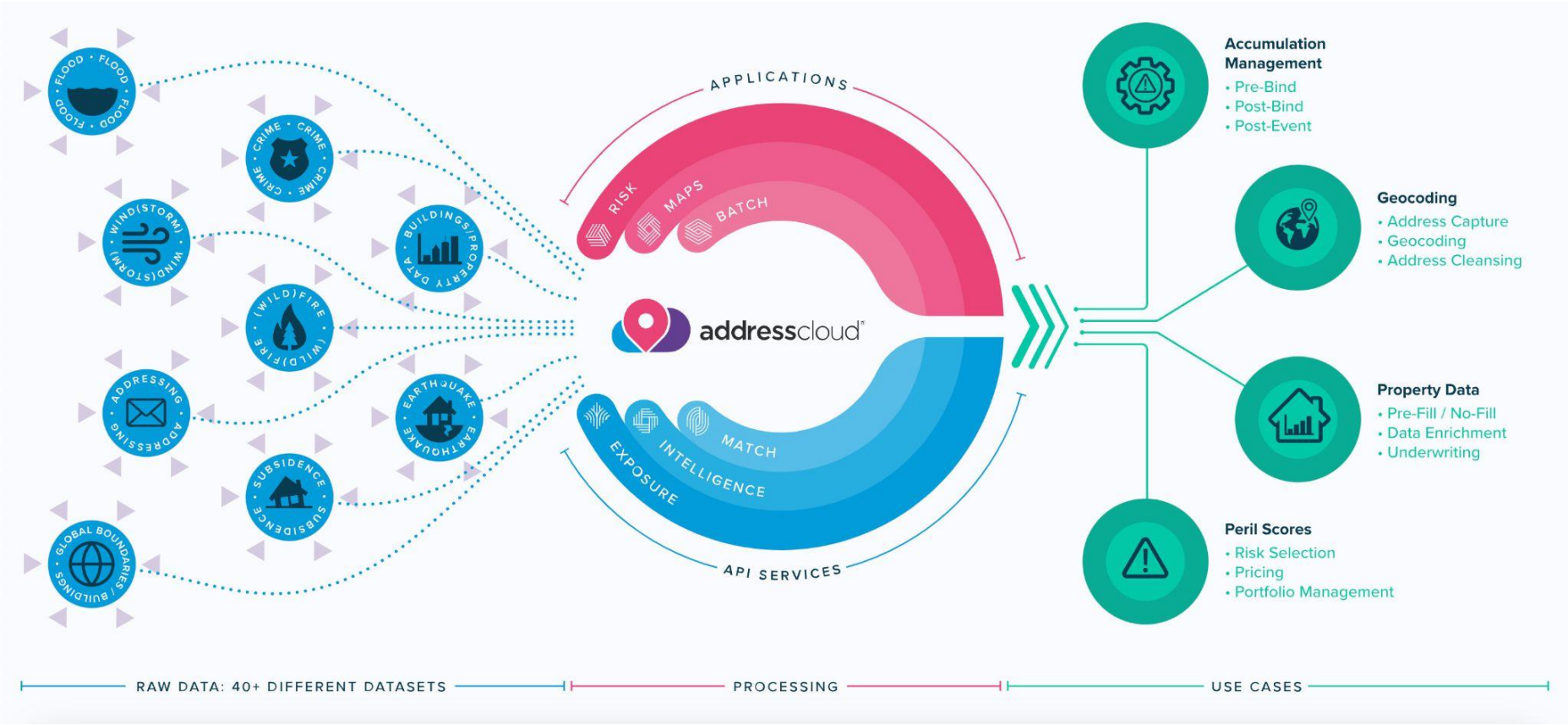
addresscloud®

FOSS4GUK Bristol 2024
Unlocking Overturemaps with DuckDB Spatial

Matt Travis

addresscloud®



© 2022
Addresscloud®

addresscloud®

Scaling to global data

Current methods of extracting buildings data.

- QGIS QuickOSM plugin
- OSMnx
- Download data Geofabrik website > OSMOSIS / pgsql2osm

Problem with the first two is that they are or not scalable as they rely on access to the OSM overpass API

The OSMOSIS/pgsql2osm is problematic due to how slow it can be.

© 2022
Addresscloud®

addresscloud®

Step 1: Extract the OSM Planet file ~144GB in size.

Even with good download speed this can take a while…

© 2022
Addresscloud®

addresscloud®

Step 2: Load into Postgres…(2-3 days)

© 2022
Addresscloud®

OVERTURE MAPS

FOUNDATION

Founded in 2022 under the Linux Foundation Overture is dedicated to the development of reliable, easy-to-use, and interoperable open map data that will power current and next-generation map products

OVERTURE MAPS FOUNDATION

Overture Members

addresscloud®

Components of Overture

1. Use the best sources of open map data

- Crowdsourced
- Government
- AI Generated
- Other

2. Market grade quality and validation

- Conflation
- Deduplication
- Validation

3. Stable, linkable format

- Data Schema
- GERS*

© 2022
Addresscloud®

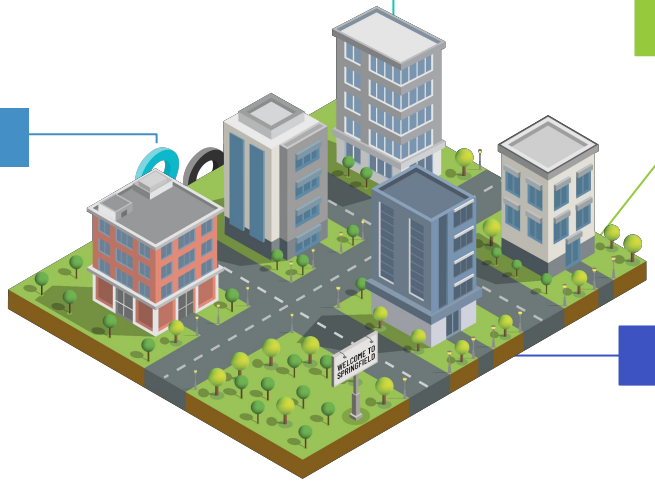
OVERTURE MAPS FOUNDATION

Data Themes

1. Addresses
2. Base (includes Land Cover)
3. Buildings
4. Divisions (Admin Boundaries)
5. Places
6. Transportation

OVERTURE MAPS FOUNDATION

GERS



3D Models

Zoning Data

Tenant Data

Building GERS

Traffic Flow

Speed Limits

ADAS Data

Transportation GERS

Social Signals

PUDO Locations

Operating Hours

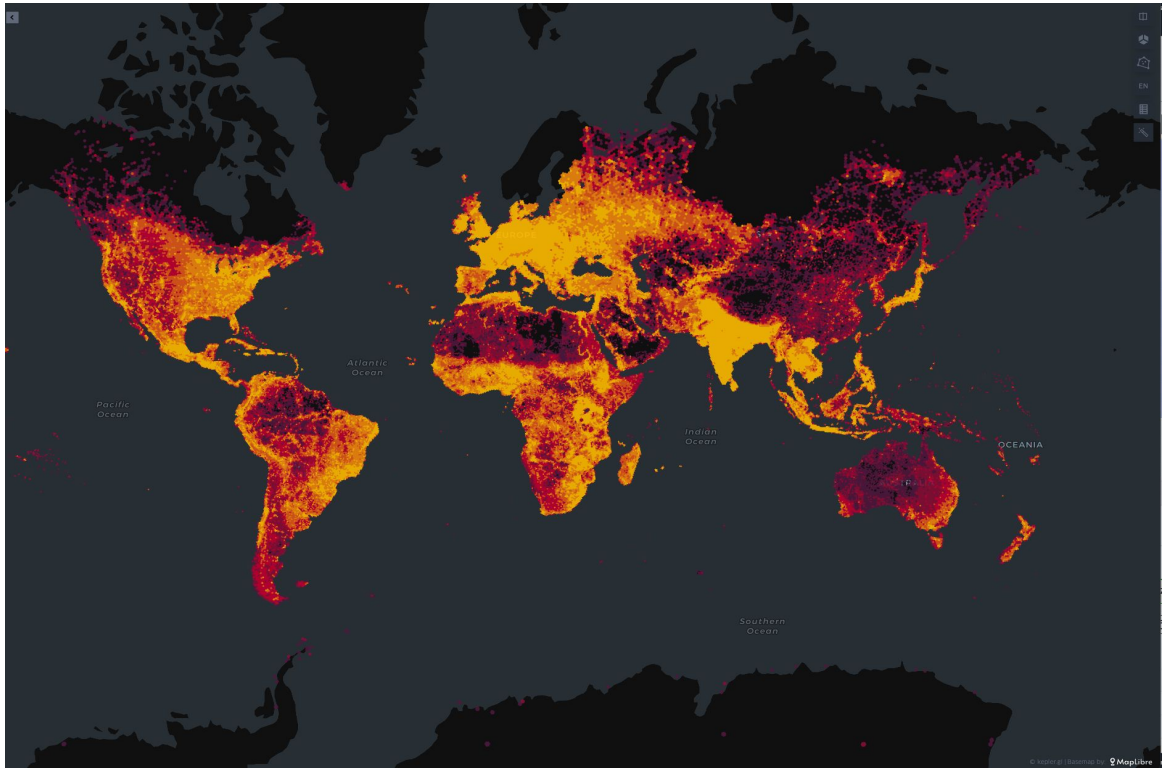
Places GERS

Environmental Data

Demographic Data

Admin GERS

The October release of Overture had over 2.3B buildings



© 2022
Addresscloud®

OVERTURE MAPS FOUNDATION

This is an example of how to download Overture data using the AWS client

```
aws s3 cp --region us-west-2 --no-sign-request --recursive
s3://overturemaps-us-west-2/release/2024-10-23.0/theme=places/type=place/*
```

You can substitute theme/type from the list below.

```
type_theme_map = {
  "address": "addresses",
  "building": "buildings",
  "building_part": "buildings",
  "division": "divisions",
  "division_area": "divisions",
  "division_boundary": "divisions",
  "place": "places",
  "segment": "transportation",
  "connector": "transportation",
  "infrastructure": "base",
  "land": "base",
  "land_cover": "base",
  "land_use": "base",
  "water": "base"
}
```

What is Geoparquet

Why GeoParquet?

- **Standard Geospatial Data in Parquet**

  Following GeoParquet's structure enables interoperability between any system that reads or writes spatial data in Parquet

- **Columnar Data for Geo**

  Data science workflows benefit from columnar data formats, and geospatial analysis can tap into its innovations

- **Cloud Data Warehouse Interoperability**

  Snowflake, BigQuery, RedShift, DataBricks can all work together seamlessly with the same geospatial data format

Who is involved in GeoParquet?

© 2022
Addresscloud®

addresscloud®

DuckDB

Simple: DuckDB is easy to install and deploy. It has zero external dependencies and runs in-process in its host application (eg python/R) or as a single binary (CLI) DuckDB runs on Linux, macOS, Windows

Feature-rich: DuckDB offers a rich SQL dialect. It can read and write file formats such as CSV, Parquet, and JSON, to and from the local file system and remote endpoints such as S3 buckets.

Fast: DuckDB runs analytical queries at blazing speed thanks to its columnar engine, which supports parallel execution and can process larger-than-memory workloads.

Free and Open Source: DuckDB and its core extensions are open-source under the permissive MIT License.

© 2022
Addresscloud®

addresscloud®

DuckDB Spatial

Read/Write

- The spatial extension integrates the GDAL allowing users to read and write data in vector file formats that you would access when using ogr2ogr

GDAL based COPY function

- enables exporting DuckDB tables to different geospatial vector formats through a GDAL based COPY function.

```
COPY ⟨table⟩ TO '.gpkg'
WITH (FORMAT GDAL, DRIVER 'geopackage',
LAYER_CREATION_OPTIONS 'WRITE_BBOX=YES');
```

© 2022
Addresscloud®