Query 7 × | Query 8 × | Query 9 × | Query 10 ×

+ | ▾

```

1 SELECT *
2 FROM daylight_earth
3 WHERE theme = 'building'
4 AND JSON_EXTRACT_SCALAR(metadata, '$.landuse') = 'developed:industrial'
5 LIMIT 100
    
```

SQL Ln 5, Col 10



Run again

 Explain
Cancel
Save ▾
Clear
Create ▾

Query results
Query stats

✔ Completed
 Time in queue: 168 ms Run time: 6.46 sec Data scanned: 15.19 GB

Results (100+)

Copy

Download results

<
1
...
>

# ▾	geometry_id ▾	class ▾	subclass ▾	metadata ▾	original_source_tags ▾
1	w496485510@1	general	building	{ "building": "yes", "landuse": "developed:industrial", "quadkey": "031333211101132", "surface_area_sq_m": 268.9 }	{ "building": "yes", "source": "Eusko Jaurlaritza-Gobierno Vasco. GeoEuskadi." }

Earth Table Metadata

Normalize attributes into metadata such as height

Append additional information from other themes for overlapping features

`earth_table`

Way: 519759719

Version #1

Various changes and additions

Edited almost 5 years ago by goadelic
Changeset #51557981

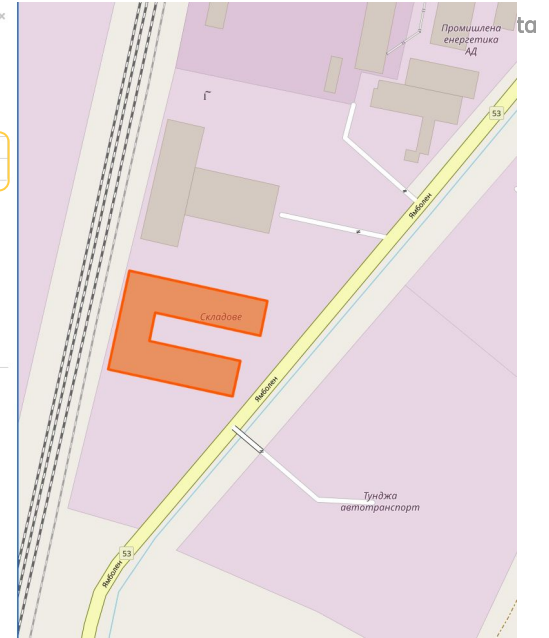
Tags

building	warehouse
building:levels	6

Nodes

▼ 9 nodes
5067554746
5067554747
5067554748
5067554749
5067554750
5067554751
5067554752
5067554753
5067554746

[Download XML](#) - [View History](#)



source: **osm**
theme: **building**
class: **industrial**
subclass: **warehouse**

height: **20.52 meters**
area: **4962 square meters**
landuse class: **developed,**
landuse subclass: **industrial**
wkt: POLYGON ((26.483607 ...))

Theme: landuse

Landuses from OSM.

Geometry Type	features
LineString	471,658
MultiLineString	2,882
MultiPolygon	97,018
Point	7,794
Polygon	38,817,988

Metadata keys

key name	description of values
landuse	value of landuse tag in OSM
layer	value of layer tag in OSM
level	value of level tag in OSM
length_m	Length of feature in meters (if a line)
natural_tag	value of natural tag in OSM
sport	value of sport tag in OSM
surface_area_sq_m	Area of feature in square meters (if a polygon)
surface	value of surface tag in OSM
wikidata	Wikidata ID (if present in OSM)

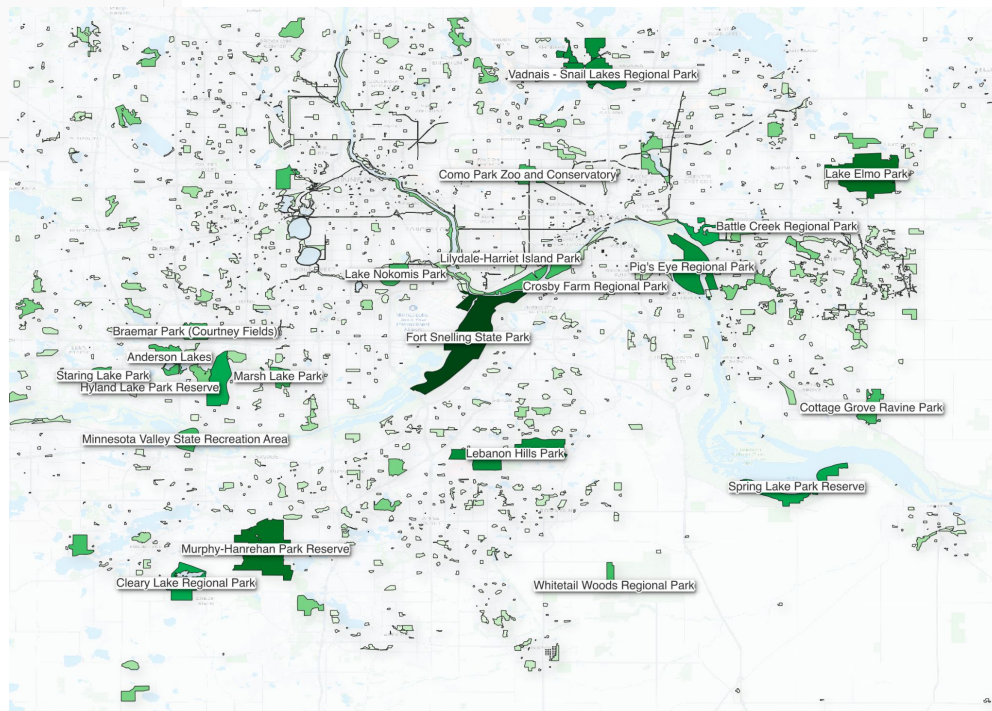
Feature counts per class and subclass in Daylight v1.18

Class	Subclass	Feature Count	Example Feature in OSM
agriculture	farmland	7,993,053	w953845420@1
agriculture	farmyard	1,124,138	w765544240@1
agriculture	meadow	362,980	w392437867@3
agriculture	animal_keeping	4,189	w762969249@2
airport	helipad	34,099	w433701531@1
airport	aerodrome	18,302	w564067476@8

```

1 SELECT
2   geometry_id,
3   class,
4   subclass,
5   JSON_EXTRACT_SCALAR(names, '$.local') as name,
6   CAST(JSON_EXTRACT_SCALAR(metadata, '$.surface_area_sq_m') AS double) as area,
7   wkt
8 FROM daylight_earth
9 WHERE release = 'v1.18'
10  AND theme = 'landuse'
11  AND class = 'park'
12  AND subclass <> 'grass'
13  AND quadkey like '021333011%'
    
```

geometry_id	class	subclass	name	area	wkt
w43377655@2	park	park	Snail Lake Marsh Park	170479.69	POLYGON ((-93.1
w522449122@5	park	park	Bunker Hill Park	62740.8	POLYGON ((-93.1
w216244611@2	park	park		22752.91	POLYGON ((-93.1
w47652678@7	park	park	East Phillips Park	29728.08	POLYGON ((-93.1
w336407468@1	park	park		1591.29	POLYGON ((-93.1
w35079781@1	park	park		19078.45	POLYGON ((-93.1
w329050291@5	park	park	Erickson Park	180152.37	POLYGON ((-93.1
w1085852524@1	park	park		200599.63	POLYGON ((-93.1
w314677340@2	park	park		44831.66	POLYGON ((-92.5
w912557268@1	park	park		7700.07	POLYGON ((-93.1
w635018766@1	park	park	Wargo Court	1806.61	POLYGON ((-93.1
w960891048@1	park	park		5114.3	POLYGON ((-93.1
w36657367@8	park	park	Cedar Manor Lake Park	60402.83	POLYGON ((-93.1
w893835529@4	park	park		115032.29	POLYGON ((-92.1
w894770103@3	park	park		35738.3	POLYGON ((-92.5
w1003253264@2	park	park	Dean Parkway	17493.68	POLYGON ((-93.1
w886329369@2	park	park	Edgerton Park	71072.74	POLYGON ((-93.1




```

1 SELECT
2   geometry_id,
3   class,
4   subclass,
5   JSON_EXTRACT_SCALAR(names, '$.local') as name,
6   wkt
7 FROM daylight_earth
8 WHERE theme = 'water'
9 AND class = 'river'
10 AND ST_CONTAINS(
11   ST_GEOMETRYFROMTEXT('POLYGON((-109.06409916585731 41.008769739828494,
12     ,-109.07238952929733 37.000860715568436, -109.06409916585731 41.00
13   )

```

SQL Ln 5, Col 51

Run again

Explain [↗](#)

Cancel

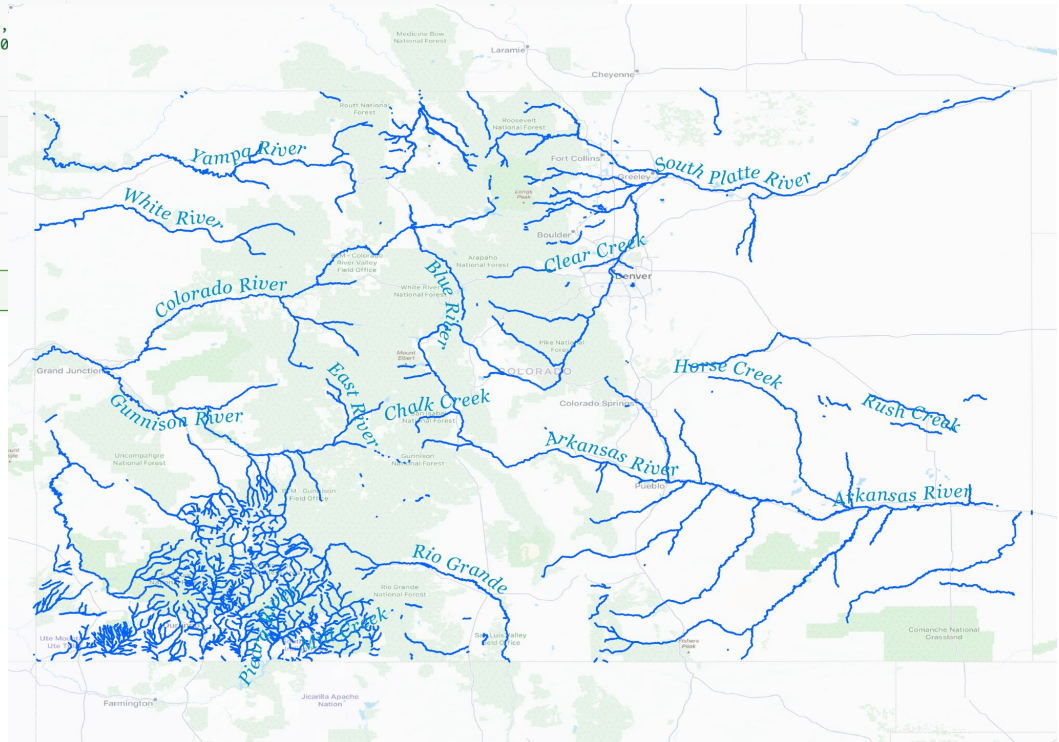
Save ▼

Clear

Create ▼

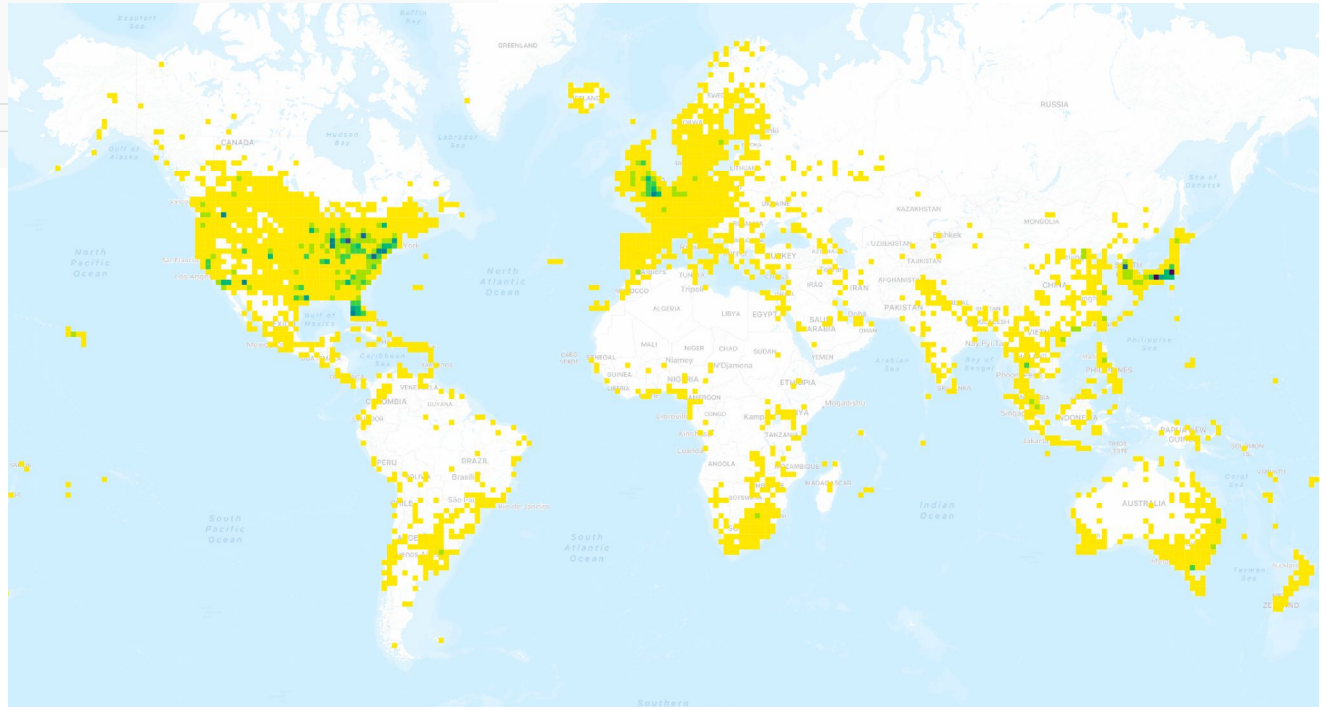
Query results

Query stats

 Completed


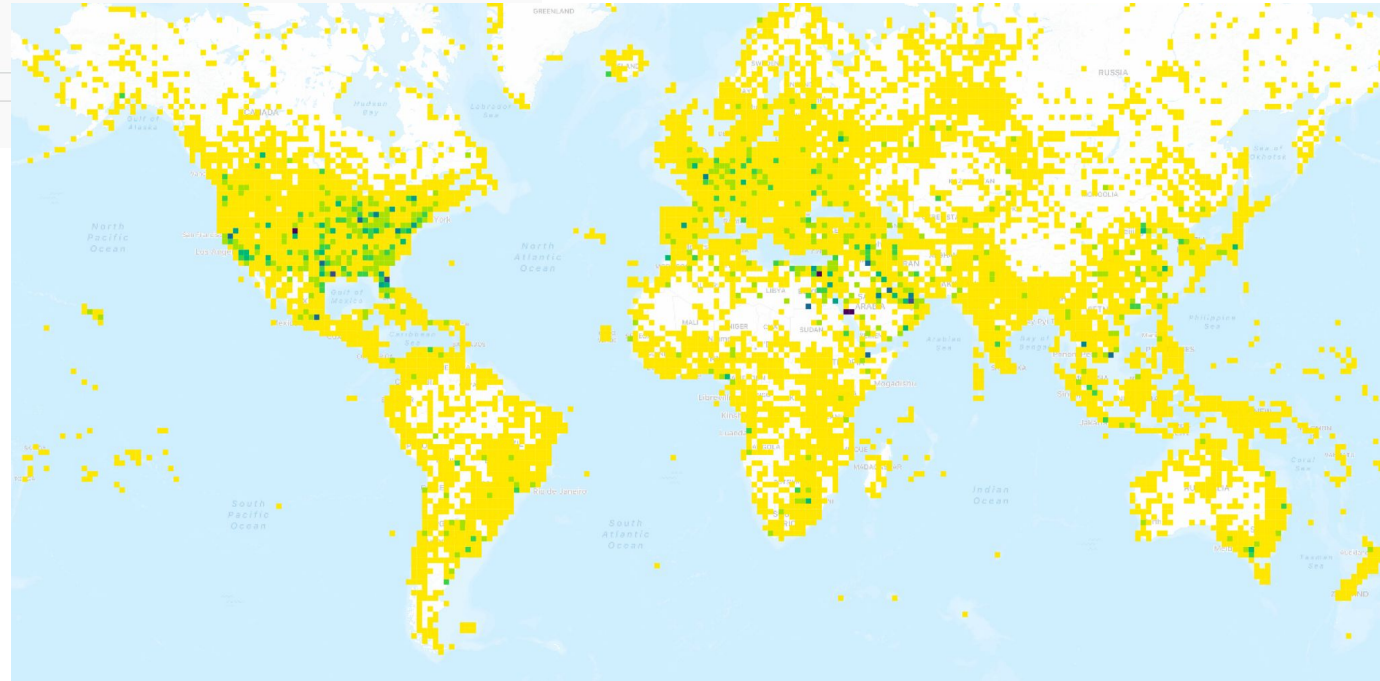
✔ Query 3

```
1 select BING_TILE_POLYGON(BING_TILE(substr(quadkey, 1, 8))),
2     sum(
3     cast(
4     JSON_EXTRACT_SCALAR(metadata, '$.surface_area_sq_m')
5     as double)
6     ) / 1000000 AS sq_km_golf
7 from daylight_earth
8 where theme = 'landuse'
9     and class = 'golf'
10 group by substr(quadkey, 1, 8)
```



✔ Query 3

```
1 select BING_TILE_POLYGON(BING_TILE(substr(quadkey, 1, 8))),
2     sum(
3     cast(
4         JSON_EXTRACT_SCALAR(metadata, '$.surface_area_sq_m')
5         as double)
6     ) / 1000000 AS sq_km_airport
7 from daylight_earth
8 where theme = 'landuse'
9     and class = 'airport'
10 group by substr(quadkey, 1, 8)
```



Thank you

daylightmap.org/earth

jenningsa@meta.com

jonahadkins@meta.com