



“The future of data is more open” – How Overture standardizes the use of geospatial data

Jennings Anderson

Meta Platforms, Inc.

February 21, 2025

 10.6094/grid.2025.001

This conversation with Jennings Anderson explores the evolving landscape of open geospatial data, focusing on ethical considerations, corporate involvement, and digital transformation. As a Software Engineer at Meta, he provides insights into the projects of the Overture Maps Foundation, highlighting potential impact on developers and map creators.

In the halls of Friedrich-Alexander Universität Erlangen-Nürnberg, Germany, we debated the best Nuremberg gingerbread and weekend museum visits, our chatter mingling with echoes of the new students' academic excitement. Our exchange with Jennings Anderson, the guest speaker at the Geospatial Worlds Conference, was seamlessly integrated with our shared enthusiasm for exploring the intricacies of digital transformation, the ethical implications of data ownership and licensing, and the role of corporations in aligning with open data principles.

Jennings Anderson is currently a Software Engineer at Meta (Facebook's parent company). His research focuses on geospatial data analysis, with a particular emphasis on the community project OpenStreetMap (OSM) and disaster mapping, where he has employed and researched the power of geographic information to support humanitarian efforts (Anderson et al. 2018; Anderson, Sarkar, and Palen 2019; Herfort et al. 2021; Palen et al. 2015). His work is currently focused on a new open geospatial data initiative, the Overture Maps Foundation. This initiative, which is also a key topic in this interview, was launched in December 2022 and aims at providing open geospatial data

for developers and map creators. We wanted to have a conversation with Jennings to gain a deeper insight into his expertise on the latest developments in open geospatial data.

Can you briefly say what the Overture Maps Foundation is?

I think the Overture Maps Foundation website describes it best, so I'll paraphrase: The mission is to create open, interoperable map data by combining and pulling open map data from the best available sources. It's specifically designed for developers building map services to use open data more easily.

How is Overture Maps Foundation organized, and is it similar to OpenStreetMap or Google Maps?

The Overture Maps Foundation is an open data project within the Linux Foundation. The founding steering members were Meta, Amazon, Microsoft, and TomTom, with Esri as a general member. Today there are also a number of other contributor members. Google Maps is a commercial mapping product and OpenStreetMap is a community-built open mapping project driven by individual mappers. Overture is an open data project led by its member companies. It combines data from multiple sources, including OpenStreetMap. It's not geared towards individual mappers, but rather on pooling contributions from larger organizations to create a comprehensive, interoperable dataset.

The literal meaning of overture is “an introduction to something more substantial” or “a piece at the beginning of an opera”. Is Overture an introduction to something bigger?

That's a great question! As you said, the word “overture” suggests a beginning, and I think that fits here. It's about doing what hasn't been done before: combining various open data sets into a consistent schema, which could lead to a new ecosystem of map services and tools built around a high-quality base map of open data. It's the start of a new era of map tooling, where data is easier to use and more interoperable.

What are the possibilities that can emerge from Overture? From your perspective, what excites you the most?

What excites me most about Overture is the potential to pool resources around open mapping and create a 'flywheel effect' with open data. As Overture grows, it encourages more data contributions, leading to the development of increasingly higher-quality datasets. This growth not only attracts more contributors but also prompts organizations, local governments, and others to open up their data and contribute to Overture,

either by directly including it or referencing it through the global entity reference system.

The exciting part is the potential for future growth. As the project evolves, we have an opportunity to address the current challenges of fragmented geospatial data. Rather than dealing with millions of disparate datasets that are difficult to work with, Overture could help create more interoperable, standardized datasets. That's the vision I'm excited about—bringing together diverse data sources into a more cohesive and accessible ecosystem.

You talked about the steering committee and the members of Overture, but then there's also a data infrastructure that Overture data is built upon. Can you elaborate on that?

Overture organizes its data into six themes: transportation, buildings, addresses, places, administrative boundaries, and the base layer (land, water, and other essential features). The data comes from a variety of sources, including OpenStreetMap, Microsoft, Meta, and Google's open building datasets. This allows us to offer more complete datasets, like 2.3 billion buildings, combining OpenStreetMap with machine learning-derived data. But if you just want the OpenStreetMap buildings, you could filter for just those buildings, because you might trust them to be a different level of quality, whereas the other buildings could be added in there, filling in the map where OpenStreetMap doesn't have coverage. Another important theme is the places theme, whose data is coming from a combination of Meta and Microsoft sources right now. This is information from locations from public Facebook pages and similar sources. This newly produced dataset hasn't been available before.

For people who are new to Overture Maps, how can they access the data, and for what use cases is it suitable? For which actors is it good to use?

Overture Maps data is organized in a strict schema and is released as a series of GeoParquet files that are available on platforms like Azure Blob Storage and AWS S3. It's free to access and can be used through a variety of tools that can interface with Geo Parquet data for different applications. It's like using OpenStreetMap data, but in a more structured way: The data has already been processed and put into a consistent schema. The documentation on overturemaps.org provides all the necessary information for accessing and working with the data. For example, if you want to use all of OpenStreetMap buildings, then Overture is a great resource because, in Overture's data, the buildings are already turned into actual polygons instead of nodes and ways.

With Overture it's possible to combine different sources of data that can also have different licenses. How does Overture handle different licenses from data sources like OpenStreetMap, Microsoft, or Meta? And what are the implications for users?

Each open data source carries its own license. OpenStreetMap data is under the ODBL license, while Overture's non-OSM data uses the more permissive CDLA-V2 license. Unlike the ODBL license, it doesn't trigger the share-alike clause. Each theme in Overture has its own license, and users need to be aware of this when working with different datasets. For example, the buildings theme is under ODBL because it uses OSM data, although it also mixes data from other sources such as Microsoft Buildings and Google Open Buildings that are also released under the ODBL. The addresses theme, by contrast, doesn't come from OpenStreetMap, but from a variety of different open data sources, including many government datasets. Each license is listed on the attribution page to help users understand where the data is coming from and what licenses are in place.

Could another open data project compete with Overture by establishing another standard?

That's a great point. But Overture is not trying to become a new geospatial standard. Instead, it's focused on providing an open map dataset for use by map service providers, relying on existing geospatial standards like those from the Open Geospatial Consortium. This ensures compatibility with established standards. If Overture's schema is adopted elsewhere, that would be beneficial. The project is focused on making its data organized and machine-readable, which means that it would be easy to translate to other areas. But we want to comply with existing standards, not reinvent the wheel.

Why would companies like Microsoft and Meta, which are competitors, contribute their data openly?

These companies have been involved in open data for a long time. Each one of them had to do their own thing. Overture provides a way for them to pool their resources in a transparent, open framework. The future of data is more open, and collaboration allows them to do more together than by working in silos. TomTom's 2022 announcement of their Orbis Maps Platform shares this sentiment of an improved open collaboration ecosystem that includes OpenStreetMap. They outlined that there is great value in open data, and I think that is a shared sentiment across much of the geospatial industry.

From a user perspective, many different actors, different purposes could be using this open data for many different purposes. Are there procedures or policies to govern the use of this open data? And is there a guarantee that the data will always be accessible?

Open data can be used by anyone, as long as it follows the license. There aren't many restrictions beyond that. As for data accessibility, there is a guarantee that the foundation is building infrastructure to ensure that this isn't just a one-time release of open data. Instead, the data will be continually updated and improved. Hosting on AWS or Microsoft Azure as part of open data programs provides some guarantees of long-term accessibility.

In contrast to OSM where tags are very heterogeneous, Overture is normalizing or standardizing, which also means cutting off specific tags and then choosing other tags. Does this mean that there might be a loss in the data to make it more processable?

Exactly. That has always been the trade-off. OpenStreetMap is less of a map and more of a database of openly tagged geospatial objects and that is what has made it so valuable over time. However, those trying to build services from it have to force it into something that is machine readable and consistent. Overture is basically doing that and then making that result openly available. Overture is making a lot of decisions around what that schema should be and how it's most useful to the map makers. But there are definitely trade-offs where just not all of the OSM tags are making it in. If you want to use Overture data, but you need the missing tags from OSM, you can always use the source id of that feature to join it back to OpenStreetMap data. As an open data project, we would also hope that you would raise that issue in one of our discussion forums on GitHub, so that maybe we could add that tag as well. In our documentation, we try to show exactly what tags get mapped to what classes in Overture.

How do you see the relation between OSM data and other open data? Do you think Overture will surpass OSM data in scale? And isn't it possible that over time, the amount of OSM and ODBL data in Overture will shrink relative to other data from corporate actors?

That's a great point. OSM currently contains about 600 million buildings, while machine learning datasets already surpass that with over a billion buildings. While Overture can incorporate ML data, the challenge is the workload needed to quality-check and input this data into OSM, especially if we want human validation. While Overture uses open licenses like CDLA V2, which means this data can flow back into OSM, the process of manually adding billions of buildings is daunting. This is where artificial intelligence

comes in. The Humanitarian OSM Team (HOT), which is a non-profit organization dedicated to disaster and crisis mapping and other groups are exploring how AI can relieve the burden of human labor in tasks like building tracing, which AI is well-suited for. Overture provides the flexibility to choose data sources. If users prefer community-validated OSM data without AI contributions, they can filter the Overture dataset accordingly. However, as more high-quality ML datasets come out, the ratio of OSM data to AI-generated data within Overture may shift. On the other hand, some AI data could eventually flow back into OSM as human contributors review and integrate it, causing the ratio to fluctuate.

Who ultimately makes these decisions about data selection in Overture, and is this process transparent?

Yes, the decision-making process is very transparent. Overture works collaboratively with its member companies and others to design the logic for handling data. This process is openly documented. For instance, building tags in OSM might translate into specific subtypes within Overture's schema, and this mapping process is detailed in the documentation.

While the core decisions are made by the companies involved in Overture, the project welcomes feedback from the community. Users can contribute to discussions, suggest changes, and point out any issues with the data via GitHub. This open data approach ensures that decisions remain open to scrutiny and that there's room for community input.

What advice would you give to beginners in this field?

A: For beginners, especially those in academia or research, my main advice is to engage deeply with the resources and new formats emerging in this field. Overture is pioneering in terms of how we release data being cloud-native and working with cutting-edge formats so it's quite different from traditional GIS approaches. This shift has created some challenges, but we are actively working to improve our documentation, including tutorials and getting-started guides, to help users adapt.

I would encourage beginners to explore our documentation as a starting point. It's constantly evolving to provide clear use cases and examples. One of Overture's strengths is its ability to bring together multiple datasets into one unified, global dataset, which is invaluable for research. For example, the Microsoft Buildings dataset, with over a billion buildings, has been widely used in academic research. Overture builds on this concept, offering even more comprehensive and combined datasets, making it a great resource for those seeking global data completeness.

Regarding data quality, some users may have concerns about specific datasets like Microsoft's or Google's open data. In Overture, we take special care during the conflation process, prioritizing higher-quality data wherever possible. Since these datasets are constantly being updated — Microsoft refreshes its data every month as new imagery becomes available — the quality will continue to improve over time. So, if you're looking for reliable, evolving geospatial data, Overture is an excellent resource to incorporate into your workflow.

Interview and Editing: Lama Ranjous, Susanne Schröder-Bergen, Adila Brindel, Max Münßinger

References

- Anderson, Jennings, Dipto Sarkar, and Leysia Palen. 2019. "Corporate Editors in the Evolving Landscape of OpenStreetMap." *ISPRS International Journal of Geo-Information* 8 (5): 232. <https://doi.org/10.3390/ijgi8050232>.
- Anderson, Jennings, Robert Soden, Brian Keegan, Leysia Palen, and Kenneth M. Anderson. 2018. "The Crowd Is the Territory: Assessing Quality in Peer-Produced Spatial Data During Disasters." *International Journal of Human-Computer Interaction* 34 (4): 295–310. <https://doi.org/10.1080/10447318.2018.1427828>.
- Herfort, Benjamin, Sven Lautenbach, João Porto De Albuquerque, Jennings Anderson, and Alexander Zipf. 2021. "The Evolution of Humanitarian Mapping Within the OpenStreetMap Community." *Scientific Reports* 11 (1): 3037. <https://doi.org/10.1038/s41598-021-82404-z>.
- Palen, Leysia, Robert Soden, T. Jennings Anderson, and Mario Barrenechea. 2015. "Success & Scale in a Data-Producing Organization: The Socio-Technical Evolution of OpenStreetMap in Response to Humanitarian Events." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 4113–22. Seoul Republic of Korea: ACM. <https://doi.org/10.1145/2702123.2702294>.



 Licensed under CC-BY 4.0

Permanent link:

<https://gridisnotajournal.de/articles/anderson-2025-overture>