

The Path to High-Quality Point-of-Interest (POI) Data

How to Validate and Update POI Databases at Scale

Key Takeaway

- Users of POI data should constantly be questioning the quality of their data. POI data can be extremely valuable but is only as good as its accuracy.
- With the onset of COVID-19, there is an increasing turnover in POI data.
- Crowdsourcing can help you validate and update your POI databases at scale.

Introduction

A Point/Place-of-Interest (POI) most commonly refers to a point on the earth representing a place. For example, at geographic coordinate latitude 40.689167 and longitude -74.044444 stands the Statue of Liberty. A POI database is a collection of points representing places such as tourist attractions, restaurants, hotels, banks and more. It is relatively easy to drop pins on the map to create a POI database, but validating, maintaining and updating a POI database can be difficult, time-consuming and costly.

Chances are you have a smartphone, and—know it or not—chances are you use a POI database on nearly a daily basis. For example, many of us rely on driving direction applications to take us from our current location to a specific place—the destination location is part of a POI database. Unfortunately, there are times one has arrived at the destination only to find that the place is permanently closed or moved. Sadly, due to the COVID pandemic, now more than ever, restaurants and stores are closing on a regular basis. As a result, places in POI databases are increasingly perishable. The Premise platform and its contributor network is an ideal platform for validating and updating POI databases.

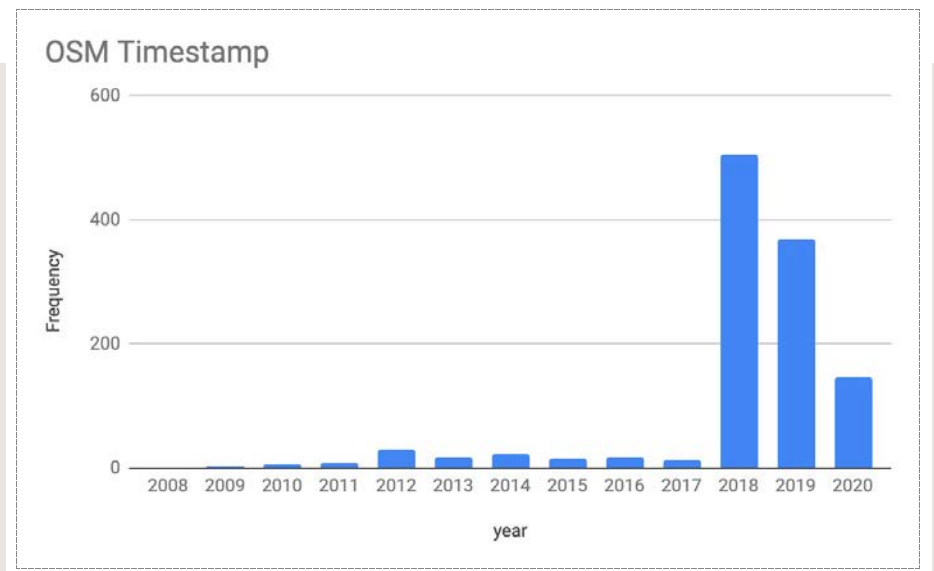
The Perishability of Places

A POI database is only as good as the quality of the information it contains. Premise's ability to crowdsource geolocated data from across the globe provides a unique ability to get near real-time information on locations critical to maintaining an accurate database. To demonstrate Premise's ability to efficiently and effectively validate and update POIs, we ran a test using OpenStreetMap (OSM) data in Kenya.

First, we pulled all ATM and bank POIs in Kenya from our two repositories—OSM has 1,158 POIs, and Premise has 1,129 POIs as of September 23, 2020. Here is a breakdown of the two databases:

- Of the 1,158 ATMs and banks in the OSM Kenya POI database, 46.2% are unique to OSM. The remainder is OSM's Kenya POI database overlap with points in Premise's POI database.
- Of the 1,129 ATMs and banks in the Premise Kenya POI database, 41.9% are unique to Premise. The remainder overlap with OSM's POI database.

Looking at the collection as a whole, the vast majority of OSM's ATMs and banks in Kenya have a timestamp from 2018 or newer, as illustrated in the chart below. Given the data's relative recency, the Premise team initially did not assess that updates would be required to the database. Surprisingly our experiment showed we were wrong, thereby underscoring just how perishable POI data can be, as we will illustrate shortly.



To test our hypothesis that recent OSM data is valid and current, the Premise team distributed a sample of unique ATM locations from the OSM database to our contributors as a directed location-based discovery task. This means a contributor on the Premise app selects the task and is then expected to go to the physical location to verify the ATM is indeed there. Upon arriving at the specified location, the Premise contributor will have three options:

1. If an ATM is at the specified location, and the name on the ATM matches what is in the OSM database, then the Premise contributor will take photos of the ATM that verify the information.
2. If an ATM is at the specified location, but the name does not match what is in the OSM database, then the Premise contributor will take photos of the ATM and enter the modified name of the ATM at that location.
3. If no ATM exists at the specified location, then the Premise contributor will take photos of the place where an ATM was specified in the OSM database and report the absence.

Given the ATM data's recency, as illustrated in the chart above, we thought we would only see data from the first option above. However, after just a week of the tasks running in Nairobi, we can see a significant number of submitted tasks from the second and third options.

Validate and Update POI

When we began this experiment, we were expecting to largely find ATMs at their specified location and with a name matching the one in the OSM database. Instead, we found 35.7% ATMs at the specified location without a matching name and 21.4% where the ATM did not exist at the specific location.

Here is an example of when an ATM is at the specified location, but the name does not match the OSM database. The OSM POI database (image 1) and Google Maps (image 2) both show a Chase Bank at virtually the same location in Nairobi, which was distributed to Premise Contributors for validation. Of note, Google Maps' most recent photo is from a Google Local Guide taken in August 2017 (image 3), as shown in the image below.

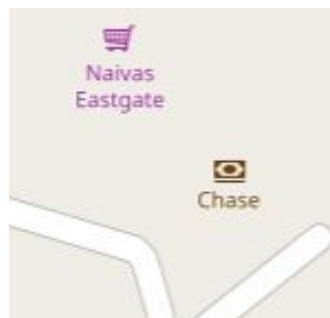


Image 1 - OSM POI Database

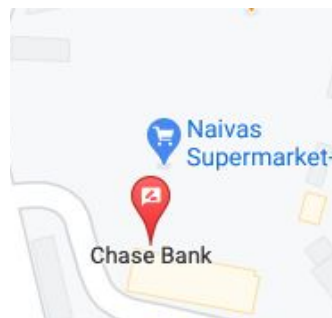


Image 2 - Google Maps



Image 3 - Google Local Guide

When the Premise Contributor went to that location, they noticed a bank was still present, but it was no longer a Chase branch as initially reported. To confirm the change, they could take photos and update the name from Chase Bank to SBM Bank. According to open-source research, SBM Holdings acquired Chase Bank (Kenya) in August of 2018.



Image 4 - Premise Contributor submission

To illustrate, when no ATM exists at the specific location, the OSM POI database (image 5) and Google Maps (image 6) show a Barclays Bank at practically the same place in Nairobi, which was also distributed to Premise Contributors for validation.

In this instance, Google StreetView (image 7) shows that Barclays is still active as of April 2018, the date on the most recent StreetView image. However, when a Premise Contributor went to that location on September 25, 2020, they found that the bank was no longer there. Their photos (image 8) clearly show the building that once housed the bank no longer does.

Get Data at the Speed of Life

Points/Places-of-interest are the foundational building blocks of spatial data science, including turn-by-turn routing, footfall traffic, AI/ML models, etc. As such, the accuracy and integrity of the POIs are critical for anyone using this data. Unfortunately, the most popular open-source options cannot keep up with the pace of change on the ground.

Premise's ground truth platform and ML algorithms are an ideal fit to maintain data provenance and continually validate, on a recurring update schedule, the open-source data such as OpenStreetMaps and Google StreetView, as well as proprietary POI datasets.

In this experiment, the Premise platform and its contributors validated and updated dozens of Nairobi, Kenya OSM points in less than 48 hours. With a global network of 2M+ contributors in 100+ countries, Premise specializes in on-demand, crowdsourced information—from sentiment (e.g., safety, event impact, personal opinion) and demographic data (e.g., language, religion, ethnicity) to location information, gathering nearly 3.5 million submissions in May 2020 alone.

To learn more about the Premise platform and how we can provide your organization with better POI data, please reach out to us at info@premise.com.



Image 5 - OSM POI Database

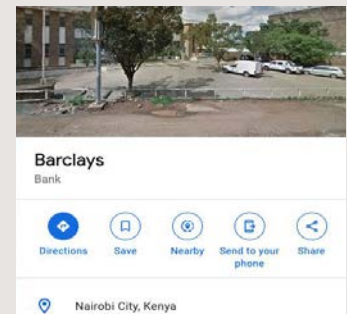


Image 6 - Google Maps



Image 7 - Google StreetView



Image 8 - Premise Contributor submission

© 2020 Premise Data Corporation.
All Rights Reserved.

