Daylight Earth Tables

Jennings Anderson, Jonah Adkins, Jacob Wasserman
Maps at Meta

- OpenStreetMap
- Daylight Map Distribution
- Earth Table

Maps in Meta Products
- Instagram
- Facebook Marketplace
- And more…

map data © OpenStreetMap
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
- State of the Map 2018 & 2019
- Mobius Logical Changesets (LoChas)

2020
- March 2020: Announcing Daylight v0.1

2021
- April 2021: Daylight v1.0 released

2022
- January 2022: Daylight published on AWS Registry of Open Data as PBF and Analysis-Ready parquet files

Today
- October 2022: 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 Earth Table

<table>
<thead>
<tr>
<th>Buildings</th>
<th>Kilometers of roads / paths</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>525 M</td>
<td>78 M</td>
<td></td>
</tr>
</tbody>
</table>

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

v1.18 Summary

Most Notable 4

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

Fixed broken relation for Kapchagay Reservoir, in Almaty Oblast, Kazakhstan

Fixed a broken relation for Área Natural Protegida La Auxiliadora in Departamento de Sonsonate, El Salvador

Fixed the Área de la Bionfera de Transicion 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…
<table>
<thead>
<tr>
<th>amenity</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>baby_hatch</td>
<td>A place where a baby can be, out of necessity, anonymously left to be safely cared for and perhaps adopted.</td>
</tr>
<tr>
<td>clinic</td>
<td>A medium-sized medical facility or health centre.</td>
</tr>
<tr>
<td>dentist</td>
<td>A dentist practice / surgery.</td>
</tr>
<tr>
<td>doctors</td>
<td>A doctor's practice / surgery.</td>
</tr>
<tr>
<td>hospital</td>
<td>A hospital providing in-patient medical treatment. Often used in conjunction with emergency to note whether the medical centre has emergency facilities (A&amp;E (int.) or ER (am.))</td>
</tr>
<tr>
<td>nursing_home</td>
<td>Discouraged tag for a home for disabled or elderly persons who need permanent care. Use amenity=social_facility + social_facility=nursing_home now.</td>
</tr>
</tbody>
</table>

**Daylight Earth Table**

**Value**

- healthcare=alternative
- healthcare=audiologist
- healthcare=birthing_centre
- healthcare=blood_bank
- healthcare=blood_donation
- healthcare=centre
- healthcare=clinic
- healthcare=community_health_worker
- healthcare=counselling
- healthcare=dentist
- healthcare=dialysis
- healthcare=doctor
- healthcare=hospice
- healthcare=hospital
- healthcare=labatory
- healthcare=midwife

**Way: Riverside Regional Medical Center (298161573)**

**Version #12**

Roads and Sidewalks

Edited almost 2 years ago by sarthomster
Change set #96688275

**Tags**

- address.city: Newport News
- address.housenum: 500
- address.postcode: 23601
- address.state: VA
- address.street: J. Clyde Morris Boulevard
- amenity: hospital
- beds: 450
- emergency: yes
- healthcare: hospital
- name: Riverside Regional Medical Center
- opening_hours: 24/7
- operator: Riverside
- phone: +1 757 584 2000
- website: https://www.riversideonline.com/rmmc

**theme**

- poi: medical
- landuse: medical

**subclass**

- hospital

**metadata**

- "amenity":"hospital","is_area":true,"quadkey":"032010233003132"
- "quadkey":"032010233003132","surface_area_sq_m":267897.27

**map data © OpenStreetMap**

**wkt**

- POINT (-76.4)
- POLYGON (null)
Earth Table Schema

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

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
<table>
<thead>
<tr>
<th>Data Source</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earth Table Data Source</td>
<td>83%</td>
</tr>
<tr>
<td>Internal Data Sources that are “Earthified”</td>
<td>16%</td>
</tr>
<tr>
<td>Curated Cartographic Map Features</td>
<td>1%</td>
</tr>
</tbody>
</table>
Earth Table Schema

map data © OpenStreetMap
Accessing Daylight Earth Tables

s3://daylight-openstreetmap/earth
Accessing Daylight Earth Tables

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

<table>
<thead>
<tr>
<th>theme</th>
<th>classes</th>
<th>subclasses</th>
<th>features</th>
</tr>
</thead>
<tbody>
<tr>
<td>building</td>
<td>19</td>
<td>225</td>
<td>525,248,488</td>
</tr>
<tr>
<td>building_detail**</td>
<td>1</td>
<td>1</td>
<td>2,232,481</td>
</tr>
<tr>
<td>infrastructure</td>
<td>6</td>
<td>45</td>
<td>11,220,445</td>
</tr>
<tr>
<td>land</td>
<td>9</td>
<td>35</td>
<td>40,033,408</td>
</tr>
<tr>
<td>landuse</td>
<td>26</td>
<td>97</td>
<td>39,307,348</td>
</tr>
<tr>
<td>placename</td>
<td>3</td>
<td>11</td>
<td>3,438,464</td>
</tr>
<tr>
<td>poi</td>
<td>22</td>
<td>235</td>
<td>38,777,358</td>
</tr>
<tr>
<td>road</td>
<td>18</td>
<td>32</td>
<td>236,958,857</td>
</tr>
<tr>
<td>transit</td>
<td>6</td>
<td>32</td>
<td>6,716,466</td>
</tr>
<tr>
<td>water</td>
<td>9</td>
<td>22</td>
<td>36,481,694</td>
</tr>
</tbody>
</table>

** The building detail theme includes all building variants. 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
Getting started with Amazon Athena

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

```sql
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.

daylightmap.org/earth
Accessing Daylight Earth Tables

Daylight Earth Table

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

SQL: Ln 1, Col 1

Run again Explain Cancel Save Clear Create

Query results

Completed

Time in queue: 108 ms  Run time: 4.453 sec  Data scanned: 5.04 GB

Results (10)

<table>
<thead>
<tr>
<th>#</th>
<th>theme</th>
<th>_col1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>building</td>
<td>525248408</td>
</tr>
<tr>
<td>2</td>
<td>road</td>
<td>216890017</td>
</tr>
<tr>
<td>3</td>
<td>land</td>
<td>40831400</td>
</tr>
<tr>
<td>4</td>
<td>landuse</td>
<td>39357340</td>
</tr>
<tr>
<td>5</td>
<td>poi</td>
<td>38777950</td>
</tr>
<tr>
<td>6</td>
<td>water</td>
<td>36401694</td>
</tr>
<tr>
<td>7</td>
<td>infrastructure</td>
<td>11226463</td>
</tr>
</tbody>
</table>
SELECT * FROM daylight_earth WHERE theme = 'building' AND JSON_EXTRACT_SCALAR(metadata, '$.landuse') = 'developed:industrial'

LIMIT 100
Normalize attributes into metadata such as height

Append additional information from other themes for overlapping features
### Accessing Daylight Earth Table

#### Daylight Earth Table

[Image](daylightmap.org/earth/landuse)

### Theme: landuse

Landuses from OSM.

<table>
<thead>
<tr>
<th>Geometry Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>LineString</td>
<td>671,658</td>
</tr>
<tr>
<td>MultiLineString</td>
<td>2,862</td>
</tr>
<tr>
<td>MultiPolygon</td>
<td>97,285</td>
</tr>
<tr>
<td>Point</td>
<td>7,794</td>
</tr>
<tr>
<td>Polygon</td>
<td>36,817,998</td>
</tr>
</tbody>
</table>

### Metadata keys

<table>
<thead>
<tr>
<th>key name</th>
<th>description of values</th>
</tr>
</thead>
<tbody>
<tr>
<td>landuse</td>
<td>value of landuse tag in OSM</td>
</tr>
<tr>
<td>layer</td>
<td>value of layer: tag in OSM</td>
</tr>
<tr>
<td>level</td>
<td>value of level: tag in OSM</td>
</tr>
<tr>
<td>length_m</td>
<td>Length of feature in meters (if a line)</td>
</tr>
<tr>
<td>natural:tag</td>
<td>value of natural: tag in OSM</td>
</tr>
<tr>
<td>sport</td>
<td>value of sport: tag in OSM</td>
</tr>
<tr>
<td>surface_area_sq_m</td>
<td>Area of feature in square meters (if a polygon)</td>
</tr>
<tr>
<td>surface</td>
<td>value of surface: tag in OSM</td>
</tr>
<tr>
<td>wikidata</td>
<td>Wikidata ID (if present in OSM)</td>
</tr>
</tbody>
</table>

### Feature counts per class and subclass in Daylight v1.18

<table>
<thead>
<tr>
<th>Class</th>
<th>SubClass</th>
<th>Feature Count</th>
<th>Example Feature in OSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>agriculture</td>
<td>farmland</td>
<td>7,993,053</td>
<td><a href="#">w538454201</a></td>
</tr>
<tr>
<td>agriculture</td>
<td>farmyard</td>
<td>1,124,138</td>
<td><a href="#">w7655424881</a></td>
</tr>
<tr>
<td>agriculture</td>
<td>meadow</td>
<td>362,960</td>
<td><a href="#">w3925476793</a></td>
</tr>
<tr>
<td>agriculture</td>
<td>animal_killing</td>
<td>4,189</td>
<td><a href="#">w5729924892</a></td>
</tr>
<tr>
<td>airport</td>
<td>helipad</td>
<td>34,099</td>
<td><a href="#">w4337053101</a></td>
</tr>
<tr>
<td>airport</td>
<td>aerodrome</td>
<td>18,302</td>
<td><a href="#">w5610057208</a></td>
</tr>
</tbody>
</table>
SELECT geometry_id, class, subclass, JSON_EXTRACT_SCALAR(names, '$.local') as name, CAST(JSON_EXTRACT_SCALAR(metadata, '$.surface_area_sq_m') AS double) as area, wkt FROM daylight_earth
WHERE release = 'v1.18'
AND theme = 'landuse'
AND class = 'park'
AND subclass != 'grass'
AND quadkey like '021330011%'
Accessing Daylight Earth Table

```sql
SELECT
    geometry_id,
    class,
    subclass,
    JSON_EXTRACT_SCALAR(names, '$.local') as name,
    wkt
FROM daylight_earth
WHERE theme = 'water'
    AND class = 'river'
    AND ST_CONTAINS(ST_GEOMETRYFROMTEXT('POLYGON((-109.06409916585731 41.008769739828494,
                                      -109.0723895292973 37.000860715568436,-109.06409916585731 41.00))'),
                                      wkt)
```

map data © OpenStreetMap
Query 3

```sql
SELECT BING_TILE_POLYGON(BING_TILE(substr(quadkey, 1, 8))),
      sum(cast(JSON_EXTRACT_SCALAR(metadata, '$.surface_area_sq_m') as double)) / 1000000 AS sq_km_golf
FROM daylight_earth
WHERE theme = 'landuse'
  AND class = 'golf'
GROUP BY substr(quadkey, 1, 8)
```
Query 3

```sql
select BING_TILE_POLYGON(BING_TILE(substr(quadkey, 1, 8))),
       sum(cast(JSON_EXTRACT_SCALAR(metadata, '$.surface_area_sq_m') as double)) / 1000000 AS sq_km_airport
from daylight_earth
where theme = 'landuse'
  and class = 'airport'
group by substr(quadkey, 1, 8)
```
Thank you
daylightmap.org/earth

jenningsa@meta.com
jonahadkins@meta.com