OSM Quality Assurance
thru cross checking statistics and
external datasets

Christian Quest - @cq94
cquest@openstreetmap.fr
The missing needles in the haystack

Found It!!!

Congratulations, it only took you 65298 seconds
A larger and larger haystack...
False perception of coverage

A visually « full » map hides potential missing data...

Something missing here
False perception of coverage

Confirmed by comparing with external data...
Using external data

- Some examples of useful data in France:
  - POI statistics by municipality « Base Permanente des Equipements » from INSEE (National Statistics)
    - tells for example how many restaurants in 1 city
    - problem: freshness of data…
  - Population statistics on 200m squares
    - tells where people live
    - should match with housings
    - should match with nearby roads and streets
« QA rendering »

→ Uses population data on 200m squares
« QA rendering »

Uses population data on 200m squares.
« QA rendering »

+ comparing street names counts with opendata
+ show highway=* without name=*
Other source: addresses data

BANO project: more than 15 millions address nodes

BANO = Open National Address Database

OpenStreetMap
opendata
cadastre + OSM
cadastre only

We can improve OSM on all red areas!
"BANO rendering"

Example of missing roads/streets…

→ needs tracing
Example of missing roads/streets name…
Visible effect on contributions...

Population n’ayant pas de route avec un tag name dans OSM en France métropolitaine issue du carroyage Insee - type de cadastre (image/vecteur)

BANO start...
New street name rate doubled!
Visible effect on contributions...

Milliers de Km de voie avec name dans OSM en France métropolitaine

- **Voies Rapide et Rn**
  - Y-axis range: 20 to 100
  - Data points showing a trend over time.

- **Rd**
  - Y-axis range: 13.5 to 17.5
  - Data points showing a trend over time.

- **Locale**
  - Y-axis range: 200 to 250
  - Data points showing a trend over time.

- **Autre**
  - Y-axis range: 48 to 62
  - Data points showing a trend over time.

**Date**
- Data points are marked with arrows indicating trends.

SOTM-2014 Buenos-Aires
Osmose QA tool

- Also uses several open datasets to show missing OSM data
  - Post Office
  - Museums
  - Historic Monuments
  - Schools
  etc.

- Geocoding needed in many cases...
The trap!

- Reference / external data may be:
  - Wrong
  - Outdated

- Survey needed on case of doubt / difference
Questions?

Christian Quest

OSM : cquest / twitter : @cq94
Email : cquest@openstreetmap.fr