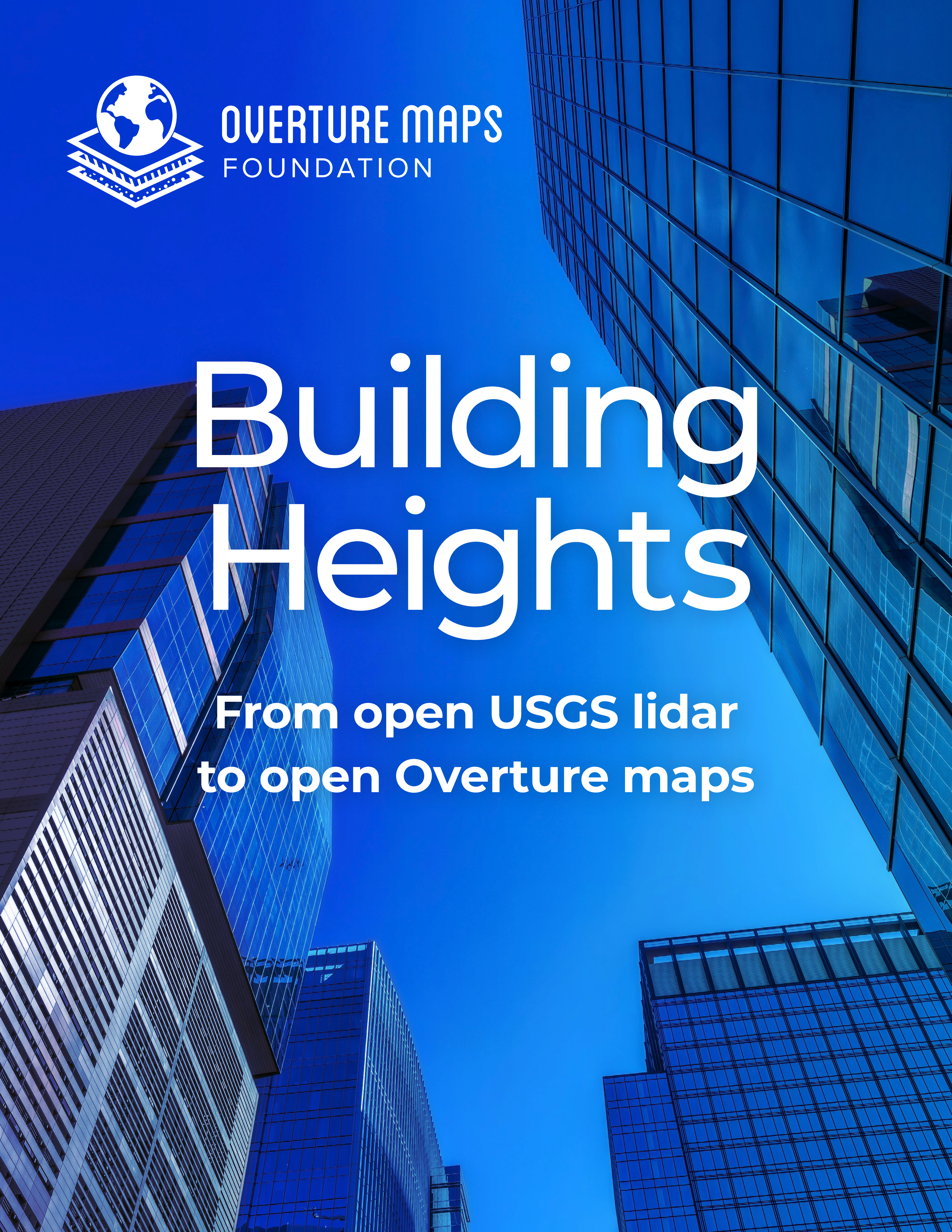


OVERTURE MAPS FOUNDATION

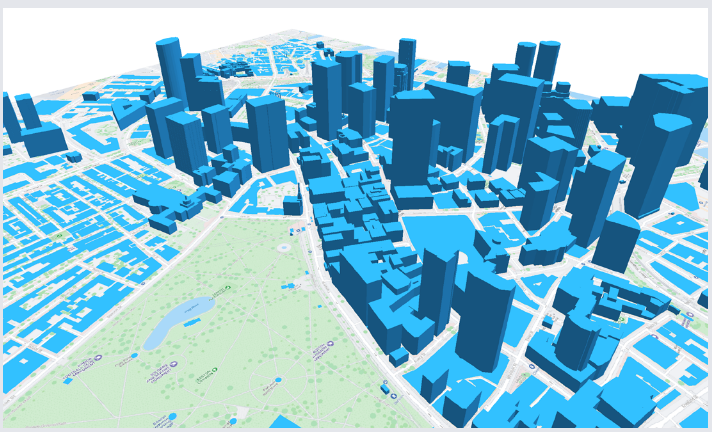
Building Heights

From open USGS lidar to open Overture maps

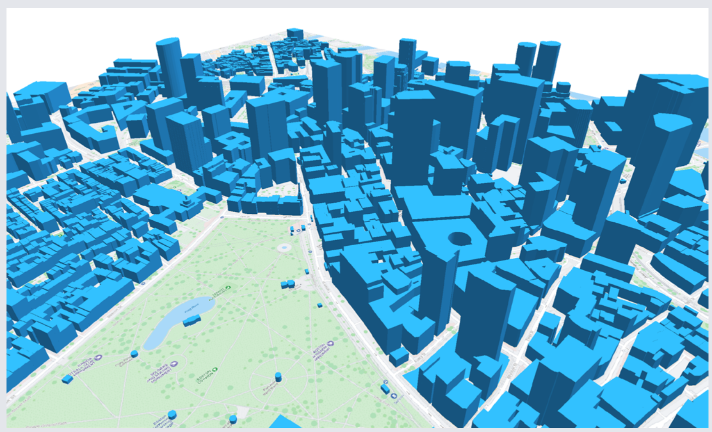
Summary

As part of the initial Overture Maps Foundation data release, we are publishing heights for over six million buildings estimated using over 34,000 km2 of open 3D Elevation Program (3DEP) lidar data from the US Geological Survey (USGS) — enabling GIS analysis and richer map visualizations. The estimated heights apply both to buildings in OpenStreetMap (OSM) as well as to buildings detected from imagery by Microsoft. This is the initial installment of estimated building heights; over the next several months, we plan to release estimated heights for the majority of US buildings in OpenStreetMap. In addition to increased coverage, we will also continue to refine our process in order to increase the accuracy of our estimations.

Boston without building heights



Boston with building heights



MAP DATA FROM OPENSTREETMAP

Many map visualizations already make use of building heights. By increasing the percentage of open building heights in the US, we can enable a richer map experience, whether through better visualizations or through improved infrastructure for Visual Positioning and other advanced technology.

Motivation

Per-building height estimates can enrich GIS analysis and enhance map visualization. While the USGS has made elevation data available both as point clouds and as digital elevation rasters, having the height attached to a particular building greatly simplifies applications and analyses. Currently in OpenStreetMap, approximately 10% of the buildings in the US are tagged with estimated heights or number of levels. In the past, some of these height tags have been populated by local mappers and GIS departments. By using open 3DEP data, we have an opportunity to populate them at scale.

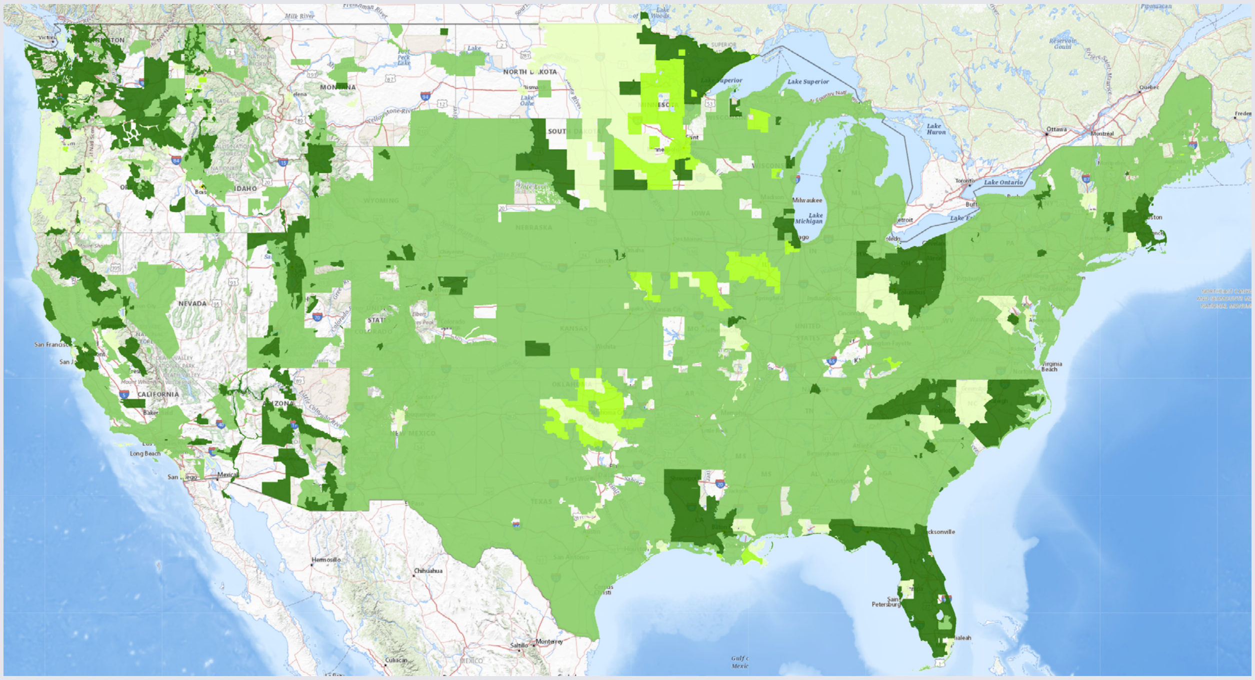
Scale of coverage

For this initial release, we have focused on a small set of pilot areas. Over the course of the year, we plan to continually increase our coverage over the entire contiguous US. This release includes building heights in:

- Eastern Massachusetts (Boston)
- Orange County, FL (Orlando)
- Cook County, IL (Chicago)
- Maricopa and Pinal Counties, AZ (Phoenix)
- King County, WA (Seattle)
- Santa Clara County, CA (San Jose)

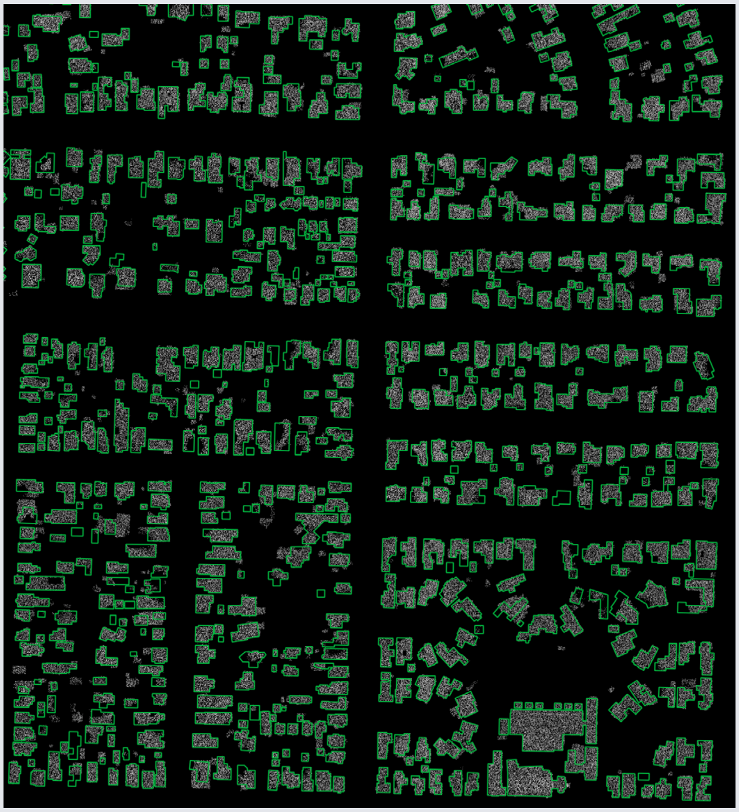
The goal of the 3DEP program is to acquire lidar across the nation and "provide the first-ever national baseline of elevation data." That open lidar data is the key input to our building estimation pipeline and determines the areas that we can process. The map below shows the scale of current 3DEP lidar available from the USGS color coded by quality level (QL).

This initial release focused on the 3DEP data with highest quality level (QL1 in dark green), but we plan to process QL1 and QL2 regions (light green) covering the majority of OSM buildings in the US.



Height estimation

To estimate the height of a building, we use the building footprint (from OSM or from imagery detections) and select lidar samples within that footprint (See figure). We then calculate statistics on those lidar samples to provide an estimate of the absolute elevation of the tallest part of the building. The lidar is an excellent source to measure this elevation due to its high accuracy; the [QL1 and QL2 data is specified to have vertical error not exceeding 10cm RMSE.] However, instead of elevation, we want to provide a building's height above the ground. To do this we need to subtract the elevation of the ground. We currently sample bare earth terrain sources (including 1m USGS digital terrain models and/or [Mapzen terrain tiles]) along the building's perimeter.



MAP DATA FROM OPENSTREETMAP

We estimate height by sampling LiDAR points within building footprints.

Quality assurance

In order to validate the estimations, analysts inspect a sample of buildings using aerial imagery. With an internal tool, they select the building's highest point from multiple views and compute an absolute elevation of the building by triangulation. This alerts us to quality issues and gives us confidence in the performance of the algorithm. Quality assurance (QA) is currently performed on each lidar project (a project is a USGS region collected at the same time and under the same conditions) by taking random building samples, stratified across urban density. During the runup to this initial release, manual checks on more than 3,000 buildings were performed.

Technical challenges and future improvements

Any automated process is going to produce some errors. A large part of our effort has been to catch these errors and improve incrementally. Further improvements will be reflected in future releases. Technical challenges and planned improvements include:

- **Alternative bare earth terrain sources.** Where available, we are using 1m bare earth terrain from the USGS. This is a flagship product of the 3DEP program and is extremely reliable. However, in some areas where this is not available, we currently rely on Mapzen terrain tiles. While globally complete, these offer lower spatial resolution and can occasionally produce spurious values which, when subtracted from roof elevation, directly affect the height estimate. In future releases, we plan to use the point cloud to estimate the elevation of the terrain instead.

- **Point clouds without building classifications.** For some USGS projects, the vendor has classified building points in the point clouds. In these cases, height estimation is very straightforward since we can restrict our statistics to these points. However in about 85% of projects, building points are not classified separately. We notice this especially when a tree overhangs the footprint of a building. Since these points are included in the statistics, that building's height will be overestimated. In future releases we will continue to refine our approach including a combination of alternative estimators, taking into account the spatial distribution of the points, using last return lidar points, performing our own point classification, and possibly using coregistered imagery to segment trees.

- **Decreased point density.** This current release prioritized using lidar of QL1 – the most dense point clouds. As we scale to using less dense QL2 lidar, we may have to discard smaller buildings. It is also possible that we will undersample small rooftop structures like spires and thus underestimate the maximum building height.

- **Misalignment between building footprints and lidar.** Since the building footprints and the lidar point cloud are collected separately, they can sometimes be misaligned. When the scale of the misalignment is on the order of the distance between buildings, it is possible that samples from neighboring buildings can contaminate the sample of a height measurement. Here again, we are considering more sophisticated estimators that use the spatial distribution of the points.

- **New buildings.** Similarly, there are cases when we have a new building footprint, but the building did not exist at the time of the lidar collection. In these circumstances we will often omit any estimation, but it is possible that old, high vegetation points in the footprint could be considered. Point classification improvements will help here, too.

OVERTURE MAPS FOUNDATION

Open data, open maps

The Overture Maps Foundation intends to use collaborative map building to generate high quality open map data. This project is an example of the benefits of this type of collaboration. It is the result of community contributions, industry partnerships, and government-supported open data. No single party in this collaboration could have achieved this outcome by itself, but by combining data, technology and engineering, new open map datasets can be built. We hope in the future to spur similar innovation and collaboration on all types of map data.