5.2 OMK Field Data Cleaning Workflow

After field data collection with OMK, clean the data before upload to OSM. Below is an example version of the OMK Field Data Cleaning Procedure. This is provided for guidance but should be modified as needed.

Skills and Technology Needed * JOSM

Step 1. Create folders

Create the folder structure for the data cleaning process.

1. Create primary working folder: `[omk_field_data_cleaning]`
2. Create a container folder: `[village_name]_[current-date]`. For example: `akweteyman_2019-01-31`
3. Create a sub-folders within the container folder:
    1. Add a folder for the raw data: `/01_raw_[yourOSMname]`
    2. Add a folder for the working data: `/02_working_[yourOSMname]`
    3. Add a folder for the final data: `/03_final_[yourOSMname]`

Step 2. Raw Files

Supervisor or team leader will download data from the server, and select sections of data for individual data cleaners to clean up.

For Supervisor:

- *Each mapper's device id (IMEI) and username should be registered, to keep track of their submissions to the server.*
- *Using the filter option of the server, filter server submissions according to dates, download data for a specific data, divide it into subsections and give those to the data cleaning team.*

1. Create a sub folder `[server_downloads]` for all downloads from the server naming downloads according to the date of download, i.e `akweteyman_2019-01-31_server_download.osm`
2. While dividing out this data to the different team members, **purge** everything else except what you want to give to a specific team member, and then 'Save As…' that section. Do not copy a section of the data downloaded from the OMK server into another new file, as this will not sync while downloading data from OSM database to start cleaning field data. Read about the purge action: https://josm.openstreetmap.de/wiki/Help/Action/Purge
3. Do this for every team member you are giving data to clean.

For data cleaners:

1. The file your supervisor gives you will be named as: `[projectArea_name]_[current-date]_raw_[yourOSMusernam`
   For example: `akweteyman_2019-01-31_raw_lusdavo.osm`
2. Copy the file your team leader gives you to the `[projectArea_name]_[current-date]` → `01_raw_[your_OSMusername]_[id]` folder. Format dates as YYYY-MM-DD.

For example:

`akweteyman_2019-01-31` → `01_raw_lusdavo` → `akweteyman_2019-01-31_raw_lusdavo_10.osm`

NB: The represents project area subdivisions you are supposed to be working on.

Step 3. Add OSM Files to JOSM

Add file to JOSM

1. Open JOSM

2. Drag the .osm file into JOSM

Step 4. Sync Field Data with OSM Data

This step covers a series of checks on the data to check if there are any issues with the data.

1. You will need several *plugins* during this cleaning process. Read about plugins here: https://wiki.openstreetmap.org/wiki/JOSM/Plugins. Add the following plugins:

- `todo`
- `opendata`
- `utilsplugin2`
- `buildings_tool`

2. You will be using several filters during your cleaning process. Read about Filters here: https://josm.openstreetmap.de/wiki/Help/Dialog/Filter. Add the following filters: 'type:node untagged' 'new OR modified'

3. Prepare your data for cleaning:

- Select field collected data using the JOSM filter: `type:node untagged`
- Select all results (Ctrl+A)
- Add the selection to the TodoList plugin. In the plugin window, click '+ Add'
- Deselect results by clicking in the 'Map View' window
- Download OSM data in the area of the field collected data, to ensure we merge properly with existing OSM data
  - Press '2' on the keyboard to zoom to the extents of your field collected data
  - Select 'File -> Download in current view'
- Now save this file to your 'working folder'
  - File → Save As → `[projectArea_name]_[current-date]_working/[working_yourOSMusername]_[id]`
  - For example: `akweteyman_2019-01-31_working_lusdavo_10.osm`

Step 5. Cleaning Data

1. Select/enable the filter 'new OR modified'
2. Select only new or modified data using the JOSM filter: 'new OR modified', tick 'E' and 'I' to differentiate between field data and OSM Server data. Check if everything looks okay.
3. You are now going to clean the 'working' file in JOSM.
4. Add Imagery being used to clean data.
5. Uncheck the 'type:node untagged' filter to enable you edit features.
6. Convert all building nodes to polygons by using replace geometry feature.
7. From the ToDoList, inspect each feature (select the first one in the TodoList)

- Are the locations correct?
- Convert comments into possible tags as possible else change the key to fixme
- Are there any spelling mistakes in the attributes of this feature?
- Are the tags capitalised correctly?
- Are the tags fields filled correctly?
- Are we not duplicating existing OSM data?
- Expand all text and correct all text, examples below:
  - St. → Street
  - Ave → Avenue
  - Ltd. → Limited
  - BANK OF GHANA → Bank of Ghana
  - GOIL → Goil, etc

8. If okay, click 'Mark' in TodoList to proceed to the next item

5.1 Cleaning Apartments

Use the `Terrace a Building` tool to divide a building into the different number of apartment units that building has, then copy and paste attributes from an apartment point to the respective building terrace. If a building is 5 or more sided, manually draw the different apartments connecting them together.

5.2 Record Errors Encountered

Capture common problems that come up while cleaning data.

1. Data cleaners should record any errors and problems encountered (both while mapping and validating).
2. Please record this in a text file with screenshots if possible
3. Save edits to `<projectArea_name>_<current_date>/02_working'` folder.

Step 6. Save Final Files

> NOTE: Some projects may require that datasets are split into public and private datasets, where the full dataset remains private, and personal/private data is removed from the partial dataset uploaded to OSM. For similar situations, use 6.1 and 6.2. For projects where all data collected is public and can be uploaded to OSM, skip to 6.2

1. Right-click the working layer and select 'Save As..'
2. You are to 'Save As..' to the final folder. Team Leaders are to review their team's data validation. If satisfied, get the Group Supervisor to double check the data.
3. Disable all filters used during data cleaning.
4. Run 'Validation' in the 'Validate Results' window. Correct any errors that are detected, and go through the warnings.
5. Notify supervisor for review.

6.1 Private Data

If full data set includes private/personal information: 1. Save to the final layer: `[projectArea_name]/[current_date]/` for example `akweteyman/2019-01-31/03_final_lusdavo` folder. * File name: `akweteyman_2019-01-31_final_lusdavo_` 2. From the saved file to be uploaded to OSM, remove the private/personal attributes before uploading to OSM (enable the type:node untagged and new OR modified filters). * For example: * building:population * building:sleeping_spaces * building:rooms * Building:rooms_painted

NOTE: Do not overwrite the _private file. These changes will be saved to a new layer in step 6.2.

6.2 Public Data for upload to OSM

1. Save to the final layer: `[projectArea_name]/[current_date]/03_final`, for example `akweteyman/2019-01-31/03_final_lusdavo` folder.
   - File name: `akweteyman_2019-01-31_final_lusdavo_10.osm`
2. Notify supervisor for review

Step 7. Upload to OSM

If Supervisor authorises, upload the validated data to OSM

1. Add the **source** tag to all field data. At minimum, `Field Survey`
   - Example: source= Open Cities Accra - Field Survey
2. Disable the 'type:node untagged' and 'new OR modified' filters

3. With your OSM Username and Password added to JOSM preferences, upload to OSM adding the appropriate tags for the project as a changeset comment.
   - Example: `#OpenCitiesAfrica #OSMGhana #MobileWebGhana #HOTOSM` Specify the data source for change as survey.

Step 8. Quality Assurance using OSMCha|QGIS|JOSM|Osmose

This is intended to be used by supervisors to ensure data quality

Once data is cleaned and uploaded to OSM, several tools will be used to monitor Quality Assurance. * OSM Cha: Will be used to monitor the edits made by users uploading data to OSM, Filters will be made using UserIDs of data to give a summary of modified, deleted, added nodes, which the QA person will review and either flag as bad or good, sending feedback to the person who uploaded the data. Example of filtered changesets * JOSM: JOSM Validation tools, filters and customised MapCSS will be used to flag data quality issues, then give feedback to the data entry people for improvement. * Osmose

For more information, see Quality Assurance Tools.