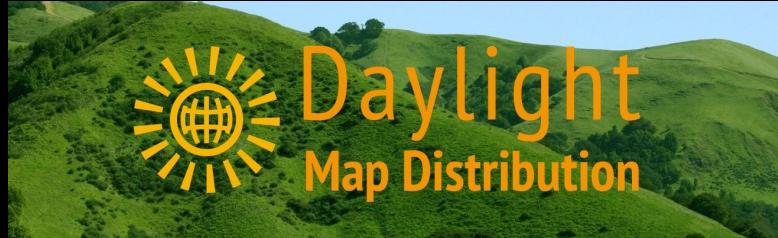


Increasing OpenStreetMap Data Accessibility with the Analysis-Ready Daylight OSM Distribution: Demonstration of Cloud-Based Assessment of Global Building Completeness

Jennings Anderson and Timmera Whaley Omidire



01 Daylight



What is Daylight?

Daylight is made of OpenStreetMap Data

Validated by the mobius team to be accurate, recent, and
vandalism-free

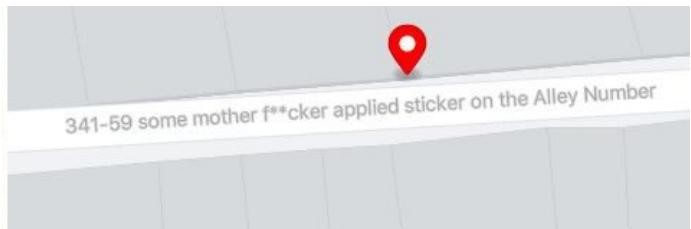
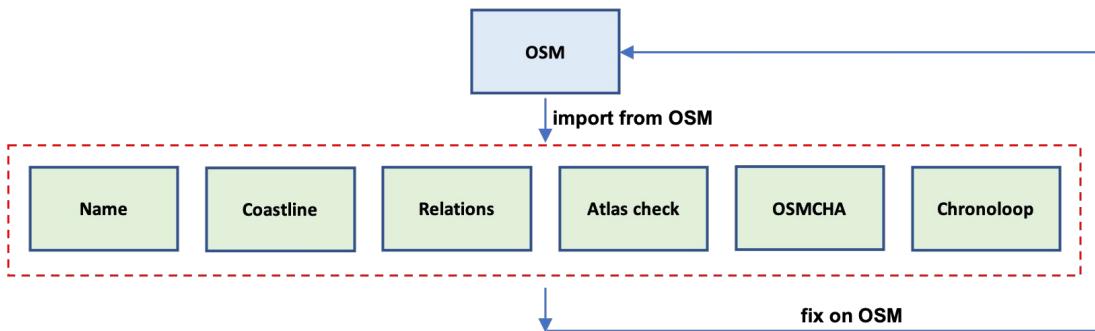
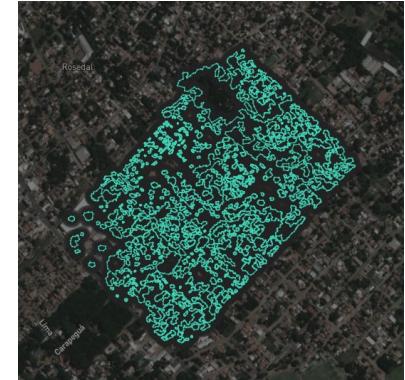
Daylight is a distribution of OpenStreetMap Data
published with additional datasets such as microsoft
buildings and experimental ML roads.

Daylight Process: Find-Fix-Import Loop

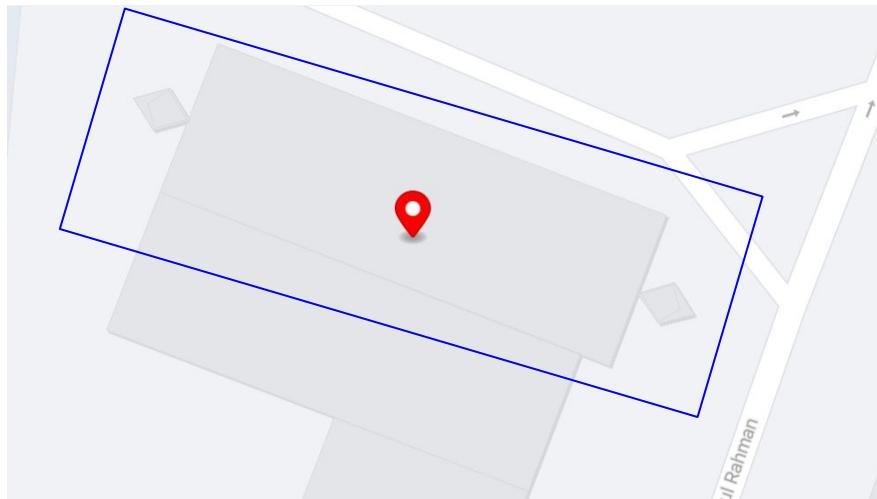
Find: Discover errors or other issues anywhere on the map

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

Import: Apply fixes from OSM into the Daylight map

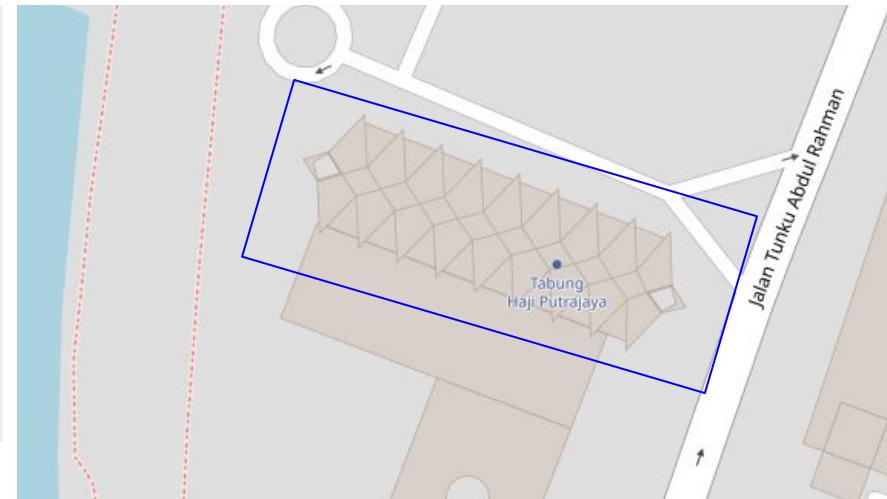


OSM Relations Example



Quality issue

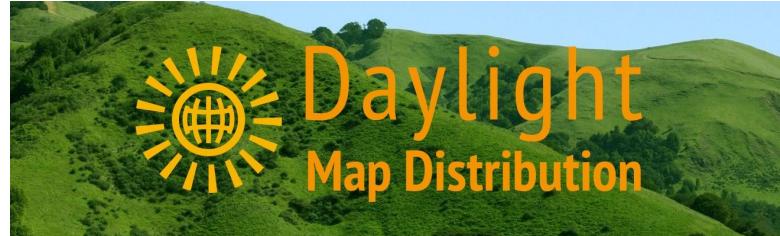
Geometric roof of Tabung Haji Putrajaya not rendering on map



Quality fix

Dissolved parts of relation and added roof tags to each individual way
Changeset 123292182

Timeline of Daylight
OpenStreetMap Distribution



< 2019 2020 2021 2022 Today



State of the Map
2018 & 2019

Mobius Logical
Changesets
(LoChas)

March 2020

Announcing
Daylight v0.1

April 2021

Daylight v1.0
released

January 2022

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

August 2022

Daylight v1.16

<https://daylightmap.org>

Daylight v1.16

517 M

Buildings

78 M

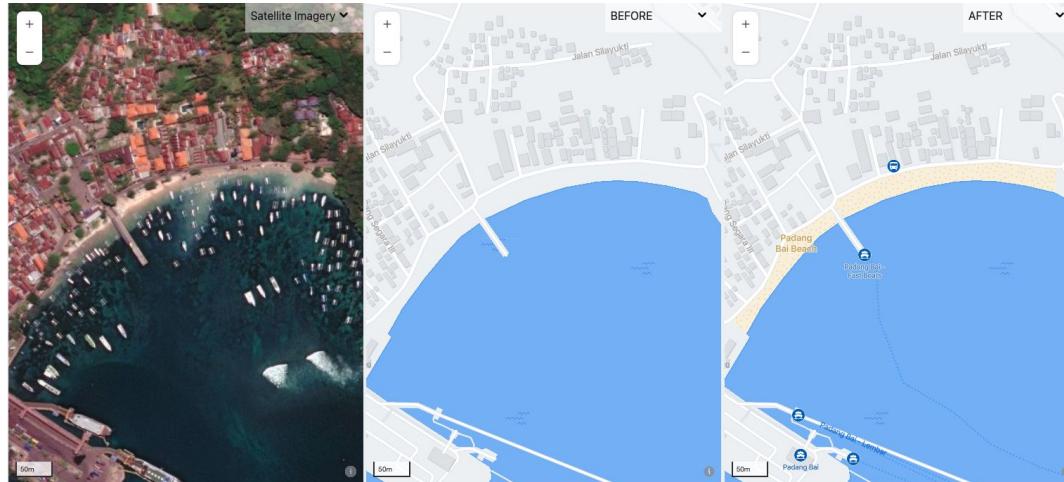
Kilometers of roads / paths

100%

OpenStreetMap Data

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

Repaired beach relation for Padang Bai Beach in Bali, Indonesia



02 Analysis-Ready Daylight OpenStreetMap Distribution

A Generalized OpenStreetMap Data Analysis Pipeline



- OSM Planet PBF
- Overpass
- OSM Galaxy / HOT Export Tool
- OSHDB
- GeoFabrik Extracts

First analysts obtain
OSM data from any
number of sources

A Generalized OpenStreetMap Data Analysis Pipeline



- OSM Planet PBF
- Overpass
- OSM Galaxy / HOT Export Tool
- OSHDB
- GeoFabrik Extracts

First analysts obtain OSM data from any number of sources

- osmium extract
- osmium tag filter

It can be more efficient to pre-process the data depending on the format and use case.

Osmium tag-filter, for example is extremely fast.

A Generalized OpenStreetMap Data Analysis Pipeline



- OSM Planet PBF
- Overpass
- OSM Galaxy / HOT Export Tool
- OSHDB
- GeoFabrik Extracts

First analysts obtain OSM data from any number of sources

- osmium extract
- osmium tag filter

It can be more efficient to pre-process the data depending on the format and use case.

Osmium tag-filter, for example is extremely fast.

- osm2pgsql
- ogr2ogr

At some point, nodes, ways, and relations need to be converted into Points, Lines, and Polygons and the OSM tags turned into columns or key/value pairs

GIS tools for analysis

A Generalized OpenStreetMap Data Analysis Pipeline



- OSM Planet PBF
- Overpass
- OSM Galaxy / HOT Export Tool
- OSHDB
- GeoFabrik Extracts

First analysts obtain OSM data from any number of sources

Pre-process as desired

- osmium extract
- osmium tag filter

It can be more efficient to pre-process the data depending on the format and use case.

Osmium tag-filter, for example is extremely fast.

Convert to Analysis-Ready Format

- osm2pgsql
- ogr2ogr

At some point, nodes, ways, and relations need to be converted into Points, Lines, and Polygons and the OSM tags turned into columns or key/value pairs

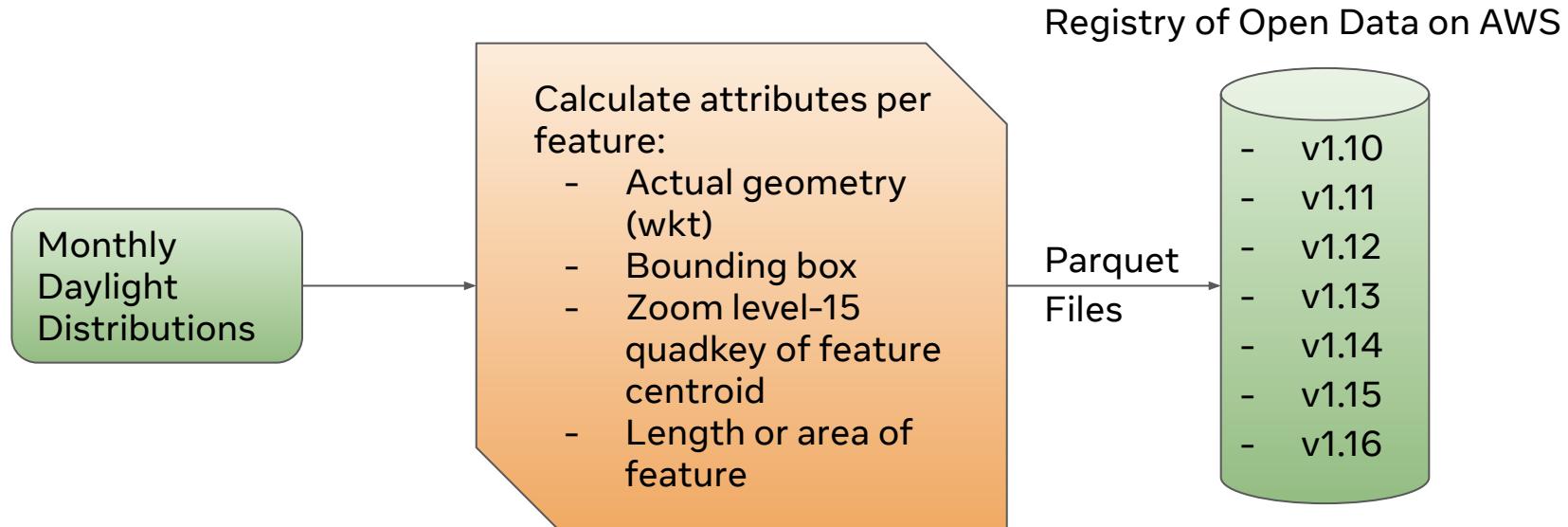
GIS tools for analysis

- QGIS
- Jupyter Notebooks
- Kepler.gl
- R Studio

Now we're ready to explore our data

Analysis Ready OSM Pipeline





Relation: Plesso
Didattico Viale
Morgagni (8128765)

Version #3

Additions and Fixes - Centro Didattico Morgagni

Edited about 4 years ago by

ananonymousmapper

Changeset #59101379

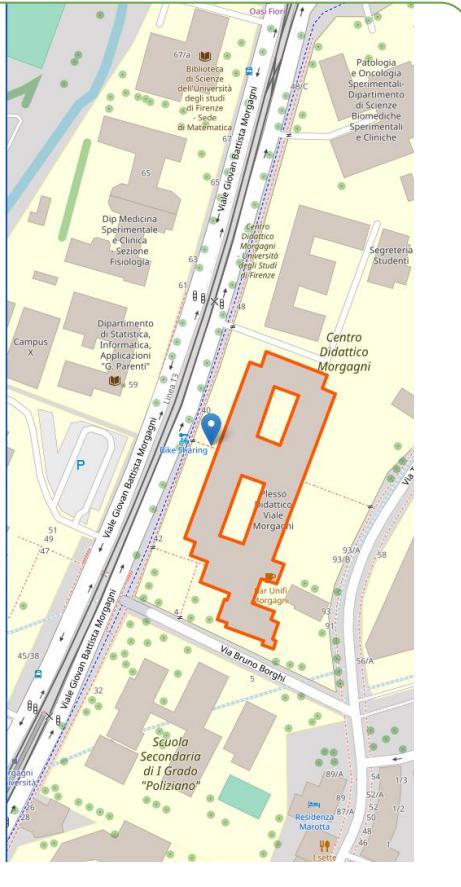
Tags

addr:city	Firenze
addr:housenumber	44-48
addr:postcode	50134
addr:street	Viale Giovan Battista Morgagni
building	university
building:levels	3
email	reception.morgagni4 4@polibiotec.unifi.it
name	Plesso Didattico Viale Morgagni
name_1	Centro Didattico Morgagni
phone	+39 055 275 1016
phone_1	+39 055 275 1015
type	multipolygon

Members

- ▼ 3 members
 - Way 57728745 as outer
 - Way 571092456 as inner
 - Way 571092453 as inner

[Download XML](#) · [View History](#)



Relation: Plesso Didattico Viale Morgagni (8128765)

Version #3

Additions and Fixes - Centro Didattico Morgagni

Edited about 4 years ago by

anonymousmapper

Changeset #59101379

Tags

addr:city	Firenze
addr:housenumber	44-48
addr:postcode	50134
addr:street	Viale Giovan Battista Morgagni
building	university
building:levels	3
email	reception.morgagni4@polobiotec.unifi.it
name	Plesso Didattico Viale Morgagni
name_1	Centro Didattico Morgagni
phone	+39 055 275 1016
phone_1	+39 055 275 1015
type	multipolygon

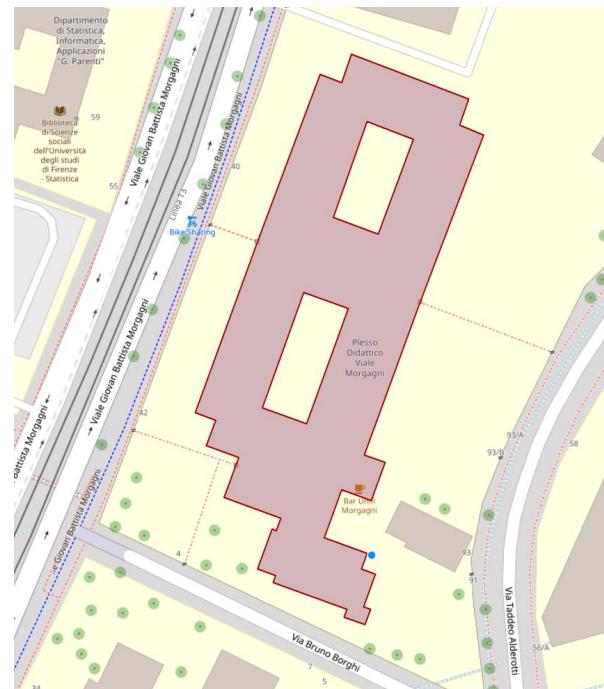
Members

▼ 3 members

Way 57728745 as outer
Way 571092456 as inner
Way 571092453 as inner

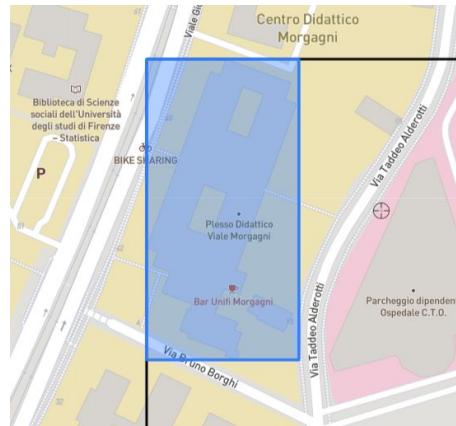
[Download XML](#) · [View History](#)

Standard OSM Attributes



WKT: MULTIPOLYGON(((11.2449916
43.8006028,11.2450614 43.8005845,11.24....

Area: 5431.48 m²



Envelope: (11.2449916, 43.800078)
(11.2459798, 43.8014907)

Analysis-Ready Attributes

Analysis-Ready Daylight in Action

How many buildings are in Daylight OSM?

Query 1

```
1 SELECT COUNT(*)
2 FROM daylight_osm_features
3 WHERE tags [ 'building' ] <> 'no'
4 AND release = 'v1.16'
```

How many buildings are in Daylight OSM?

Query 1

```
1 SELECT COUNT(*)
2 FROM daylight_osm_features
3 WHERE tags [ 'building' ] <> 'no'
4 AND release = 'v1.16'
```

SQL Ln 4, Col 24

Run again Explain Cancel Save Clear Create

Query results | Query stats

Completed Time in queue: 117 ms Run time: 10.091 sec Data scanned: 13.05 GB

Results (1)

Copy Download results

Search rows

_col0

#	_col0
1	517664923

How many buildings are in Daylight OSM?

Query 1

```
1 SELECT COUNT(*)
2 FROM daylight_osm_features
3 WHERE tags [ 'building' ] <> 'no'
4 AND release = 'v1.16'
```

SQL Ln 4, Col 24

Run again Explain Cancel Save ▾

Query results Query stats

Completed

Results (1)

Search rows

_col0

1 517664923

 taginfo

KEYS · TAGS · RELATIONS · PROJECTS · REPORTS · ABOUT

building

To mark the outline of a building, a man-made structure with a roof, standing more or less permanently in one place

Overview Values Combinations Similar Chronology Map Wiki Projects Characters

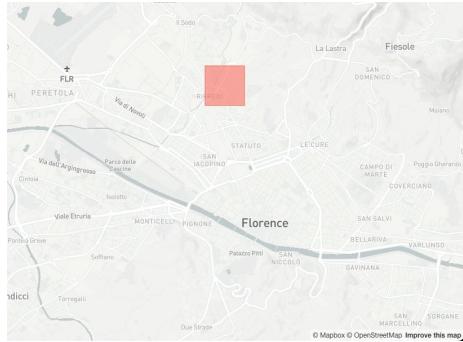
Overview

Type	Number of objects	Number of values
All	523 970 916 (6.00%)	7 978
Node	1 125 142 (0.57%)	1 122
Way	521 940 603 (59.38%)	7 197
Relation	905 171 (8.94%)	755

Objects with this key were last edited by 710 432 different users.

Global Building Counts

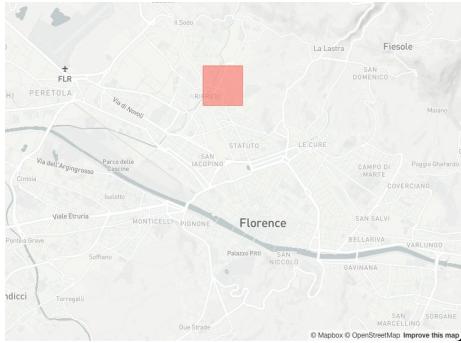
```
SELECT BING_TILE_POLYGON(  
    BING_TILE(  
        SUBSTR(  
            quadkey, 1, 7  
        )),  
        COUNT(*)  
    )  
FROM daylight_osm_features  
WHERE release = 'v1.13'  
    AND tags ['building'] <> 'no'  
GROUP BY SUBSTR(quadkey, 1, 7)
```



quadkey = 120223131311311

Global Building Counts

```
SELECT BING_TILE_POLYGON(  
    BING_TILE(  
        SUBSTR(  
            quadkey, 1, 7  
        )),  
        COUNT(*)  
    )  
FROM daylight_osm_features  
WHERE release = 'v1.13'  
    AND tags ['building'] <> 'no'  
GROUP BY SUBSTR(quadkey, 1, 7)
```



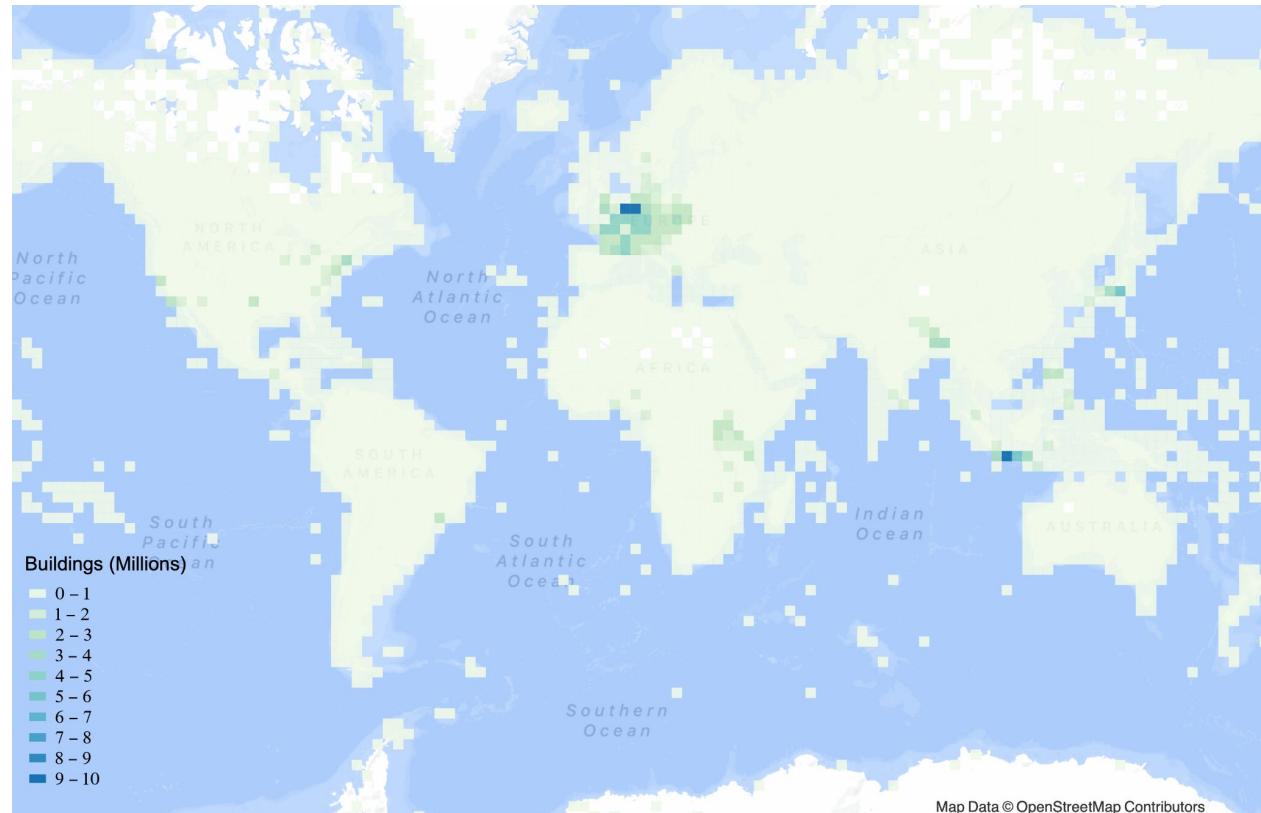
quadkey = 120223131311311



quadkey = 1202231

Global Building Counts

```
SELECT BING_TILE_POLYGON(  
    BING_TILE(  
        SUBSTR(  
            quadkey, 1, 7  
        )),  
        COUNT(*)  
    )  
FROM daylight_osm_features  
WHERE release = 'v1.13'  
    AND tags ['building'] <> 'no'  
GROUP BY SUBSTR(quadkey, 1, 7)
```



Map Data © OpenStreetMap Contributors

What about using spatial
boundaries?

```
SELECT tags [ 'highway' ] AS
highway_tag,
      SUM(linear_meters) /
1000 as total_km
FROM daylight_osm_features
WHERE release = 'v1.16'
      AND tags [ 'highway' ]
          IS NOT NULL
      AND linear_meters > 0
      AND ST_CONTAINS( <PR>
ST_GEOGRAPHYFROMTEXT(wkt)
      )
GROUP BY tags [ 'highway' ]
ORDER BY total_km DESC
```



```
ST_GEOGRAPHYFROMTEXT(
  'POLYGON((-67.51 17.65,-67.44 18.78,-65.16 18.63,-65.06 17.99,-66.40
17.57,-67.51 17.65))')
```

What about using spatial boundaries?

```
SELECT tags [ 'highway' ] AS
highway_tag,
      SUM(linear_meters) /
1000 as total_km
FROM daylight_osm_features
WHERE release = 'v1.16'
      AND tags [ 'highway' ]
      IS NOT NULL
      AND linear_meters > 0
      AND ST_CONTAINS( <PR>
ST_GEOFFOMETRYFROMTEXT(wkt)
      )
GROUP BY tags [ 'highway' ]
ORDER BY total_km DESC
```

#	highway_tag	total_km
1	residential	21907
2	track	5734
3	tertiary	5626
4	service	3558
5	secondary	1772
6	unclassified	1589
7	motorway	888
8	path	611
9	primary	541
10	trunk	395
11	footway	323
12	motorway_link	315
13	trunk_link	64
14	cycleway	34
15	primary_link	33
16	secondary_link	29
17	tertiary_link	28
18	construction	24

What about using spatial boundaries?

```
SELECT id,  
       tags [ 'highway' ],  
       linear_meters,  
       wkt AS geometry  
FROM daylight_osm_features  
WHERE release = 'v1.16'  
      AND tags [ 'highway' ]  
      IS NOT NULL  
      AND linear_meters > 0  
      AND ST_CONTAINS( <PR>  
        ST_GEOGRAPHYFROMTEXT(wkt)  
      )
```

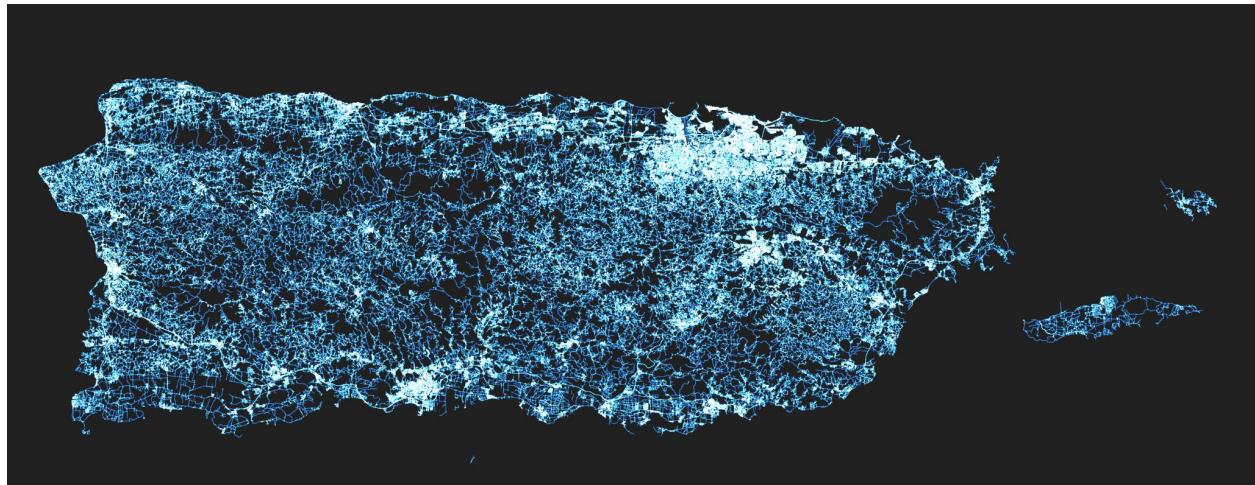
Puerto_rico_roads.csv

id	highway	meters	geometry
22242081	residential	172	LINESTRING(-66.026714 18.375809,-6
22115107	residential	284	LINESTRING(-67.125756 18.087181,-6 18.0882486,-67.12494 18.0888079,-6
543251238	service	73	LINESTRING(-66.8028986 18.4519378,
717378341	residential	412	LINESTRING(-67.0995413 18.03515,-6 18.038776)

39mb, 166k rows

What about using spatial boundaries?

```
SELECT id,  
       tags [ 'highway' ],  
       linear_meters,  
       wkt AS geometry  
FROM daylight_osm_features  
WHERE release = 'v1.16'  
      AND tags [ 'highway' ]  
          IS NOT NULL  
      AND linear_meters > 0  
      AND ST_CONTAINS( <PR>  
        ST_GEOGRAPHYFROMTEXT(wkt)  
    )
```



Puerto_rico_roads.csv

id	highway	meters	geometry
22242081	residential	172	LINESTRING(-66.026714 18.375809,-6
22115107	residential	284	LINESTRING(-67.125756 18.087181,-6 18.0882486,-67.12494 18.088079,-6
543251238	service	73	LINESTRING(-66.8028986 18.4519378,
717378341	residential	412	LINESTRING(-67.0995413 18.03515,-6 18.038776)



39mb, 166k rows

Incorporating other Datasets

OSM Changesets have been available on AWS via the Registry of Open Data since 2017

The screenshot shows a web browser window displaying the AWS Registry of Open Data. The URL in the address bar is `https://registry.opendata.aws/osm/`. The page title is "OpenStreetMap on AWS". Below the title, there are several tabs: "disaster response", "geospatial", "mapping", "osm", and "sustainability".

Description
OSM is a free, editable map of the world, created and maintained by volunteers. Regular OSM data archives are made available in Amazon S3.

Update Frequency
Data is updated weekly

License
<https://www.openstreetmap.org/copyright>

Documentation
<https://github.com/awslabs/open-data-docs/tree/main/docs/osm-pds>

Managed By
Pacific Atlas
See all datasets managed by [Pacific Atlas](#).

Contact
<https://github.com/mojonada/osm-pds-pipelines/issues>

How to Cite
OpenStreetMap on AWS was accessed on `DATE` from <https://registry.opendata.aws/osm>.

Usage Examples
Tutorials

- [AWS Develop and Extract Value from Open Data](#) by Daniel Bernao
- [Querying OpenStreetMap Changesets with Amazon Athena](#) by Jennings Anderson

Resources on AWS

Description	Resource type	Amazon Resource Name (ARN)	AWS Region
Imagery and metadata	S3 Bucket	<code>arn:aws:s3:::osm-pds</code>	<code>us-east-1</code>
New data notifications	SNS Topic	<code>arn:aws:sns:us-east-1:80021804198:New_File</code>	<code>us-east-1</code>

AWS CLI Access (No AWS account required)
`aws s3 ls --no-sign-request s3://osm-pds/`

<https://registry.opendata.aws/osm/>

Incorporating other Datasets

OSM Changesets have been available on AWS via the Registry of Open Data since 2017

```
SELECT * FROM changesets
WHERE tags['comment']
LIKE '%#hotosm%'
```

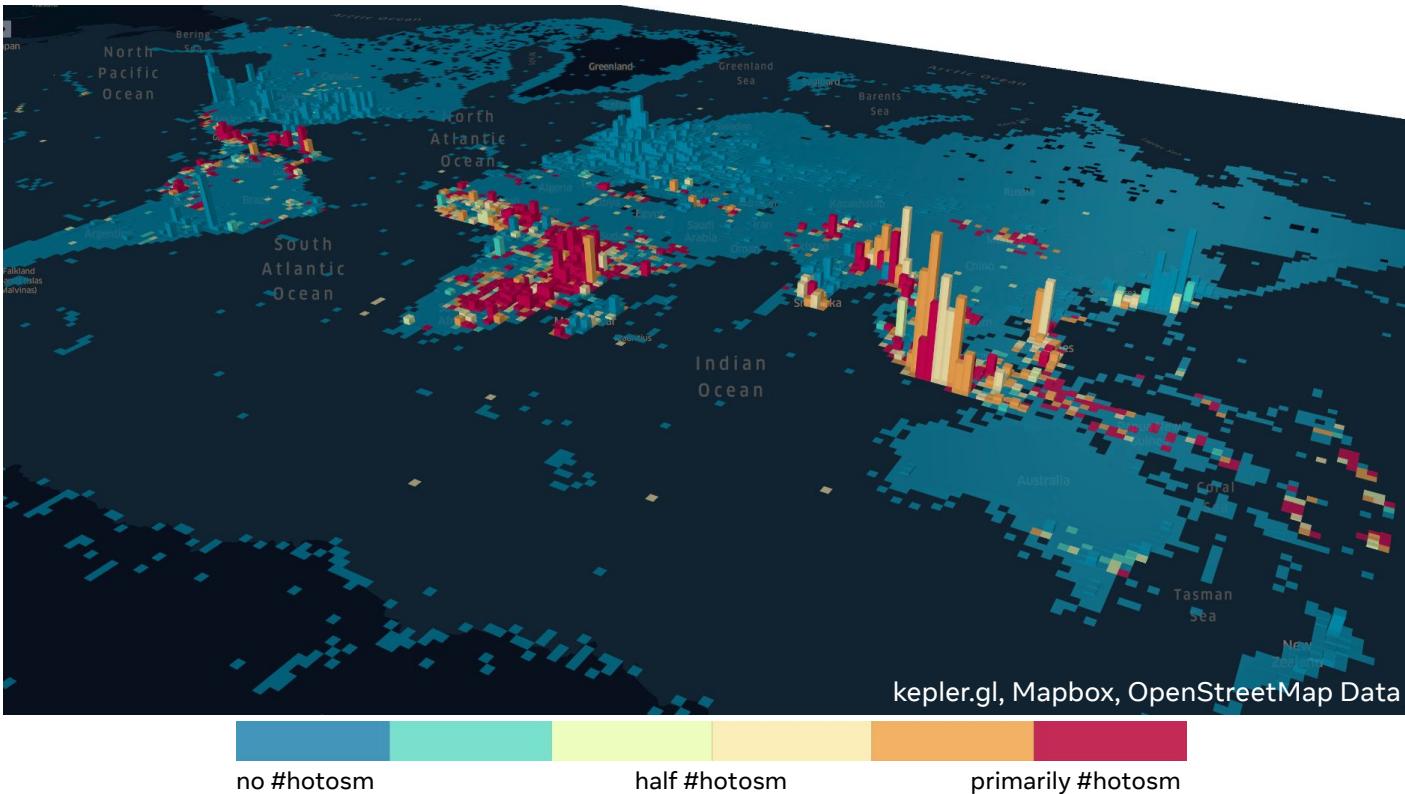
#	id	tags	created_at	uid	min_lat	max_lat	min_lon
1	124027360	{"changesets_count": "12", "comment": "#hotosm-project-12825 #OSMBD #Bangladesh #BangladeshFloodResponse2022", "created_by": "ID 2.20.2", "hashtags": "#hotosm-project-12825;#OSMBD;#Bangladesh;#BangladeshFloodResponse2022", "host": "https://tasks.hotosm.org/projects/12825/map/", "imagery_used": "Maxar Premium Imagery;gpx data file", "locale": "en", "review_requested": "yes"}	2022-07-24 23:57:52.000	16536796	24.7133899	24.7137115	91.8185726
2	124027343	{"changesets_count": "598", "comment": "#hotosm-project-12825 #OSMBD #Bangladesh #BangladeshFloodResponse2022", "created_by": "ID 2.20.2", "hashtags": "#hotosm-project-12825;#OSMBD;#Bangladesh;#BangladeshFloodResponse2022", "host": "https://tasks.hotosm.org/projects/12825/map/", "imagery_used": "Maxar Premium Imagery;gpx data file", "locale": "en"}	2022-07-24 23:56:35.000	9215893	25.1358742	25.1415300	92.1332617
3	124027335	{"changesets_count": "15", "comment": "#hotosm-project-12691 #OpenCitiesLAC #ourimpact #MMCSeoul2022", "created_by": "ID 2.20.2", "hashtags": "#hotosm-project-12691;#OpenCitiesLAC;#ourimpact;#MMCSeoul2022;#MarshSeoul22", "host": "https://tasks.hotosm.org/projects/12691/map/", "imagery_used": "Maxar Premium Imagery", "locale": "ko"}	2022-07-24 23:56:11.000	16512513	15.1015880	15.1025018	-91.6396594

<https://registry.opendata.aws/osm/>

Incorporating other Datasets

Join Analysis-Ready Daylight Distribution to OSM Changesets table from OpenStreetMap on AWS RODA.

```
SELECT COUNT( buildings )
  FROM daylight_osm d
  JOIN changesets c
    ON d.changeset = c.id
 WHERE c.tags['comment']
      LIKE '%#hotosm%'
```

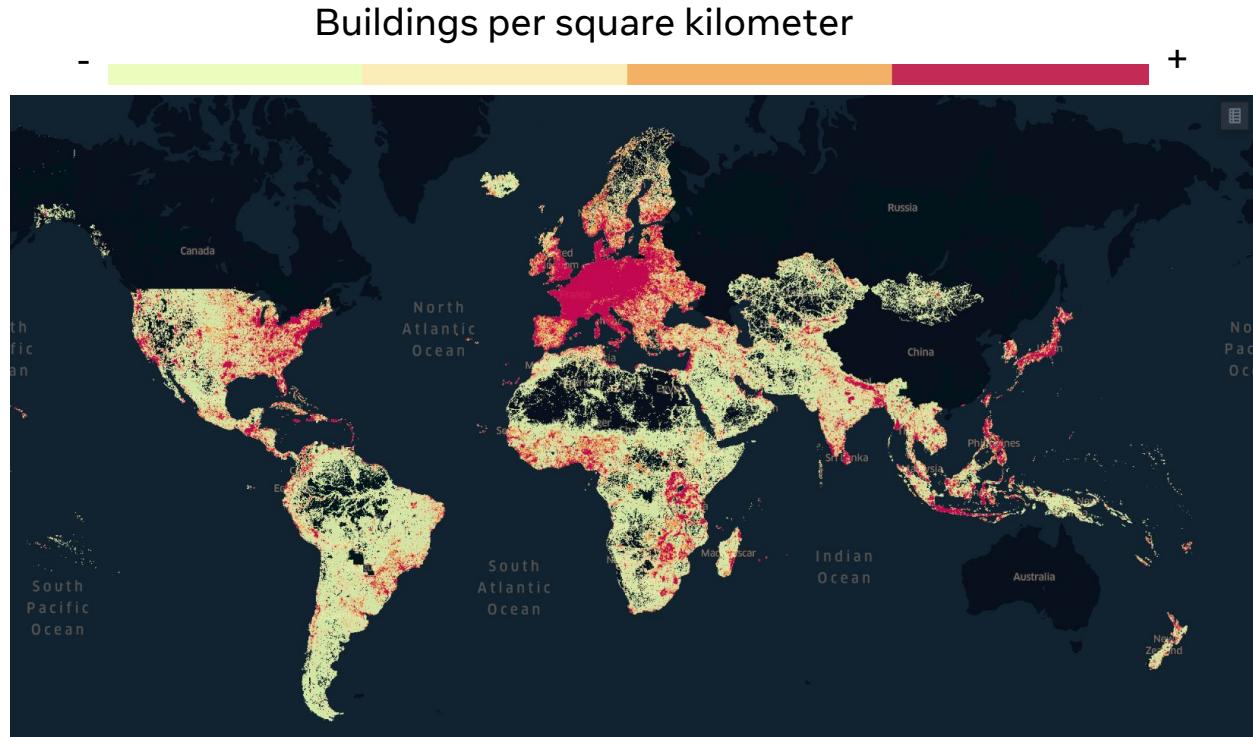


Buildings in OSM that were last edited in a changeset with #hotosm hashtag

OpenStreetMap Building Density

Calculated the same way as before, but with smaller tiles.

Additionally, for each tile, we calculate the area and compute the actual density of buildings in terms of buildings per square kilometer — otherwise our data could not be compared between different latitudes.



Kepler.gl, Mapbox, OpenStreetMap Data

High Resolution Settlement Layer (HDSL)

High Resolution population data for much of the world.

Also available as an open dataset on the Registry of Open Data on AWS

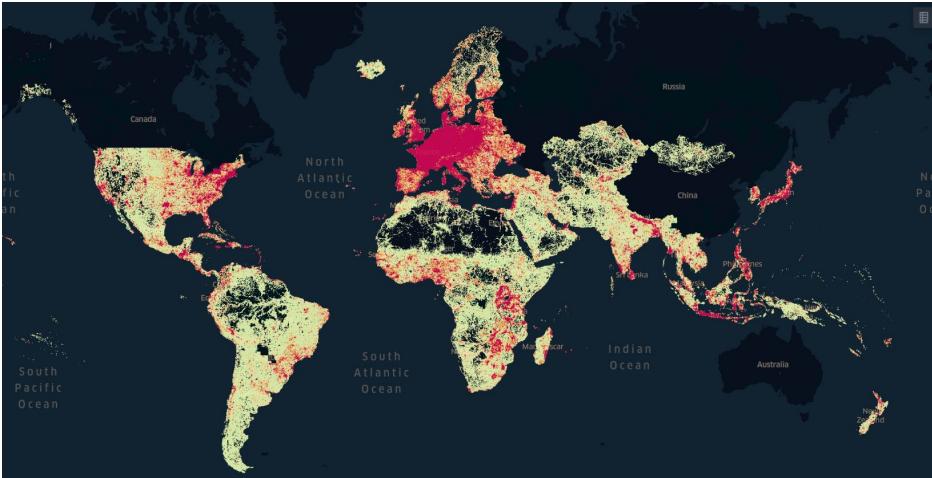
Like Analysis-Ready Daylight OSM, you can easily load it directly into athena without having to download it.

Once in Athena, **JOIN hrs1**

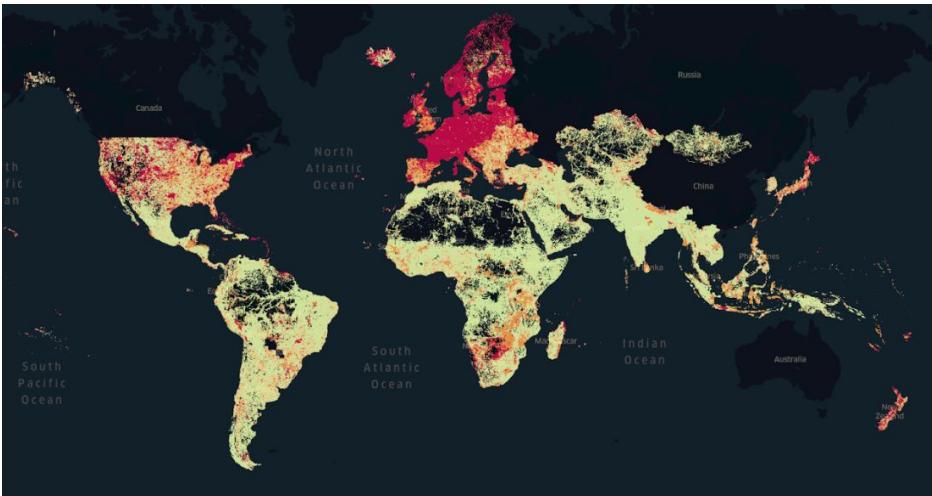
The screenshot shows a web browser window with the URL registry.opendata.aws. The page title is "Registry of Open Data on AWS". In the top right corner, there is a "aws" logo. The main content features a large heading: "High Resolution Population Density Maps + Demographic Estimates by CIESIN and Meta". Below the heading is a row of tags: aerial imagery, demographics, disaster response, geospatial, image processing, machine learning, population, satellite imagery, and sustainability. A "Description" section follows, explaining that population data is allocated to 1 arcsecond blocks and provided in CSV and Cloud-optimized GeoTIFF files, refined from CIESIN's Gridded Population of the World using machine learning models on high-resolution worldwide Maxar satellite imagery. The "Update Frequency" is listed as "Quarterly". The "License" is "CC BY 4.0", with a link to <https://creativecommons.org/licenses/by/4.0/>. The "Documentation" section links to "Project overview" and "instructions for use with AWS Athena". The "Managed By" section features the Meta logo and a link to "See all datasets managed by Meta". The "Contact" section includes an email address: disastermaps@fb.com.

OpenStreetmap Building Density Normalized by Population

- USA appears with much higher coverage overall
- Eastern Africa, Nepal, Philippines, are all not *as complete* with regards to building coverage
- Majority of Europe is still *very complete*



OSM
Buildings
per km²



OSM Buildings
per km²
per person

Summary

- Analysis-Ready Daylight Distribution of OSM enables scalable analysis of OSM data in the cloud
- Not the first global OSM building density vs population assessment, but the first to run in just a few minutes with a few dozen lines of code.
- Expanding the OSM Data Science Toolkit

03 Advantages and Current Limitations

Advantages

- Contains common geospatial features such as points, lines, and polygons
- Utilizes basic knowledge of SQL to conduct analysis quickly & cost-effectively
- Not locked into AWS due to Apache Parquet files

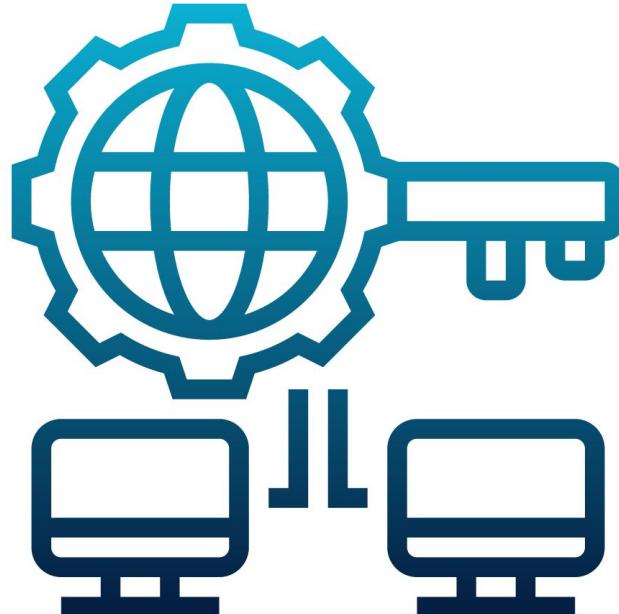
Current Limitations

- Not suitable for OSM history data
 - Exploration of OSM historical data better suited in Ohsome

Exploratory questions

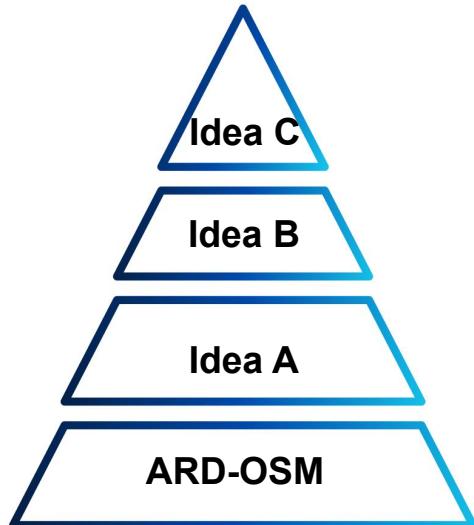
1. What type of research and innovation will emerge from utilizing ARD-OSM in the future?
2. How can the OSM research community expand on distributing OSM data through cloud-based database technologies?
3. What are the advantages of creating a pre-computed complete history in this environment vs utilizing Ohsome for investigations?

Conclusion



- Lowering barriers to entry for other researchers to increase data accessibility
- Reduce the heavy lift of data engineering
- Provide common geospatial vector formats used in GIS

Conclusion



- ARD-OSM is not a one size fits all tool (ex: Ohsome)
- No longer have to download and process the entire planet file
- Geospatial data backbone for others to build tools on

Thank you.

daylightmap.org

registry.opendata.aws/daylight-osm