

OUERTURE 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:

- Administrative Boundaries (admins)
- 2. Buildings
- 3. Places
- 4. Transportation

••• 🗆 - <	>			₾ +	G
i Overture Maps	Data Schema	Global Entity Reference System		÷ộ	6
Introduction		throduction			
Schema Themes	~	Pr	inciples		ι
Admins		Introduction	sense		- 1
Buildings					- 1
Places		Welcome to this preview of the Overture Data Schema which currently focuses on buildings, places, transportation, and			- 1
Transportation		admin data layers. We've tocused on designing a data schema that is easy for developers to quickly understand and use in building map products. You will find a higher-level overview of the schema under each Theme, or view the technical			- 1
Schema Reference	~	details and example implementations under the Schema Reference.			- 1
admins	~	Please engage in feedback on GitHub; we really appreciate your input to improve future iterations.			- 1
administrativeBound	ary				- 1
locality		Principles			- 1
buildings	\sim	1 melphoo			- 1
building		Address the core, enable the periphery. The Overture schema doesn't solve every problem. It describes fully-formed			- 1
places	~	solutions only for the most fundamental use cases (the core) while enabling less common use cases (the perpirety) via extensibility.			- 1
place		terrent energy the new Many enveloped adults a failure and the destination is been environmented a			- 1
transportation	~	already exist and are well-understood in the community. The Overture schema reuses these existing solutions to			- 1
connector		maximize compatibility and focus our efforts on solving unaddressed high-priority pain points.			
segment		Backward-compatible is forward-compatible. No design is future-proof, but good designs maintain relevance by			
		adding features without breaking existing use cases. The Overture schema can be enhanced in a backward-compatible			
		way.			
		Always open. The Overture schema and format aim for compatibility with free and open-source tools and technologies			
		and avoid depending on closed-source or proprietary tools and technologies.			
		Key Specifications			
		1. The Overture Data Schema is documented in JSON Schema, a widely used, machine-readable schema declaration			
		language for JSON documents.			



docs.overturemaps.org

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



Transportation at-a-glance

(3)

0

0

6)

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

ሰ + በ

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
```

Where is the places coverage?

```
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
```

Download some data as a CSV with Athena

```
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
```

Analyze buildings coverage globally

```
WITH buildings_with_quadkey AS (
  SELECT
    bing_tile_quadkey(
      BING_TILE_AT(
        (bbox.maxY + bbox.minY)/2
        (bbox.maxX + bbox.minX)/2,
        8
    ) AS g8,
    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
```

Download some buildings in Seattle



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!

README.md		0	Report repository		
Nelcome 1	to the Overture Maps Data Repo		Releases		
This repository includes instructions and sample queries to access Overture Maps Data.			No releases published Create a new release		
Ve also welcome te chema, is best prov	edback about Overture Maps data in the Discussions, Feedback on the data ided in the discussions in the schema repository.		Packages		
Accessing Ov	erture Maps Data	No packages published Publish your first package			
Overture Maps data Ilanet" file to be do 1pha.0 release by Data Location	is available in cloud-native Parquet format. There is no single Overture *entire whoaded. Instead, we have organized the data for the 'overture' 2023-07-26- theme and type at the following locations:		Contributors 10		
Theme	Location		w & +		
Admins	 Amazon S3: is1//overturemaps-us-west-2/release/2023-07-24- alpha.0/theme-admins Microsoft Azure: https://overturemapsetus2.blob.core.windows.net/release/2023- 07-26-alpha.0/theme-admins 				
Buildings	 Amazon S3: s3://overturemaps-us-west-2/release/2023-07-26- alpha.g/themebuildings Microsoft Azure: https://overturemapsestus2.blob.core.windows.net/release/2023- 07-26-alpha.0 themebuildings 				
Places	 Amazon S3: s3://overturemaps-us-west-2/release/2023-07-26- alpha.8/theme-places Microsoft Azure: https://overturemapswestus2.blob.core.windows.net/release/2023- 07-20-alpha.0/thems-places 				
Transportation	 Amazon S3: s3://overturemaps-us-west-2/release/2023-07-26- alpha.@/thmeetranoptration Microsoft Azure: https://overturemapsestus2.blob.core.windows.net/relase/2023- 07-28-alpha.@/themetranoptration 				

github.com/overturemaps/data





Future releases will continue improving in coverage and quality on the 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