Engineer

Geospatial

data whose meaning depends on where it is

What it is

Geospatial describes data, analysis and software in which a position on the Earth is part of the meaning. A temperature reading is only a number until it is tied to the place it was taken. A smart-meter reading becomes useful for planning once it can be placed on a feeder, inside a district or near a weather station. The word is an adjective, and it travels with whatever it qualifies: geospatial data, a geospatial join, geospatial modelling.

Spatial on its own is the wider word. A chip layout, a warehouse floor plan measured in local metres and a mechanical drawing are all spatial, and their coordinates mean nothing outside the drawing. Geospatial data is pinned to the planet through a coordinate reference system, so two datasets that were never designed together can still be laid over one another and compared.

Location reaches a dataset in several forms, and they are not equally useful.

- **Coordinates**, usually latitude and longitude from a GPS receiver or a geocoder, give a point that software can compute with directly.
- **Geometries** describe lines and areas as well as points: a cable route, a flood extent, a service territory. The Simple Features model from the Open Geospatial Consortium (OGC), also published as ISO 19125, defines the shapes that most databases and file formats share.
- **Addresses and place names** carry location for people, but they have to be geocoded, turned into coordinates, before anything can be measured, and each geocoder resolves them to its own precision.
- **Area codes**, such as postcodes, municipality numbers or grid-cell identifiers, say which area something lies in without saying where inside it.

Underneath sits a set of open standards that let geospatial data move between systems without losing its meaning. The OGC publishes the specifications for geometry, for files such as GeoPackage and for web services such as OGC API – Features. The EPSG registry gives each coordinate reference system a numeric code, so that 4326 means latitude and longitude on WGS 84 wherever it appears. PROJ, an open-source library, transforms coordinates from one of those systems to another and supports more than a hundred map projections.

Use the concept in a decision

Put “What does geospatial mean?” to work in a real decision. We can help relate the concept to the system you are building or buying. Anything that happens in a place — vessels, vehicles, plots, sensors — put on a map you can search, measure and share.

Explore our services Get in touch →

Test the assumptions behind the term

Use the explanation and linked work to frame questions for your project. We can examine the assumptions with you before a tool is chosen. Made a geographical data pipeline 50x faster — 50x faster geographical data pipelines

Made a geographical data pipeline 50x faster Get in touch →

Why it matters

Location is a join key that almost every dataset already has. Meter readings, weather observations, customer records and network assets rarely share an identifier, yet each of them happened somewhere. A geospatial join pairs rows by how their positions relate, one inside another, within a given distance or crossing, instead of by equal values. It connects data that no foreign key ever linked, and that is often the first reason an organisation starts treating its data as geospatial: the question it wants answered spans datasets that only meet on the map.

The second reason is that location changes what an average means. A national figure for consumption or for supply interruptions can hide one district that is at its limit while the rest have room to spare, and the decision that matters, where to reinforce, where to invest, where to send a crew, is about that district. Grouping by area or by distance shows the spread that a single total hides.

Location also carries weight beyond analysis. A short trail of GPS points from one phone usually shows where its owner sleeps and works, so location data about people deserves the care given to any other personal data. In practice that means collecting it at the precision the purpose needs, coarsening it before it is shared, and keeping it no longer than the purpose lasts.

What it looks like in practice

Smart meters spread across several regions produced the data behind more than 500 analysis reports written in pure SQL on PostgreSQL and BigQuery. The readings came with geolocation coordinates for each meter, which let the scripts use geospatial joins alongside CTEs and window functions to find regional trends and anomalies. Every report carried its annotated SQL so that colleagues could reproduce it, and the common data quality problems, missing GPS coordinates among them, were written down with a recommended way to handle each. Tied to their regions by those joins, the findings went on to aid regional load balancing, and the reports had a hand in the 10–15% rise in operational planning efficiency that followed.

Eleven electricity grid operators had the operational data but little geospatial analysis to base their reinforcement and investment decisions on. Their operational data was combined with geospatial modelling, so that each network could be examined on the map and asked which stretches were heading towards their limits and which areas justified investment first. The analysis was fitted to how each operator actually ran its network rather than handed out as one template.

Geospatial data also has to stay fast as it grows. In a PostgreSQL application where location queries had slowed down together with the electricity time series, GiST indexes were added for the geospatial data, and the geographical data was hash-partitioned to spread the load evenly across nodes. Geospatial lookups then ran in milliseconds.

Where it goes wrong

Unknown positions pile up at zero latitude and zero longitude. A map of customers or devices shows a dense cluster in the Gulf of Guinea, where the equator crosses the prime meridian, and regional totals include records from a place no customer can be. A device without a GPS fix, or an import that met an empty cell, wrote zeros instead of leaving the position empty. Store an unknown position as NULL, add a constraint that rejects the exact pair 0, 0, a spot in open sea, and count the missing positions in every batch so that each report states how many records it left unplaced.

Joins on coordinates find almost nothing. Two datasets describe the same substations, yet joining them on equal latitude and longitude matches only a handful of rows. Each source measured or rounded its coordinates differently, and coordinates are floating-point numbers, so two positions a few centimetres apart are simply different values. Join by distance instead, with a tolerance chosen from how accurate the sources are, through a predicate the spatial index can serve. In PostGIS, with geography columns measured in metres:

```
SELECT a.substation_id, b.asset_ref
FROM survey AS a
JOIN register AS b
  ON ST_DWithin(a.location, b.location, 15);
```

Then review rows that matched more than once, because a tolerance wide enough to absorb measurement error can also bridge two neighbouring assets.

Points at postcode centres pass for exact positions. A distance analysis reports dozens of customers living at exactly the same spot, or a heat map shows sharp peaks in the middle of postal areas. The addresses were geocoded only as far as the postcode, and each came back as the centre of its area without the analysis noticing. Record the level each point was resolved to, such as address, street or postcode, in a column beside it, and filter or weight by that level before measuring any distance shorter than the area itself.

Last year’s regional totals change after a boundary update. A report rerun on unchanged historical data gives different figures per region from the ones first published. The points were joined to the current boundary layer, and since then municipalities have merged, postcodes have split or a service area has been redrawn. Keep the raw coordinates rather than only the region they fell in, store boundary layers with the dates they were valid, join each record to the boundary in force at its timestamp, and note in every report which boundary version produced it.

Further reading

- OGC Standards — the Open Geospatial Consortium’s catalogue of open geospatial standards, each with its specification documents.
- Simple Feature Access – Part 1 — the common geometry model of points, lines and polygons that databases and formats share.

- EPSG Geodetic Parameter Dataset — the registry behind codes such as 4326 and 3857, with the full definition of each reference system.
- PROJ — the open-source library that transforms coordinates between reference systems, with its command-line tools and API.
- OGC API – Features — the OGC standard for publishing and querying geospatial features over HTTP.

Glossary

Choose for your actual constraints

Describe the workflow, the existing interfaces and the users who need the change. We can define a testable delivery and a handover that lets your team maintain it.

See what it costs Get in touch →

Make the terminology practical

Ask how “What does geospatial mean?” applies to your organisation. A concrete example from your current system is enough to start the discussion.

Explore our services

welcome@engineer.company +45 50 10 61 84 Copenhagen, Denmark

Take this page with you

Contact card

Carefully Crafted With Love by D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

Copyright © 2025 – present · Engineer ApS · All rights reserved.

Printed from https://engineer.company/glossary/geospatial/