# Hootenanny Conflation Statistics Report

**Digital Globe hootenanny.help@digitalglobe.com**

**0.2.9-191-g9a4df83-dirty, March 31, 2015**


**Prepared For:**

National Geospatial Intelligence Agency

# Contents

# 1 Introduction

Statistics for the Hootenanny conflation command tool are documented in this report. First a summary is provided of the most important statistics. This is followed by a section describing statistical aspects of the input and conflated output datasets. The last section reviews the run-time performance for each of the processing steps performed by the tool.

Please note that unsupported features are not represented in the statistics generated for this document.

# 2 Run-Time Specifications

### Input Files

- Input 1: /projects/hoot/test-files/ToyTestA.osm
- Input 2: /projects/hoot/test-files/ToyTestB.osm

### Output File

- Output: /tmp/output1.osm

### Parameters

- Parameter 1: -C UnifyingRoads.json
- Other Parameters (blank if none):

### Command Executed

- Command:

hoot conflate -C UnifyingRoads.json -D writer.include.debug=true -D stats.class=hoot::ScriptStatsComposer

-D stats.format=asciidoc -D stats.output=./report/reportBody /projects/hoot/test-files/ToyTestA.osm

/projects/hoot/test-files/ToyTestB.osm /tmp/output1.osm --stats

# 3 Summary

A summary of the most important statistics for the conflation job just executed are presented in the bar graph and table below. Shown are the number of features for both inputs and the total features in the output. Next is the number of conflated features and their percentage relative to the total number of features. Last is the number of reviewable features and their percentage relative to the total number of features. Reviewable features are the features marked for manual review.

Note that for the percentage values to be more meaningful, it should be reported as a function of the features supported by Hootenanny for conflation. We show the percentages for conflated and reviewable features relative to the total number of features at this stage of development because there is no easy way to calculate and record the number of supported features without affecting the performance of Hootenanny's conflation.

summaryChart.png

| | **Count** | **Percentage** |
|---|---|---|
| Input 1 Total Features | 4 | NA |
| Input 2 Total Features | 4 | NA |
| Output Total Features | 8 | NA |
| Conflated Features | 4 | 50% |
| Reviewable Features | 8 | 100% |

# 4 Statistical Aspects of the Input and Output Datasets

Statistics for a variety of aspects of the input and output datasets are presented in this section.

## 4.1 Summary of Basic Feature Elements

There are three basic elements of Open Street Map's conceptual data model of the physical world. These are nodes (points in space), ways (linear and area features), and relations (a high-level construct that helps coordinate how ways and nodes work together). To go along with these are tags which can associate with any of the three basic elements. In this section, a summary of the number of each element and tags are shown in the bar graph. For each, a comparison of the number of features for the two inputs and conflated output are shown.

Note: More information about the OSM Elements is provided at: wiki.openstreetmap.org/wiki/Elements

featureCountHistogram1.png

| | **ToyTestA.osm** | **ToyTestB.osm** | **Conflated file** |
|---|---|---|---|
| Nodes | 36 | 65 | 39 |
| Ways | 4 | 4 | 8 |
| Relations | 0 | 0 | 0 |
| Tags | 8 | 4 | 22 |

## 4.2 Summary of Key Hootenanny Features

Some of the key feature types that Hootenanny operates upon are *buildings*, *highways*, and *points of interest (POIs)*. These types are defined using the naming conventions that OSM assigns for these attributes. The buildings type are the area footprints for what the title indicates—buildings. The highways type is a bit more generic and refers to a collection of linear segments that include road types (highways, streets, dirt roads), cart tracks, and trails. The POIs are defined for Hootenanny as nodes, but in general are either linear or area features. The POIs are typically derived from the OSM attribute "amenity". For example, POIs may be a "place of worship", "cafe", etc. A summary of the number of buildings, highways, and POIs within the input and output datasets is shown in the bar graph.

featureCountHistogram2.png

| | **ToyTestA.osm** | **ToyTestB.osm** | **Conflated file** |
|---|---|---|---|
| Buildings | 0 | 0 | 0 |
| Highways | 4 | 4 | 8 |
| POIs | 0 | 0 | 0 |

### 4.3 Summary of Conflated Features

The statistics key to the conflation process are described in this section. The bar graph shows first the number of conflated features relative to the total number of features (see the first two bars on the left). Then a breakdown of the types of features conflated are presented in the remaining bars to the right. These include the three key feature types that Hootenanny processes: *buildings*, *highways*, and *POIs*. The bar heights show the number of each type of feature in the output dataset. The table shows the actual count values and percentages. The conflated feature percentage is relative to the total features, but the remaining percentages (denoted in parenthesis) are each relative to the number of conflated features.

conflatedFeature.png

| | **Count** | **Percentage** |
|---|---|---|
| Total Features | 8 | NA |
| Conflated Features | 4 | 50% |
| Conflated Buildings | 0 | ( 0% ) |
| Conflated Highways | 4 | ( 100% ) |
| Conflated POIs | 0 | ( 0% ) |

### 4.4 Summary of Area Features

Area features are the collection of OSM attributes that encompass a geographic area like a building footprint. Hootenanny supports conflation of building type area features to date. The statistics presented in this section show the area features, the portion of those that are buildings, and the portion of buildings that are conflated. The bar graph shows the area that each of these categories encompass across each set of items. The area provides an estimate metric for how much effort was performed by the tool. Measurements are recorded in square kilometers. The table shows the area measurements for the two inputs and conflated output for each category (where appropriate).

areaHistogram.png

| | **ToyTestA.osm** | **ToyTestB.osm** | **Conflated file** |
|---|---|---|---|
| Area Features | 0 | 0 | 0 |
| Buildings | 0 | 0 | 0 |
| Conflated Buildings | NA | NA | 0 |

## 4.5 Summary of Linear Features

Linear features are the collection of OSM attributes that have a path-like structure and do not make up an area. Hootenanny supports the highways type (described in an earlier section) to date. The statistics presented in this section show the linear features, the portion of those that are highways, and portion of highways that are conflated. The bar graph shows the aggregate length of all the features within each category. The length provides an estimate metric for how much effort was performed by the tool. The measurements are recorded in kilometers. The table shows the length measurements for the two inputs and conflated output for each category (where appropriate).

distancesHistogram.png

| | **ToyTestA.osm** | **ToyTestB.osm** | **Conflated file** |
|---|---|---|---|
| Linear Features | 1.2684 | 1.19957 | 1.27796 |
| Highways | 1.2684 | 1.19957 | 1.27796 |
| Conflated Highways | NA | NA | 1.18724 |

## 4.6 Summary of POI Features

Points of Interest (POIs) are defined as nodes in Hootenanny, and are typically derived from the OSM arribute "amenity", which may contain features like: "place of worship", "cafe", etc. The statistics presented in this section show the total number of POIs for the two inputs and conflated output, and then the number of conflated POIs in the output. The bar graph shows the number for each category. The table shows the counts for the inputs and output (where appropriate).

poiHistogram.png

| | **ToyTestA.osm** | **ToyTestB.osm** | **Conflated file** |
|---|---|---|---|
| POIs | 0 | 0 | 0 |
| Conflated POIs | NA | NA | 0 |

## 4.7 Summary of Unique Names

This section shows the number of unique names that are in the two inputs and conflated output across three categories of features. The three categories shown are: all features, highways, and buildings. The bar graph shows the number of names for each category. The table shows the specific counts for the categories (where appropriate).

uniqueHistogram.png

| | **ToyTestA.osm** | **ToyTestB.osm** | **Conflated file** |
|---|---|---|---|
| All Names | 0 | 0 | 0 |
| Building Names | 0 | 0 | 0 |
| Highway Names | 0 | 0 | 0 |

## 4.8 Summary of Tags

Tags describe specific features of the OSM data elements *nodes*, *ways*, and *relations* (defined in an earlier section). Tags are classified in this section as information and metadata. The metadata tags are the ones containing data about the feature's provenance. For example, who created it, where it came from, UUID, etc. The information tags contain the general information associated with the features (basically all the rest of the attributes). The bar graph shows the total number of tags and the breakdown of the number of information and metadata tags for the two inputs and conflated output datasets. The table shows the specific counts for each category.

tagCountHistogram1.png

| | **ToyTestA.osm** | **ToyTestB.osm** | **Conflated file** |
|---|---|---|---|
| All Tags | 8 | 4 | 22 |
| Information Tags | 8 | 4 | 14 |
| Metadata Tags | NA | NA | 8 |

## 4.9 Summary of Translated Tags

Translated tags are defined as the tags produced from a conversion from a different data source type (e.g. Shapefile). In the conversion, the attributes of the input data source are converted to OSM+ tags. The translation to the OSM+ feature may only minimally need a portion of the input attributes to fill the requirements it needs. For example, a shapefile may have ten different attributes, but the same type of feature in OSM+ might only require or define five of those attributes. This section shows three types of translated tags: populated, default, and null tags. The default tags are those described by the definition above. The populated tags are those where a tag is generated and populated for an unassigned attribute. Null tags are those generated for unassigned attributes that are not populated. A summary of the number of tags for the three tag types across the two inputs and conflated output are shown in the bar graph. The specific values for each of the tag types is shown in the table.

tagCountHistogram2.png

| | **ToyTestA.osm** | **ToyTestB.osm** | **Conflated file** |
|---|---|---|---|
| Translated Populated Tags | 16 | 12 | 62 |
| Translated Default Tags | 184 | 188 | 370 |
| Translated Null Tags | 276 | 276 | 552 |

# 5 Hootenanny Run-time Performance

A summary of the timing for each of the processing steps of the conflation are shown in the bar graph. The processing components measured are: read inputs, stats for input 1 and 2, stats for output, apply named operations, apply pre-operations, projections to planar, find matches, optimize matches, create mergers, apply mergers, apply post operations, conflation computations, projections to WGS84, and the write output times. Time measurements are shown in seconds on the bar graph and the bars are ordered from left to right based on the label ordering (top-to-bottom) shown on the right. The table shows precise details for the processing steps in minutes, seconds, and milliseconds. It also shows some aggregate timing measurements: old road conflation (if applicable), conflation (aggregate of the core processing steps), and total (aggregate of all the steps). The category *other* is the residual timing information.

timingHistogram1.png

| **Processing Step** | **Time** |
|---|---|
| Read Inputs | 0m 0s 2ms ( 0.1% ) |
| Stats for Input 1 | 0m 0s 371ms ( 16.2% ) |
| Stats for Input 2 | 0m 0s 286ms ( 12.5% ) |
| Stats for Output | 0m 0s 357ms ( 15.6% ) |
| Apply Named Ops | 0m 0s 0ms ( 0% ) |
| Apply Pre Ops | 0m 0s 0ms ( 0% ) |
| Project to Planar | 0m 0s 0ms ( 0% ) |
| Find Matches | 0m 0s 22ms ( 1% ) |
| Optimize Matches | 0m 0s 1ms ( 0% ) |
| Create Mergers | 0m 0s 0ms ( 0% ) |
| Apply Mergers | 0m 0s 5ms ( 0.2% ) |
| Apply Post Ops | 0m 0s 1ms ( 0% ) |
| Project to WGS84 | 0m 0s 0ms ( 0% ) |
| Write Output | 0m 0s 1ms ( 0% ) |
| Old Road Conflation | Not applicable |
| Other | 0m 1s 239ms ( 54.2% ) |
| Conflation (aggregate of several steps) | 0m 1s 267ms ( 55.5% ) |
| Total | 0m 2s 283ms |