

4. Data Creation & Collection: Quick Start Guide

Creating data for OpenStreetMap typically takes two forms: remote mapping and field data collection. Remote mapping is the process of tracing infrastructure such as buildings and roads into OpenStreetMap using aerial or satellite imagery. This process allows for volunteers and team members to create basemaps anywhere in the world with only a computer and internet connection.

Collecting data on the ground is an integral part of editing OSM. This data is typically richer in detail and more accurate than map data that is generated solely by remote tracing. Mappers can choose from a number of free and open tools that facilitate on-the-ground data collection in both online and offline settings.

4.1 Remote Mapping: Mapathons & Learning How to Map

Open mapping techniques, particularly editing OSM, are easy to learn and present a low barrier to entry for new contributors. Volunteer mappers can learn through in-person trainings, while attending a mapathon, and/or from a variety of online tutorials and videos. The OSM community has created a wide variety of online tools to learn how to map, including:

- LearnOSM, which includes instructions for learning how to map in 17 languages;
- HOT Toolbox;
- HOT's Learn to Map! Youtube playlist;
- MapGive, which offers extensive video tutorials on mapping for humanitarian efforts;
- Missing Map's Beginner's Guide;
- Mapbox's visual guide to editing OSM; and
- The OSM Wiki page to find out more about the communities in your area.

Humanitarian and development organizations and governments host virtual and in-person mapathon events to engage volunteers in contributing to crowdsourced mapping projects. The typical objective of a mapathon is to provide humanitarian and development efforts with improved access to data in OSM. A typical mapathon asks participants to edit OSM for a target area, provides training for new mappers, and may feature remarks or a presentation on the purpose and goals of the mapping project.

Mapathon planning resources include:

- MapGive in a Box
- Missing Maps' Event Hosting guide
- HOT's How to Host a Mapathon video.

Remote mapping is also typically the first step in the field data collection workflow. Prior to going into the field, teams are encouraged to check OSM for completeness and temporal validity. If data for an area of interest is out of date or incomplete, teams should conduct remote mapping to complete an up-to-date basemap for field activities.

Tasking Manager

The HOT Tasking Manager is a mapping tool designed and built for the Humanitarian OpenStreetMap Team's collaborative mapping process in OpenStreetMap. The purpose of the tool is to divide up a mapping project into smaller tasks that can be completed rapidly with many people working on the same overall area. It shows which areas need to be mapped and which areas need the mapping validated.

For more information and resources on the Tasking Manager, visit: <https://github.com/hotosm/toolbox/wiki/3.1-Working-with-the-HOT-Tasking-Manager>

There are several ways to edit in OpenStreetMap. The two most commonly used tools and best programs for mapping projects are iD editor and JOSM.

iD Editor

The iD editor is a user-friendly tool that allows you to directly make changes in OpenStreetMap. iD is good for:

- Simple edits

- Fast Internet access to load the imagery and save the edits.
- Following a consistent and simple tagging scheme.
- When you are restricted from installing a program on the computer you are using.

Java OpenStreetMap Editor (JOSM)

JOSM (Java OpenStreetMap Editor) is an open source editor for OpenStreetMap data. JOSM is best for:

- Adding many buildings (See buildings_tool plugin).
- Editing many polygons or lines that already exist.
- When you are on an unreliable Internet connection or offline.
- Using a specific tagging scheme (or custom presets).

For more information and resources on iD Editor and JOSM, visit <https://github.com/hotosm/toolbox/wiki/3.3-Editing-with-iD-and-JOSM>

4.2 Tools for Field Data Collection

Several mobile applications exist to assist with field data collection. The following are Data Collection Tools commonly used in the HOT workflow. A detailed overview of these and other data collection tools can be found at [insert link]. Choosing an application to use depends on mobile device capability, varying set-up requirements, and survey needs. It is also important to note that some tools require others (i.e. OMK requires use of ODK), and others can be complementary (i.e. OSMTracker or Mapillary can be used to supplement data collection with ODK, Kobo, and OMK).

I want to collect...

ODK

Kobo

OMK

Maps.

me

OSM

Tracker

Mapillary

Qualitative survey data

☐

☐

☐

☒

☒

☒

Quantitative survey data

☐

☐

☐

☒

☒

X

GPS Points

☒

☒

☒

☒

☒

X

Photos attached to GPS Points

☒

☒

X

X

☒

X

GPX Tracks

X

X

X

X

☒

☒

Streetview imagery

X

X

X

X

X

☒

Data attached to OSM points of interest

X

X

☒

☒

X

X

Data attached to OSM polygons (i.e. buildings)

X

X

☒

☒

X

X

Data attached to OSM ways (i.e. roads)

X

X

X

☒

X

X

OpenDataKit (ODK)

ODK is a free an open-source set of tools which help organizations author, field, and manage mobile data collection solutions. ODK Collect is part of ODK and is an Android app that replaces paper forms used in survey-based data gathering.

Use OpenDataKit (ODK) if:

- You have access to mobile devices but they have limited RAM & storage
- You do not need to collect data for buildings in OSM OR you are able to manually transfer data collected as points to OSM polygons after data collection.
- Someone on your team is comfortable developing or learning to develop ODK forms using spreadsheet software.

OpenMapKit (OMK)

OMK is an extension that launches directly from within ODK Collect when the OSM question type is enabled in a standard survey. It is what allows you to browse OSM features, and to create and edit OSM tags.

Use OpenMapKit (OMK) if:

- You have access to mobile devices with sufficient RAM & storage
- You need to collect data for buildings in OSM
- You have the capacity for more intensive set-up prior to data collection

KoboCollect

KoboCollect is in almost all ways similar to ODK Collect, and is built on top of the ODK platform. Kobo also has prebuilt analysis tools and is another popular option.

Use KoboCollect if:

- You have access to mobile devices with sufficient RAM & storage
- You would like to use a form builder to build your survey (as opposed to building forms in spreadsheet software, as with ODK)
- You do not need to collect data for buildings in OSM OR you are able to manually transfer data collected as points to OSM polygons after data collection.

OSMTracker

OSM Tracker is “an offline GPS tracker designed for collecting points of interest (POI) to be added to the map and for recording GPX tracks.” OSM Tracker is free and open-source. Several projects have used OSMTracker to complement ODK/OMK/Kobo survey collection.

Use OSMTracker if:

- You need to collect waypoints and GPX tracks.
- You need to collect photos with waypoints.

- You need to map points of interest.
- You do not need to conduct a survey.

Maps.Me

Maps.me is a navigation application that uses OpenStreetMap data, and can be used offline. It is suitable for collection Point of Interest (POI) information, as far as these fit within the types of data that Maps.me shows you on the map.

Use Maps.me if:

- You need to map points of interest.
- You want to upload points to OSM without data cleaning or additional processing.
- You need an application for navigation.

Mapillary

Mapillary is an open-source street view imagery capture application that integrates with OpenStreetMap. Several projects have used OSMTracker to complement ODK/OMK/Kobo survey collection.

Use Mapillary if:

- You need to collect street view imagery.

For more information and resources (including skills and tools necessary) on data collection applications, visit <https://github.com/hotosm/toolbox/wiki/4.2-Data-collection-application>

Data Collection Servers

When beginning a mapping project, many organizations ask if they need to have a server. After collecting data, you'll need to get the data from the devices. Sometimes, it works to simply collect and process data directly from the data collection devices. However, this does not scale well when you get more people collecting data, and also means your data is not backed up - if you lose the device, you can lose the data. Prior to data collection, it is important to have a data management strategy that is suited for your operations.

Using a data collection server allows for much better management of forms and deployments, collection and aggregation of responses, and can offer additional features for viewing, analysing and exporting data. Use of a server may be restricted by available resources (cost of a physical server) and/or internet connection (access to cloud server). The most used options are:

- **POSM** - a physical server that contains a set of OpenStreetMap tools, including OpenMapKit server. POSMs allow multiple users to connect and upload data from data collection devices to a central location without the need for internet access. This data can then be aggregated using the OMK Server and synced with OSM directly or downloaded for analysis and processing.
- **OpenMapKit Server** - OpenMapKit Server is a cloud based storage system for specially designed to store and compile OpenMapKit data. Data collected through OpenDataKit can also be uploaded to an OpenMapKit Server.
- **Kobo Toolbox** - An online application that allows users to build Kobo/ODK surveys as well as store, aggregate, and perform analysis of Kobo/ODK data.
- **ODK Aggregate** - ODK Aggregate is an online application (local options possible) that stores, aggregates, and allows users to perform basic analysis on ODK data.

More information and resources on servers can be found at:

If the use of a server is not available to you, it is still possible and crucial to store backups of data. In this case, data will need to be downloaded or otherwise shared with a central location, such as a laptop computer, and cloned to a secondary location such as a hard-drive or second computer. For more information about a server-less data management workflow, visit <https://github.com/hotosm/toolbox/wiki/5.1.2-Serverless-Data>