OVERTURE MAPS FOUNDATION
STARTING SOON:

Working With Overture Data: A Step-by-Step Guide

Meta Research Scientist Jennings Anderson discusses Overture’s use of cloud-native parquet format and provides a step-by-step guide to accessing and visualizing the initial Overture data release.
Marc Prioleau
Executive Director of Overture Maps
Jennings Anderson, PhD
Research Scientist at Meta
Resources

docs.overturemaps.org

github.com/overturemaps/data

github.com/overturemaps/schema
What is in the 2023-07-26-alpha.0 data release?

...and why is there no “download all the data here” button?
What is in the 2023-07-26-alpha.0 data release?

4 Themes:

1. Administrative Boundaries (admins)
2. Buildings
3. Places
4. Transportation
Administrative Boundaries (admins) at-a-glance

99,403 features

2,948 locality

i. adminLevel=2 (265) (Polygon/Multipolygon)

ii. adminLevel=4 (2,683) (Multipolygons)

96,455 administrativeBoundary (LineStrings)

iii. adminLevel=2 (18,825)

iv. adminLevel=4 (77,630)
Buildings at-a-glance

785,524,851 building features
- 27,871,758 have height information (at least 10M from USGS Lidar)
- 562,731,640 from OpenStreetMap
- 211,937,251 from Microsoft ML Buildings Dataset
- 10,855,960 from Esri Community Maps

What is the order of conflation?
OpenStreetMap > Esri > Microsoft ML
Places at-a-glance

59,175,720 place features

- 51,669,480 from Meta
- 4,681,011 from Microsoft
- 2,825,229 from both
Transportation at-a-glance

Re-segmented OSM road network:

- 294,327,662 segment features
- 330,606,087 connector features
Let’s dig in

Using Amazon **Athena** as an interactive query environment to read Overture data

Athena is just one way to access the data. Instructions are available on github.

github.com/overturemaps/data
Places Queries

SELECT COUNT(*) FROM places

SELECT * FROM places LIMIT 10

SELECT ROUND(confidence * 10)/10 AS _conf, COUNT(id) as _count
FROM places
GROUP BY ROUND(confidence * 10)
ORDER BY _conf DESC
WITH places_with_quadkey AS (  
    SELECT bing_tile_quadkey(
        BING_TILE_AT(
            ST_Y(ST_GeomFromBinary(geometry)),
            ST_X(ST_GeomFromBinary(geometry)),
            8
        )
    ) AS q8,
    id
    
    FROM places
    WHERE confidence > 0.8
)

SELECT BING_TILE_POLYGON(BING_TILE(q8)),
    count(id) as num_places
FROM places_with_quadkey
GROUP BY q8
SELECT TRY(
    FILTER(
        names [ 'common' ],
        name->name [ 'language' ] = 'local'
    ) [ 1 ] [ 'value' ]
) AS name,
categories.main AS category,
confidence,
ST_GeomFromBinary(geometry) as wkt
FROM places
WHERE confidence > 0.8
    AND bbox.minX > -126.7952
    AND bbox.maxX < -118.5453
    AND bbox.minY > 43.5453
    AND bbox.maxY < 50.4344

Download some data as a CSV with Athena
WITH buildings_with_quadkey AS (  
SELECT  
    bing_tile_quadkey(  
        BING_TILE_AT(  
            (bbox.maxY + bbox.minY)/2,  
            (bbox.maxX + bbox.minX)/2,  
            8  
        )  
    ) AS q8,  
    id,  
    CARDINALITY(  
        FILTER(sources, x -> x['dataset'] = 'OpenStreetMap')  
    )>0 AS osm_building  
FROM buildings  
)  
SELECT  
    BING_TILE_POLYGON(BING_TILE(q8)),  
    COUNT(id) as num_buildings,  
    COUNT_IF(osm_building) AS osm_building,  
    COUNT_IF(osm_building) / CAST(COUNT(id) AS double) AS percent_osm  
FROM buildings_with_quadkey  
GROUP BY q8
SELECT class, height,
    ST_GeomFromBinary(geometry) as wkt
FROM buildings
WHERE ST_CONTAINS(
    ST_GeometryFromText(
    'POLYGON((-122.36719284258956
    47.618321237733284,-122.33594394153602
    47.63240470851924,-122.2775808079059
    47.61236859966664,-122.34462990362489
    47.58012171471199,-122.36719284258956
    47.618321237733284))'
    ),
    ST_GeomFromBinary(geometry)
)
Still no download button?

Athena is just one way to access the data.

Today, we transformed, downloaded, analyzed, and grouped the data all without needing to first download and process a massive file.

If you must download all of the data, the parquet files are available!

github.com/overturemaps/data
What’s Next?

Future releases will continue improving in coverage and quality on the existing themes.

The Global Entity Reference System (GERS) will assign stable IDs to features in the 4 themes, allowing matching of external datasets and easy id-based conflation.
Thank You