

# Web Mapping & Analysis

## Data Backends I - Data Architecture & Formats

Dani Arribas-Bel

# Today

- Spatial Data
- Spatial Data Formats
- Selecting the right format

# Spatial Data

# Traditional Spatial Data

In the “old days” only:

- Mapping agencies (e.g. OS)
- Administrative organizations (e.g. Land Registry)
- Military
- Large companies (e.g. Google)
- ...

could afford the generation of spatial dataset.

# New Spatial Data

Today, this list has **expanded** to essentially **anyone** with access to (cheap) geospatial technology

Overview

Analysis

Premium 

Heart Rate

Est Power Curve

Est 25W  
Distribution

## Dani Arribas-Bel – Ride



12:54 PM on Sunday, October 2, 2016

## Lunch Ride

Add a description

STRAVA LABS  
View Flybys 

40.3 km 1:44:53 223 m

Distance Moving Time Elevation

173 W Estimated Avg Power

1,088 kJ

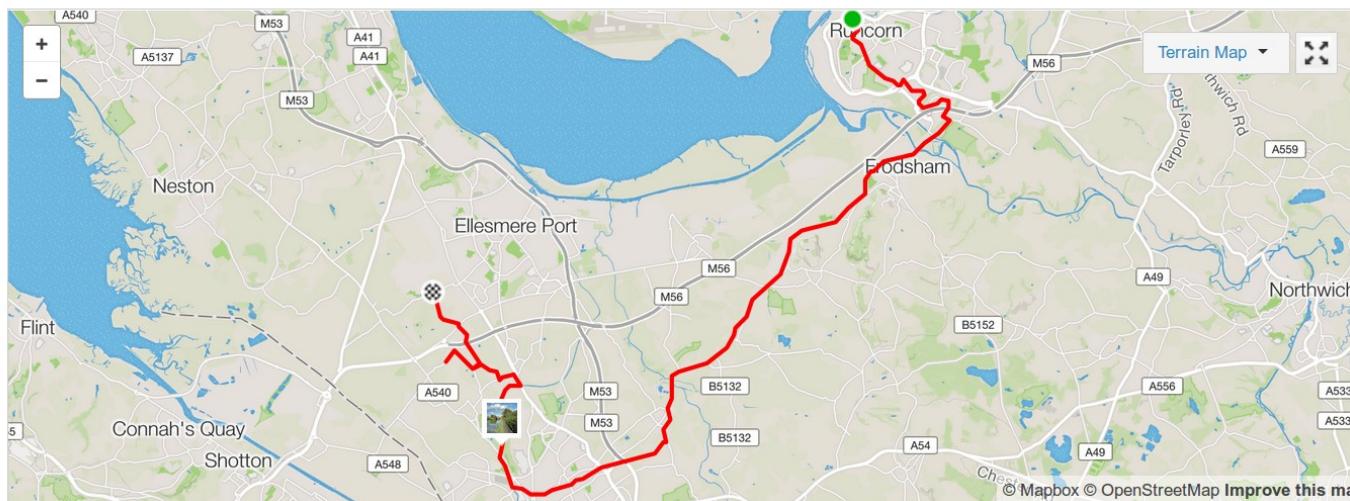
Energy Output

Speed Avg Max Show More

23.1 km/h 46.8 km/h

Elapsed Time 2:08:22

Device: Strava Android App Bike: CrossBeast



## Segments

Learn more about segments

	Name	Time	Speed	Power	VAM	HR
★	Standing start 1.1 km 23m 2%	3:01	23.3 km/h	236 W	—	—
★	RCC to Lights 0.7 km 7m 1%	2:01	21.9 km/h	168 W	—	—
★	Clifton Road 1.3 km 11m -1%	4:08	19.8 km/h	152 W	—	—
★	Coniston Close Climb 0.4 km 5m 0%	1:32	18.0 km/h	114 W	—	—
★	Bike Bible CC - Slalom De Beechwood! 1.3 km 30m -2%	4:15	19.6 km/h	125 W	—	—
★	Down Wood Ln 0.5 km 11m -2%	1:07	28.2 km/h	41 W	—	—
★	Two bridges dash 1.3 km 15m 0%	4:16	19.1 km/h	247 W	—	—

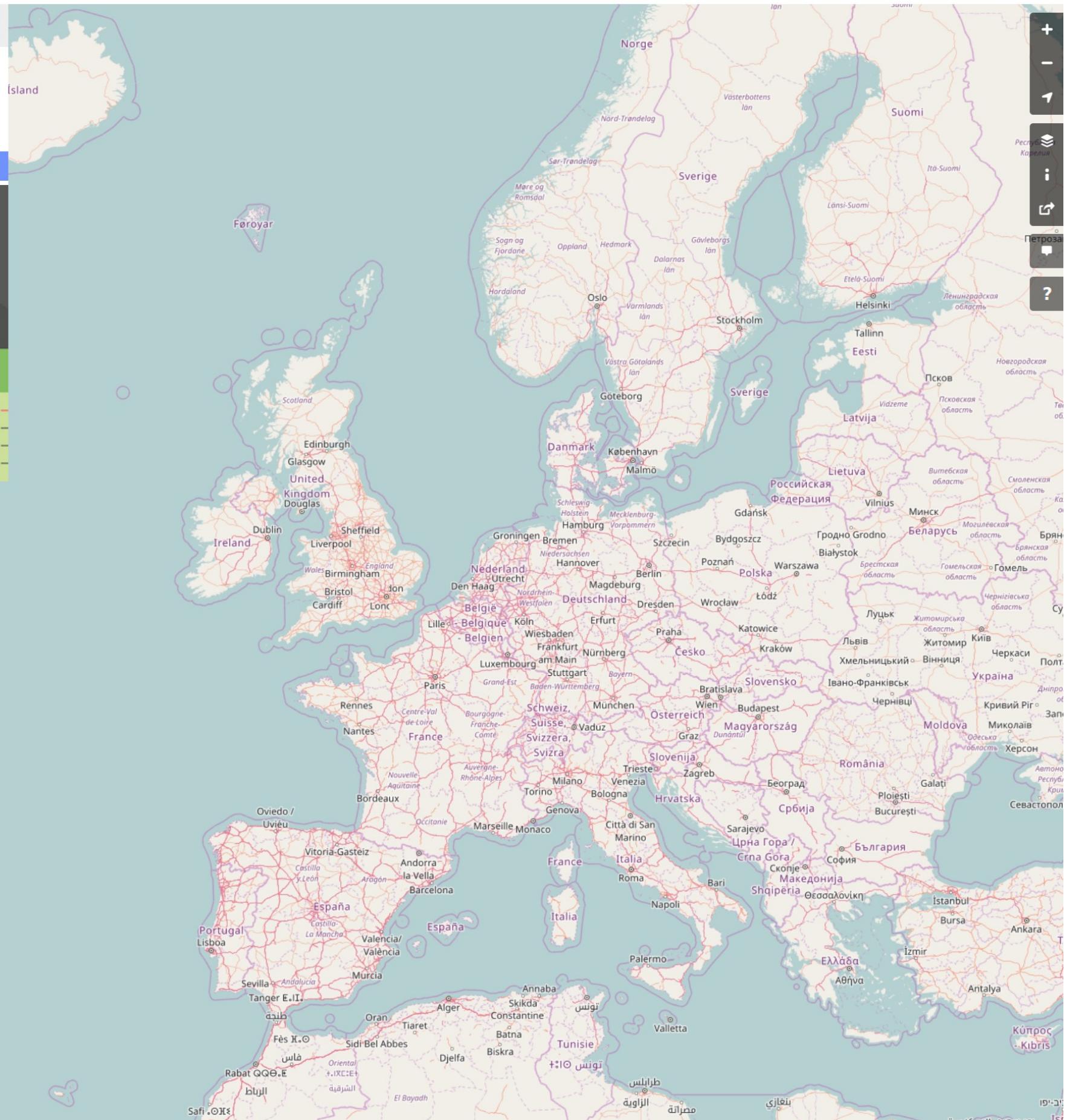
Search Where am I? Go 

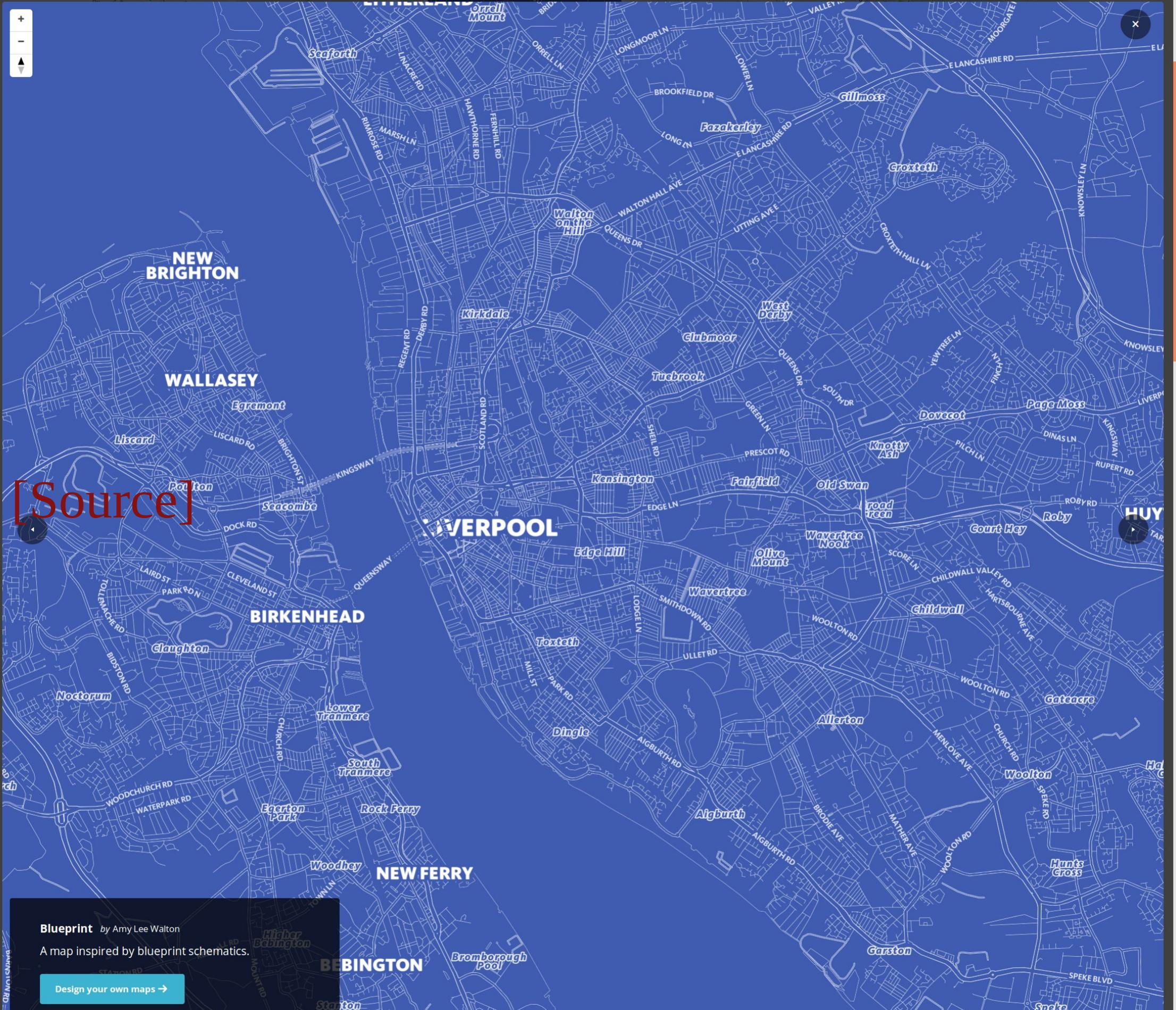
Welcome to OpenStreetMap!

OpenStreetMap is a map of the world, created by people like you and free to use under an open license.

The image features a central graphic consisting of a green puzzle piece shape containing a white circle with a red heart. A magnifying glass is positioned over the heart, with its handle pointing towards the bottom left. The background is dark grey, and the overall design is clean and modern.

# [Source]

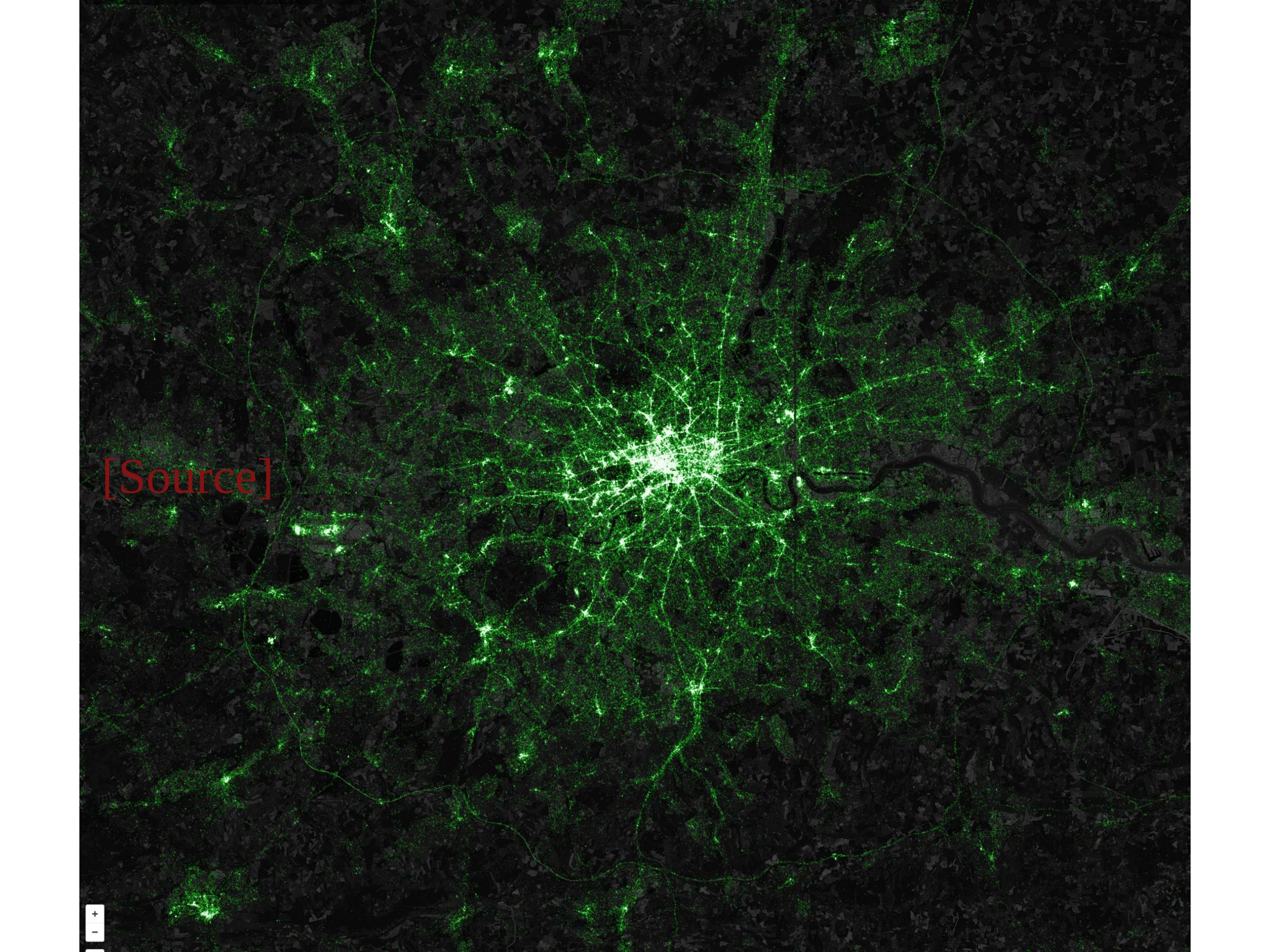




Blueprint by Amy Lee Walton

A map inspired by blueprint schematics.

[Design your own maps →](#)



[Source]





# Data Models

*How we represent the world in a computer*

- Vector
- Raster

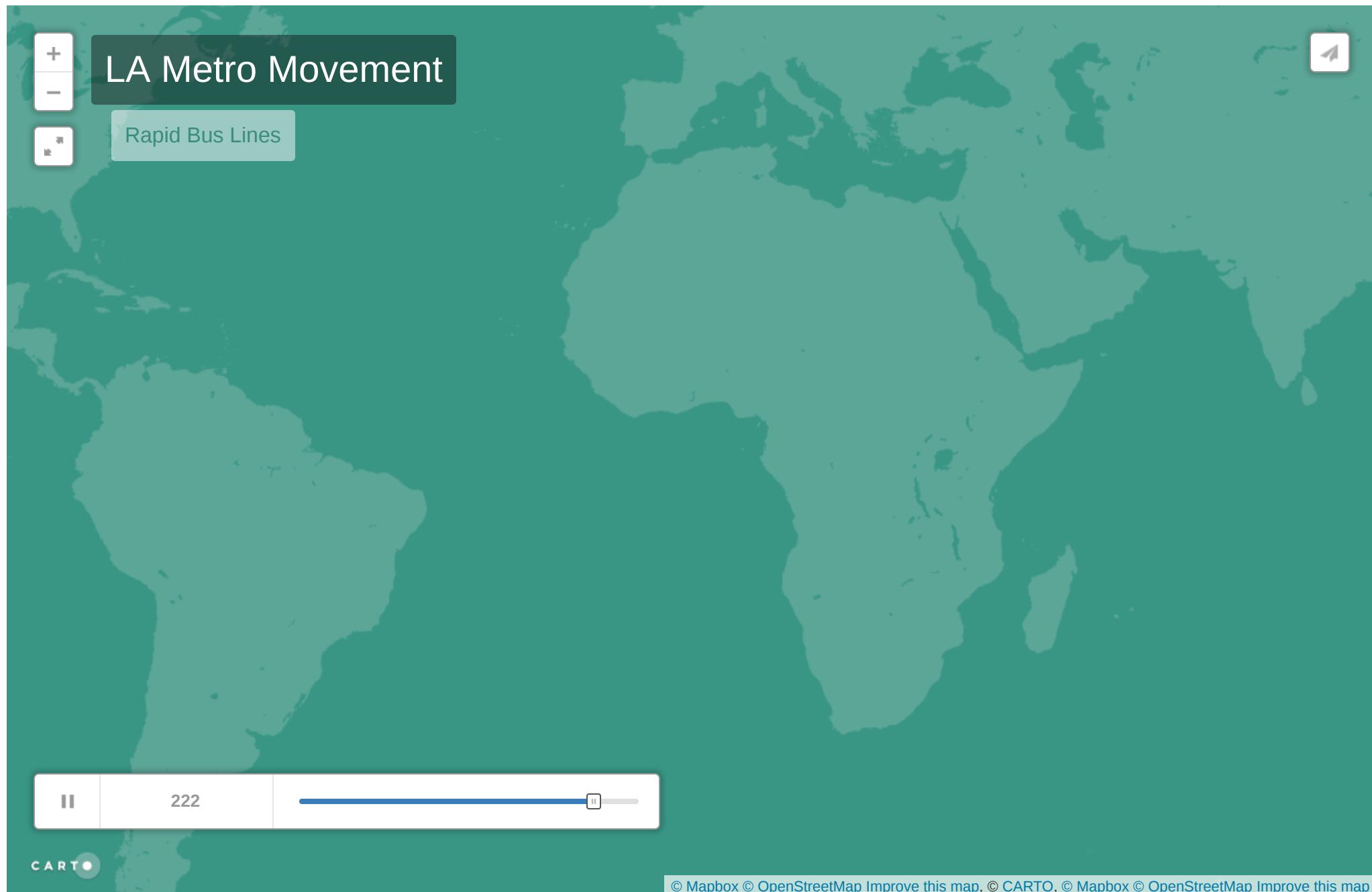
# Vector

Represent each entity with a shape or geometry.

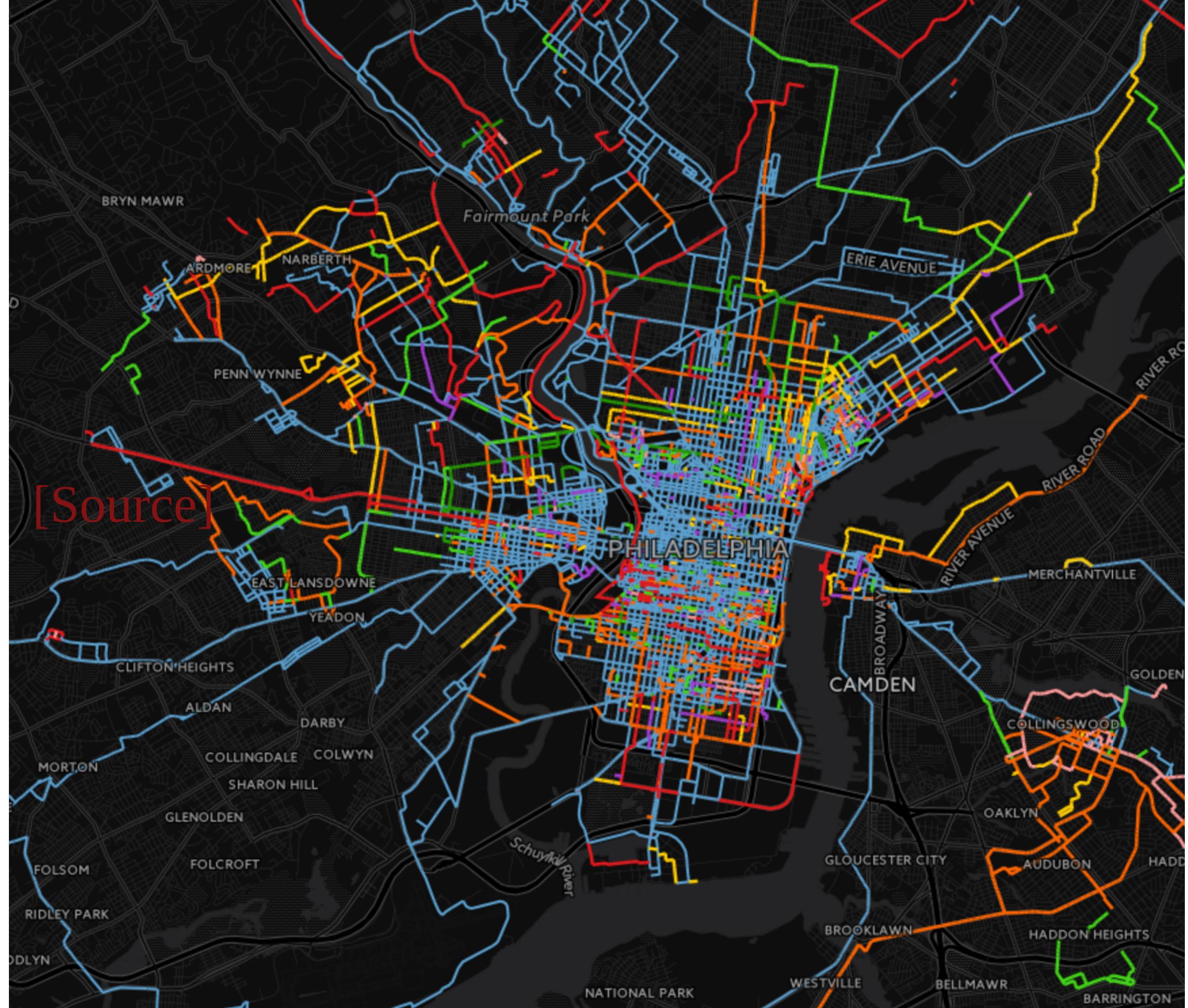
**Simple features:**

- (Multi-)Points
- (Multi-)Lines
- (Multi-)Polygons

Divide space into a **finite** set of entities



Map created by ★ [d9a](#)



[Source]

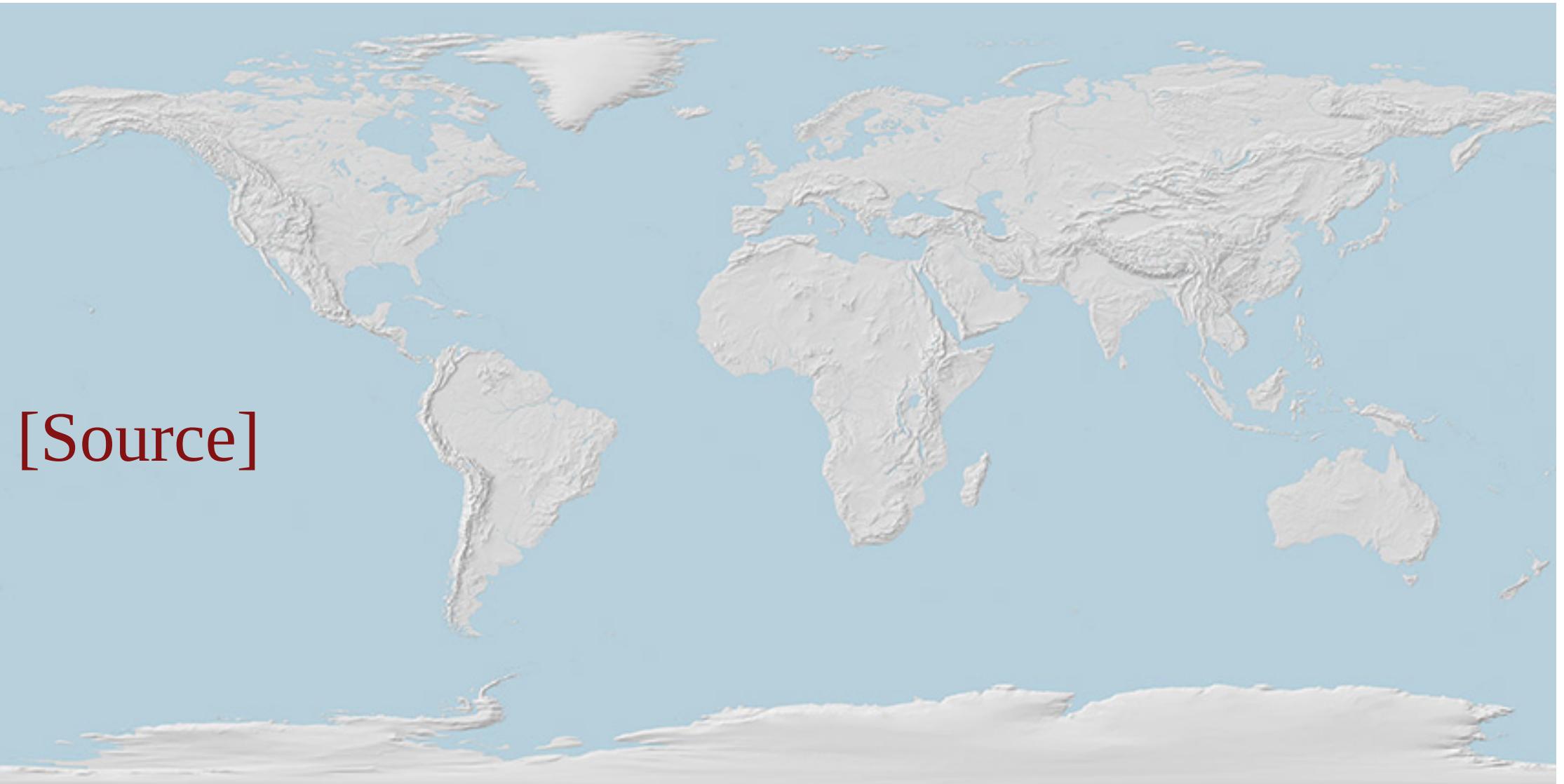


# Raster

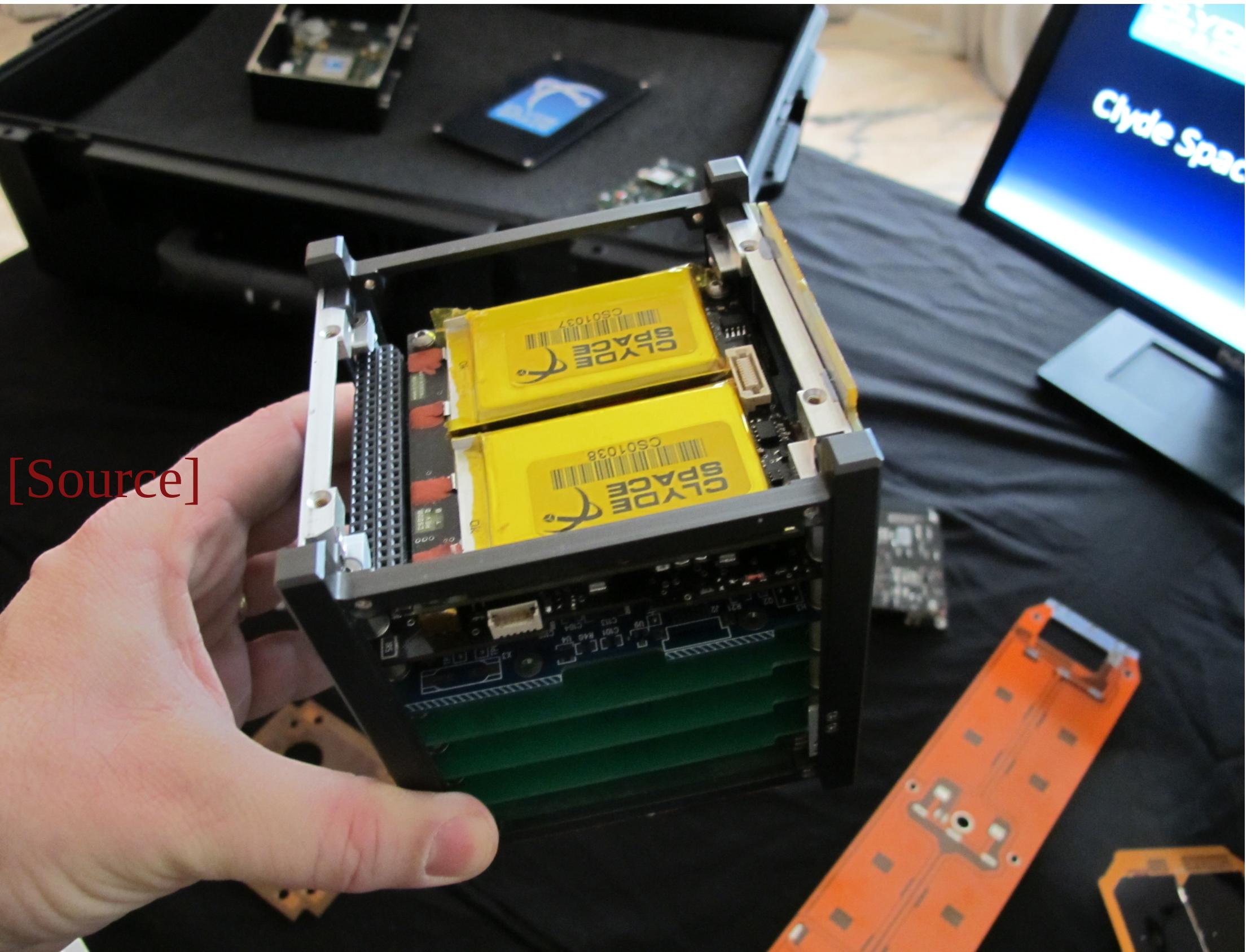
Use an **image** and control pixel colors to encode value

The value assigned for each cell represents the attribute of that cell

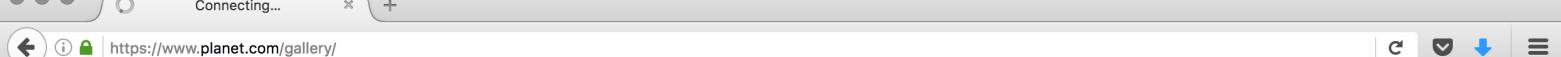
- **Continuous** variables, surfaces (temperature, density, elevation...)
- Satellite images (land cover, land use...)



[Source]



[Source]

Connecting... 

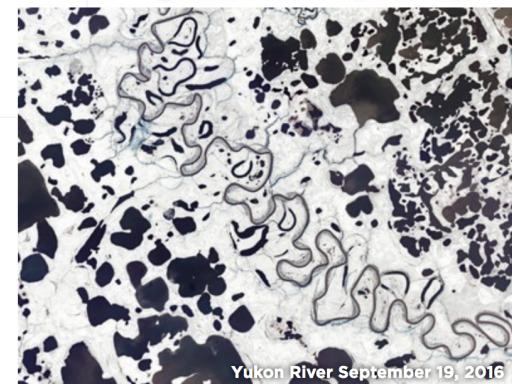
[CONTACT SALES](#) [SIGN UP](#) | [LOG IN](#)

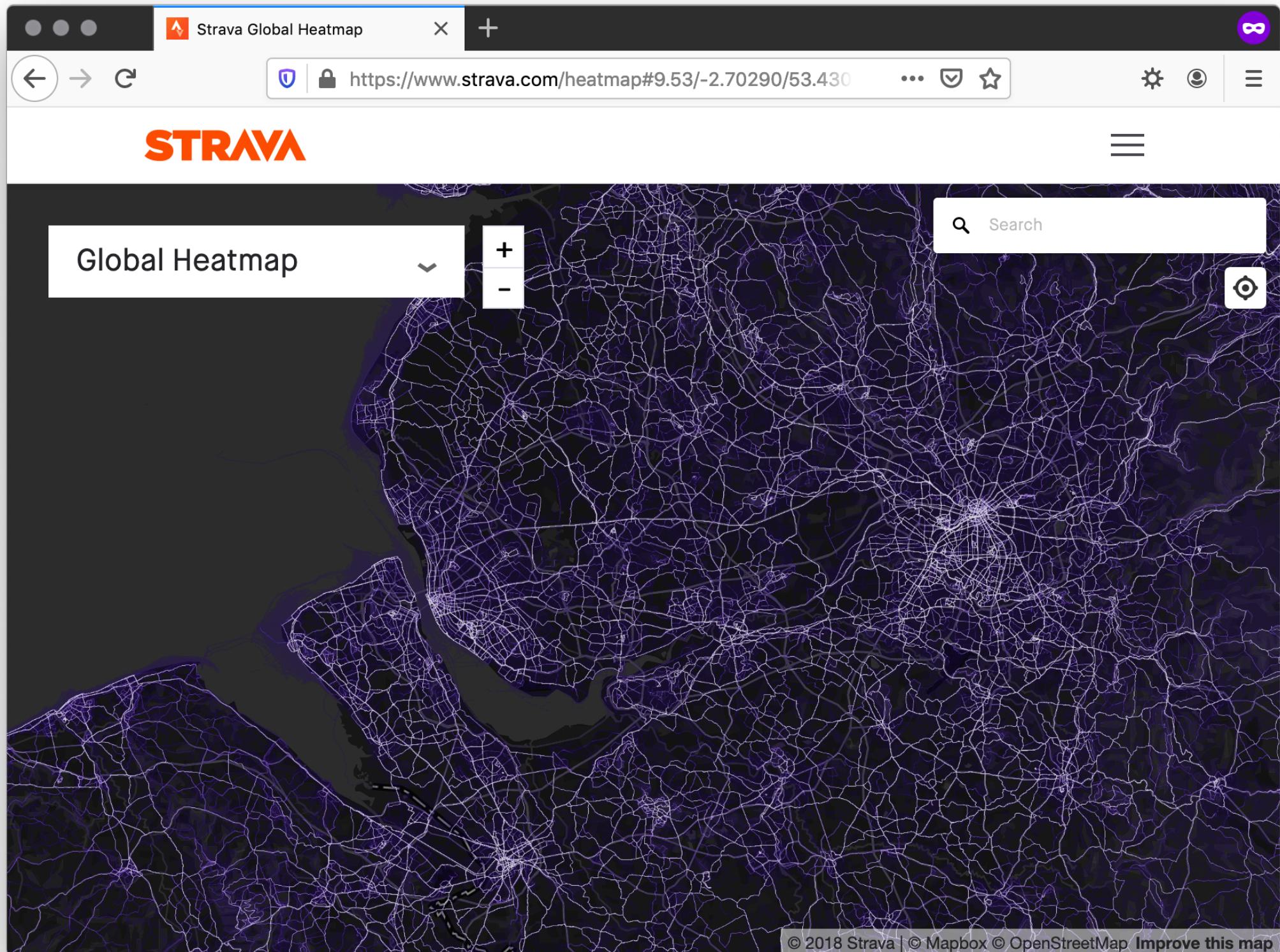
[PRODUCTS](#) [MARKETS](#) [GALLERY](#) [COMPANY](#) [BLOG](#)

## GALLERY



-   
Iguazú National Park September 23, 2016
-   
Büyükk Tokaç September 20, 2016

  
Yukon River September 19, 2016



# Spatial Data Formats

# Spatial Data Formats

In principle...

- Points, lines, polygons → Vector formats
- Images, surfaces → Raster formats

But these boundaries are blur...

Global Human Settlement - Hon X

https://ghsl.jrc.ec.europa.eu

Home Cookies  
This site uses cookies to offer you a better browsing experience. Find out more on [how we use cookies and how you can change your settings.](#)

I accept cookies I refuse cookies

Legal notice | Cookies | Contact | Search English (en)

European Commission  
Global Human Settlement

European Commission > EU Science Hub > GHSL

Home About Copernicus GEO Documents Atlases Applications Global Definition Data Tools Visualisation News

**GHSL - Global Human Settlement Layer**

**A new open and free tool for assessing the human presence on the planet**

- Produces new global spatial information, evidence-based analytics and knowledge describing the human presence on the planet
- Operates in an open and free data and methods access policy (open input, open method, open output)
- Supported by the Joint Research Centre (JRC) and the DG for Regional Development (DG REGIO) of the European Commission, together with the international partnership [GEO Human Planet Initiative](#) GEO GROUP ON EARTH OBSERVATIONS

09/10/2019 [Mapping European Cities and Settlements from Space with Copernicus data](#) - European Week of Regions and Cities - Building SQUARE - Brussels Convention Centre, Room 314+316

1975

Click here to find out about the technology behind this image

1 2 3 4 5 6

Last update: 04/07/2019 | [Top](#) | Legal notice | Cookies | Contact | Search

LUCAS - Land use and land cov X

https://ec.europa.eu/eurostat/statistics-explained/index.php/LUCAS\_-\_Land\_use\_and\_land\_cover

eurostat  
Statistics Explained

NAVIGATION ▾ ONLINE PUBLICATIONS ▾ TOOLS ▾

## LUCAS - Land use and land cover survey

No update is planned for this article

The European Union (EU) is composed of a diverse range of landscapes: it is home to a wide variety of flora and fauna and includes some of the most and least densely populated areas of the world. This background article provides information on the Land Use/Cover Area frame Survey (LUCAS), a survey that provides harmonised and comparable statistics on land use and land cover across the whole of the EU's territory.

The data collected by LUCAS provides harmonised information for studying a range of socioenvironmental challenges, such as [land take](#), soil degradation or biodiversity.

Full article + Direct access to

- Other articles
- Tables
- Database
- Dedicated section
- Publications
- Methodology
- Legislation
- Visualisations
- External links

### Notes

- ↑ <https://www.sciencedirect.com/science/article/pii/S1462901115300654> The new assessment of soil loss by water erosion in Europe
- ↑ <http://onlinelibrary.wiley.com/doi/10.1111/ejss.12499/full> LUCAS Soil, the largest expandable soil dataset for Europe: a review
- ↑ <https://ec.europa.eu/jrc/en/publication/lucas-2018-soil-component-sampling-instructions-surveyors> LUCAS 2018 - soil component: Sampling Instructions for Surveyors

Categories: [Agriculture](#) | [Agriculture background articles](#) | [Background article](#) | [Environment](#) | [Regions and cities typologies](#)

Page ISSN 2443-8219

This page was last modified on 6 March 2019, at 10:22. 2 watching users Privacy policy About Statistics Explained Disclaimers

# Traditionally

## Vector

- Single files: shapefiles, etc.
- Client-server (geo-)DBs: PostGIS, etc.

## Raster

- Mostly (single) image formats: GeoTIFF, etc.

# However...

Many of these formats were designed for an *offline* world, so display some of the following:

- Binary (“non-streamable”)
- “Unqueriable”
- Complex format structures

Shapefile must die! switchfromshapefile.org

# Switch from Shapefile

ESRI Shapefile is a [file format for storing geospatial vector data](#). It has been around since the early 1990s and is still the most commonly used vector data exchange format.

While Shapefiles have enabled many successful activities over the years, they also have a number of limitations that complicate software development and reduce efficiency.

We, members of the geospatial IT industry, believe that it is time to stop using Shapefiles as the primary vector data exchange format and to replace them with a format that takes advantage of the huge advances that have been made since Shapefile was introduced.

[Fork me on GitHub](#)

**Read more:**

- [The good side](#)
- [Shapefile is a bad format](#)
- [Shapefile alternatives](#)

---

## The good side

Shapefile does a lot of things right. Here are some reasons why Shapefile is so heavily used:

- Shapefile is by far the most widely supported format in existing software packages.
- While the format is proprietary, the [specification is open](#).
- For many use cases, it is *good enough*.
  - Index files (e.g. \*.shx) enable good reading performance.
  - It is relatively efficient in terms of file size. The resulting file, even un-zipped, is relatively small compared to some other (mostly text-based) formats.

## Shapefile is a bad format

Why is Shapefile so bad? Here are several reasons why the Shapefile is a bad format and you should avoid its usage:

- No coordinate reference system definition.
- It's a multifile format.
- Attribute names are limited to 10 characters.
- Only 255 attributes. The DBF file does not allow you to store more than 255 attribute fields.
- Limited data types. Data types are limited to float, integer, date and text with a maximum 254 characters.
- Unknown character set. There is no way to specify the character set used in the database.
- It's limited to 2GB of file size. Although some tools are able to surpass this limit, they can never exceed 4GB of data.
- No topology in the data. There is no way to describe topological relations in the format.
- Single geometry type per file. There is no way to save mixed geometry features.
- More complicated data structures are impossible to save. It's a "flat table" format.
- There is no way to store 3D data with textures or appearances such as material definitions. There is also no way to store solids or parametric objects.
- Projections definition. They are incompatible or missing.
- Line and polygon geometry type, single or multipart, cannot be reliably determined at the layer level, it must be determined at the individual feature level.
- Add more ...

# Modern formats

New formats have appeared in part “*fixing*” those issues, but also responding to web needs:

- Streamable (e.g. **GeoJSON**)
- Queriable (e.g. **PostGIS/Geopackage**)
- Single file (e.g. **.mbtiles**)

Let’s explore a bit more on a couple of them...

# GeoJSON



# GEOJSON

GeoJSON is a format for encoding a variety of geographic data structures.

```
{  
  "type": "Feature",  
  "geometry": {  
    "type": "Point",  
    "coordinates": [125.6, 10.1]  
  },  
  "properties": {  
    "name": "Dinagat Islands"  
  }  
}
```

GeoJSON supports the following geometry types: `Point`, `LineString`, `Polygon`, `MultiPoint`, `MultiLineString`, and `MultiPolygon`. Geometric objects with additional properties are `Feature` objects. Sets of features are contained by `FeatureCollection` objects.

## The GeoJSON Specification (RFC 7946)

In 2015, the Internet Engineering Task Force (IETF), in conjunction with the original specification authors, formed a [GeoJSON WG](#) to standardize GeoJSON. [RFC 7946](#) was published in August 2016 and is the new standard specification of the GeoJSON format, replacing the 2008 GeoJSON specification.

# Advantages

- Plain text, (human) readable
- Streamable
- Well integrated with web standards (JSON)

Excellent exchange format

# Disadvantages

- Plain text, inefficient
- Non-queriable
- Vector only

Tilesets (.mb/vtiles)

The screenshot shows a web browser window displaying the [Mapbox Docs](https://docs.mapbox.com) website. The URL in the address bar is <https://docs.mapbox.com>. The page title is "MBTiles". The main content area contains text about the MBTiles file format, its purpose, and its relationship to SQLite databases and tilesets. It also lists related resources, including links to the MBTiles specification and the Studio Manual geospatial data page. A search bar at the top is empty, and there is a "Sign in" button in the top right corner.

MBTiles | Glossary | Help | Map

Docs

Help **Glossary**

Search docs.mapbox.com

MBTiles

# MBTiles

**MBTiles** is a file format for storing [tilesets](#). It's designed to allow you to package up many files into a single tileset. For example, `mapbox.mapbox-streets-v8` is a single tileset that contains administrative boundaries, road networks, POIs, and other kinds of geospatial information from many different data sources.

MBTiles is an open specification based on the [SQLite](#) database. MBTiles can contain raster or vector [tilesets](#). You can upload MBTiles files directly to [Mapbox Studio](#) or use them directly in a web or mobile application.

**Related resources:**

- [MBTiles specification](#)
- Studio Manual [geospatial data page](#): transfer limits for uploading MBTiles files as tilesets

Sign in

MBTiles

Search docs.mapbox.com

MBTiles

# MBTiles

**MBTiles** is a file format for storing [tilesets](#). It's designed to allow you to package up many files into a single tileset. For example, `mapbox.mapbox-streets-v8` is a single tileset that contains administrative boundaries, road networks, POIs, and other kinds of geospatial information from many different data sources.

MBTiles is an open specification based on the [SQLite](#) database. MBTiles can contain raster or vector [tilesets](#). You can upload MBTiles files directly to [Mapbox Studio](#) or use them directly in a web or mobile application.

**Related resources:**

- [MBTiles specification](#)
- Studio Manual [geospatial data page](#): transfer limits for uploading MBTiles files as tilesets

# Map Tiles



**How does a zoomable map work?**

People are using coordinate systems and map projections to transform the shape of Earth into usable flat maps for centuries.

A map of the entire world is too big to be directly displayed in a computer so there is a clever mechanism for quick browsing and zooming on maps: the map tiles.

The world is divided into small squares, each with fixed geographic area and scale. This clever trick allows you to browse just a small part of the planet without loading the whole map - and you still get an illusion of exploring a single huge document.

Tiles à la Google Maps: Coordin X +

Tiles à la Google Maps: Coordin X +

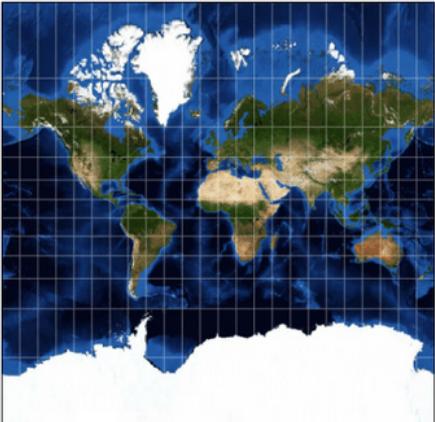
https://www.maptiler.com/google-maps-coordinates-tile-bounds-projection/



**Degrees**  
Geodetic coordinates  
WGS84 (EPSG:4326)

**Longitude and latitude**  
coordinates used by GPS devices for defining position on Earth using World Geodetic System defined in 1984 (WGS84).

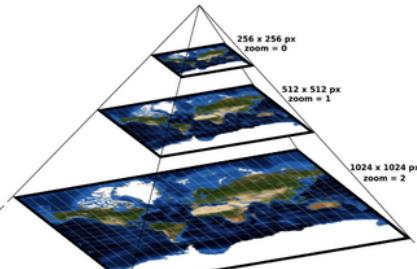
*HINT: WGS84 geodetic datum specify lon/lat ( $\lambda/\phi$ ) coordinates on defined ellipsoid shape with defined origin ([0,0] on a prime meridian).*



**Meters**  
Projected coordinates  
Spherical Mercator  
(EPSG:3857)

**Global projected coordinates** in meters for entire planet. Used for raster tile generation in GIS and WM(T)S services.

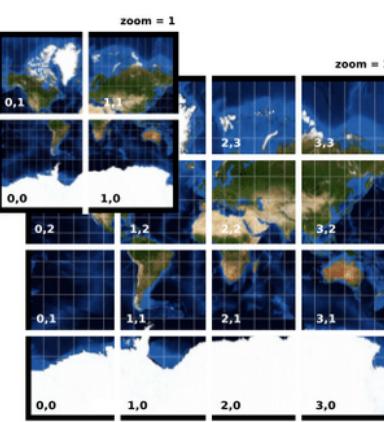
*HINT: Simpler spherical calculation are used instead of ellipsoidal. Mercator map projection deforms size (Greenland vs Africa) and never shows poles.*



**Pixels**  
Screen coordinates  
XY pixels at zoom

Zoom-specific **pixel coordinates** for each level of the pyramid. Top level (zoom=0) has usually 256x256 pixels, next level 512x512, etc.

*Devices calculate pixel coordinates at defined zoom level and determine visible viewport for area which should be loaded from servers.*



**Tiles**  
Tile coordinates  
Tile Map Service (ZXY)

Coordinates of a **tile in the pyramid**. There is one tile on the top of the pyramid, than 4 tiles, 16 tiles, etc. All tiles have the same size, usually 256x256 pixels.

*Only the relevant tiles loaded and displayed for the area of interest / viewport.*

# Advantages

- Queriable (SQLite)
- Fast access to large maps with limited resources  
(client/server model + queriable format)
- Some (vector tiles) are stylable

# Disadvantages

- Designed for *serving* not *analysing*
- A dataset needs to be stored at several zoom levels
- Once created, hard to modify (e.g. reproject)

# Selecting the right format

# Selecting the right format

No silver bullet...

- *What* type of data do you want to store? Vector, raster
- *What* are you going to do with the file? Analysis, serving
- *What* environment are you working? Locally, web

# Quiz

1. Large dataset of tweets you want to analyse

**PostGIS/Geopackage**

2. Drone imagery to make available for workshop

participants **MBTiles**

3. Street basemap to provide context to a small

dataset you want to make available on the web

**Vector tiles**

4. The small dataset from 3. **GeoJSON**



Web Mapping & Analysis by Dani Arribas-Bel is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.