

# How to globally measure OSM quality in a reproducible way?

The *ohsome* framework  
for OpenStreetMap analytics

[ohsome.org](https://ohsome.org)  
[github.com/giscience](https://github.com/giscience)



  
**HeiGIT**

HEIDELBERG INSTITUTE  
FOR GEOINFORMATION  
TECHNOLOGY

Prof. Dr. Alexander Zipf  
[zipf@uni-heidelberg.de](mailto:zipf@uni-heidelberg.de) : [zipf@heigit.org](mailto:zipf@heigit.org)

[HeiGIT.org](https://HeiGIT.org) :: [uni-heidelberg.de/gis](https://uni-heidelberg.de/gis)

# HeiGIT gGmbH : Heidelberg Institute for Geoinformation Technology



## Smart Mobility

Location-based Services  
& Navigation

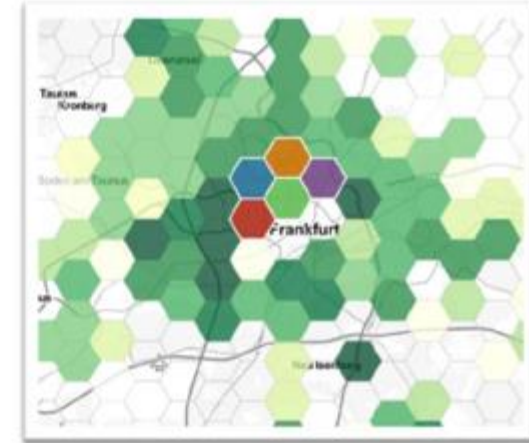
- e.g.
- Emergency Routing / Navigation
  - Healthy / Pleasant Routing
  - Specialized Wheelchair Routing
  - POI service



## Geoinformation for Humanitarian Aid

Supporting Disaster management  
& Sustainable Development

- e.g.
- Disaster Routing / Logistics
  - Improving OSM data
  - Damage assessment / Exposure Modeling
  - Healthcare Access
  - Mobile Crowdsourcing (MapSwipe)



## Big Spatial Data Analytics

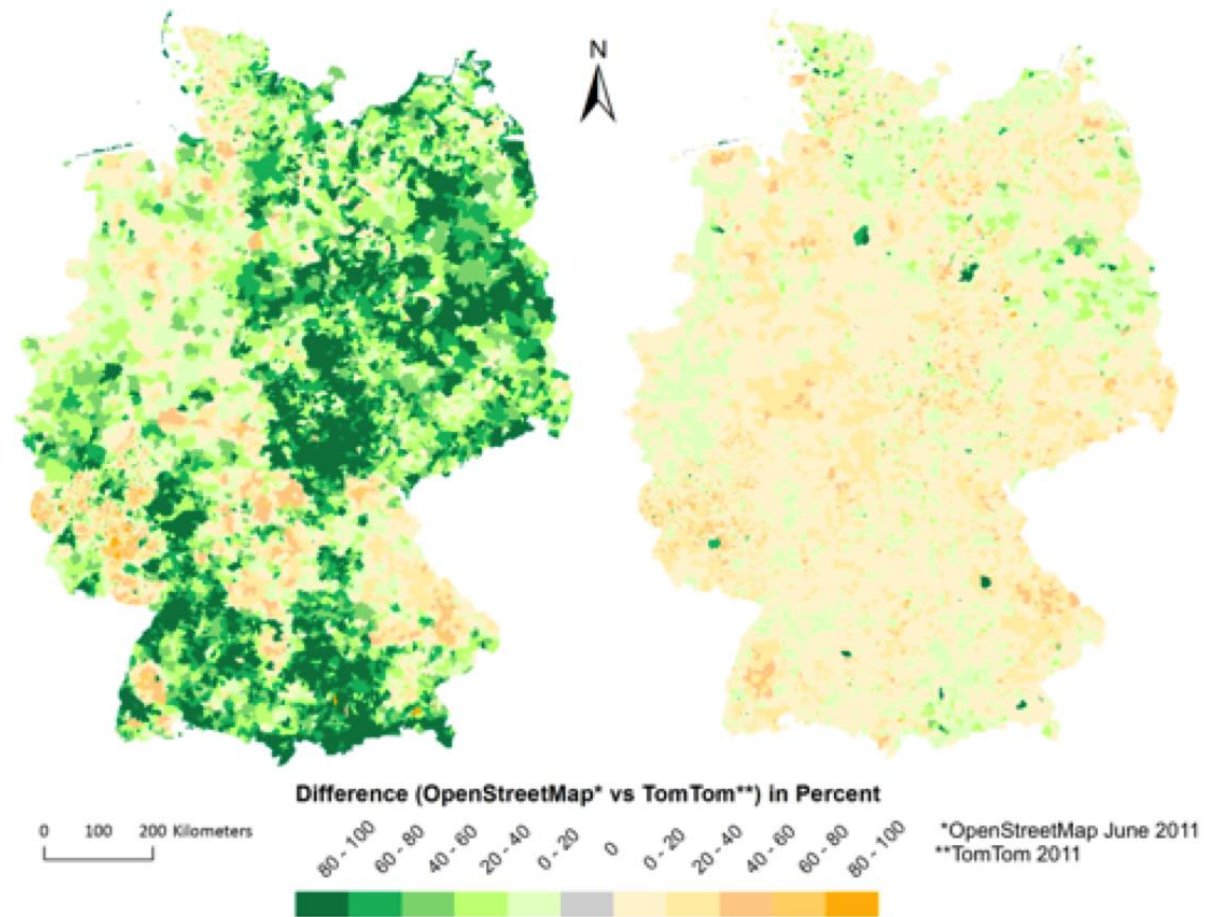
Data Mining &  
Machine Learning  
using open GI

- e.g.
- OSM data quality (analysis & improvement)
  - data products via machine learning
  - Predictive Analytics
  - Maps (e.g. climate protection map)

# OSM Data Quality Analytics

## Data Quality Dimensions

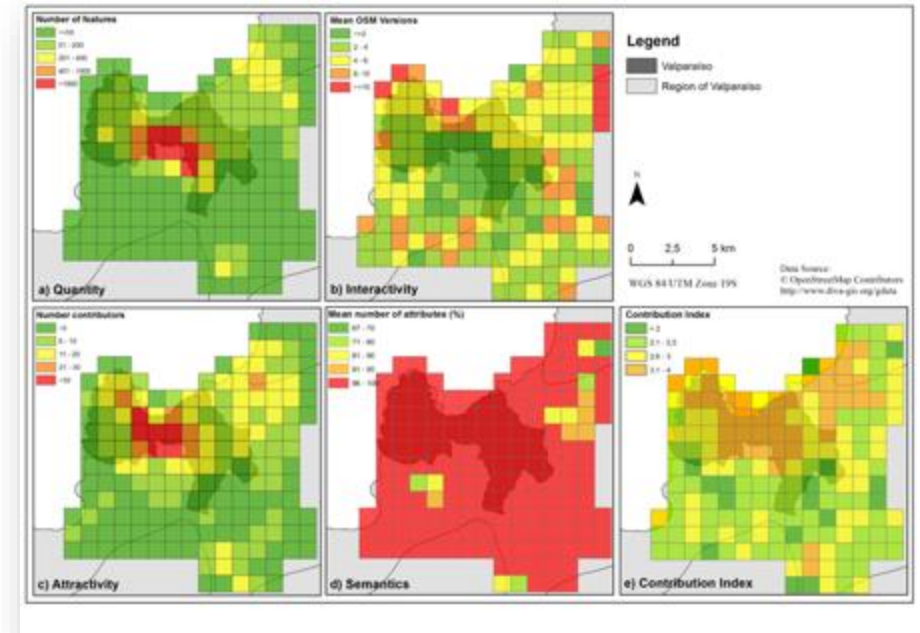
- *Completeness*
- Lineage
- Logical Consistency
- Positional Accuracy
- Attribute Accuracy
- etc...



Neis, P., Zielstra, D. & Zipf, A. (2012): [The Street Network Evolution of Crowdsourced Maps - OpenStreetMap in Germany 2007-2011](#). Future Internet.



# iOSManalyser: Intrinsic Quality Indicators



Example: User Contribution Index

Barron, C., Neis, P. & Zipf, A. (2013):  
**A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis.**  
 Transactions in GIS . DOI:10.1111/tgis.12073.

Jokar A., J., Mooney, P., Helbich, M., Zipf, A.,  
 (2015): **An exploration of future patterns of the  
 contributions to OpenStreetMap and development  
 of a Contribution Index,** Transactions in GIS.  
 DOI:10.1111/tgis.12139.

Heterogeneous OSM quality (analytics)

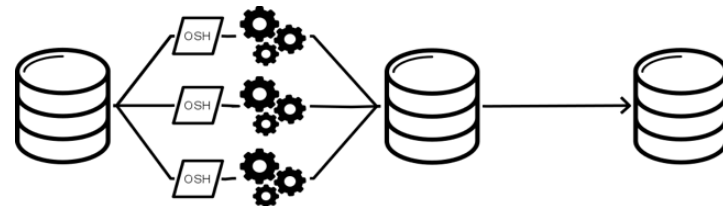
- fragmented small studies
- heterogeneity of methods / tools
- no replicability

=> Need for a **framework** supporting **global** OSM analysis

**Reproducible & sustainable**

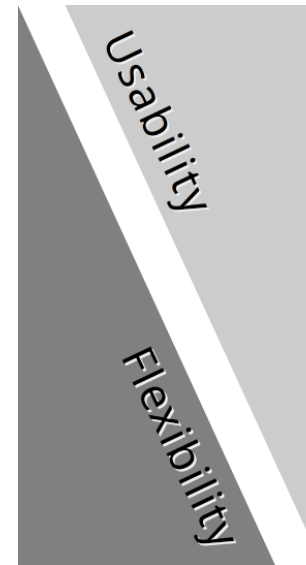
=> *OSM data quality analytics framework*

## ohsome.org



Processing using Big Data Frameworks  
Apache Ignite, Spark etc.

## OpenStreetMap History Data Analyzing Platform



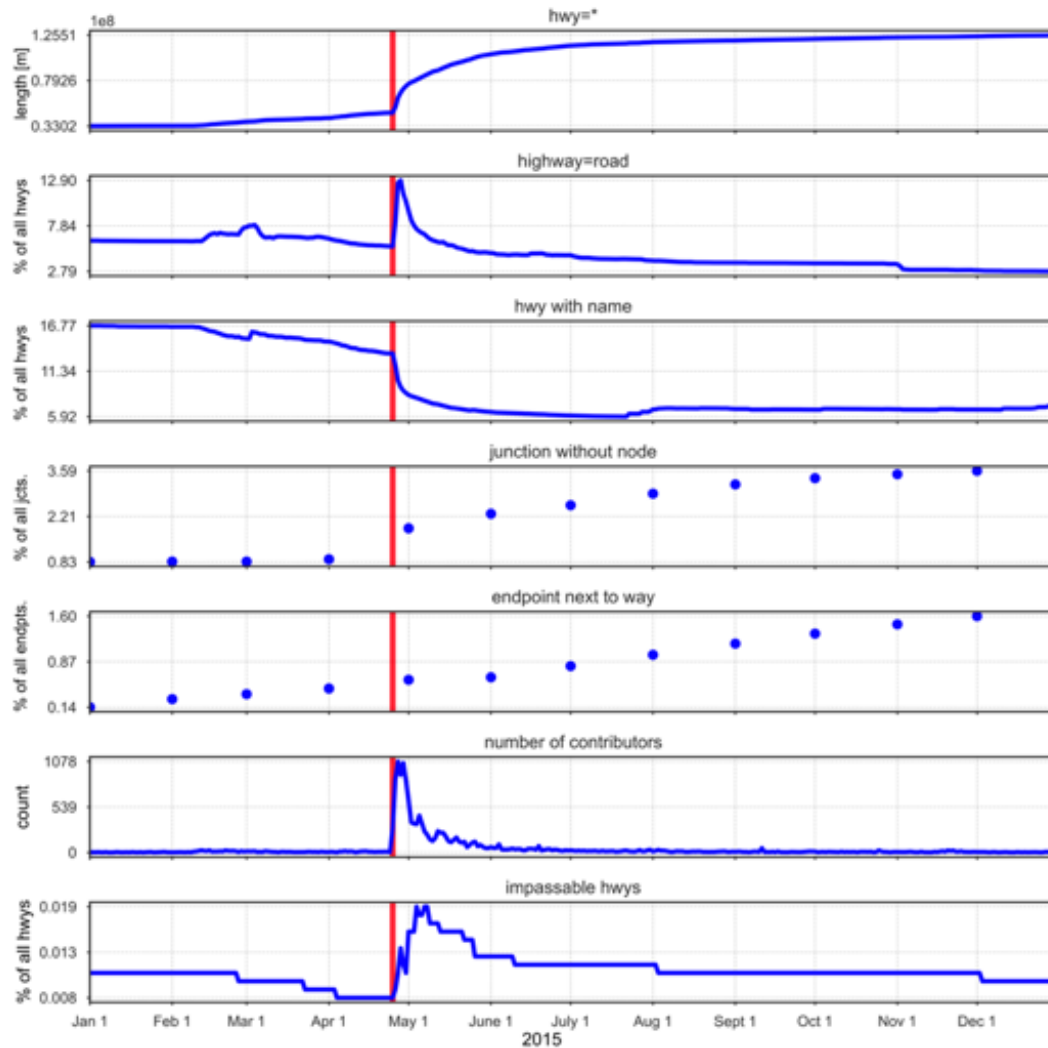
- ohsome quality analyst
- ohsome OSM History Explorer
- ohsome dashboard
- ohsome2label / QGIS
- ohsome-py / ohsome-R
- ohsome API
- OSHDB API
- OSHDB

A vertical stack of icons representing the platform's components: a screenshot of the OSM History Explorer interface, a screenshot of the OSM History Explorer dashboard, a screenshot of the OSM History Explorer map view, a screenshot of the OSM History Explorer API, a blue hexagonal icon with 'HTTP', a Java logo, and a cylinder icon labeled 'OSM History'.

Raifer, M, Troilo, R, Kowatsch, F, Auer, M, Loos, L, Marx, S, Przybill, K, Fendrich, S, Mocnik, FB & Zipf, A (2019):

[OSHDB: a framework for spatio-temporal analysis of OpenStreetMap history data](#). Open Geospatial Data, Software & Standards, Springer.

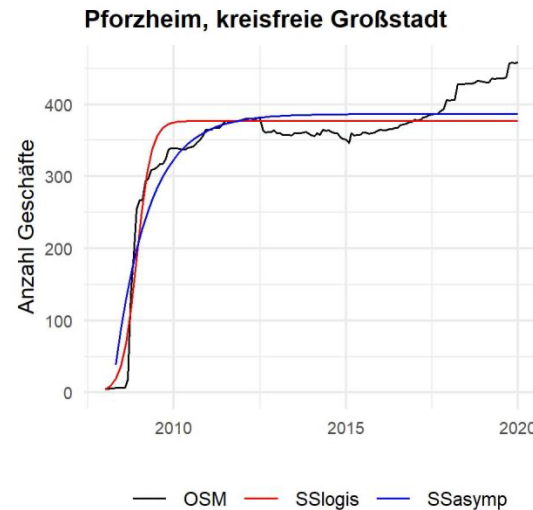
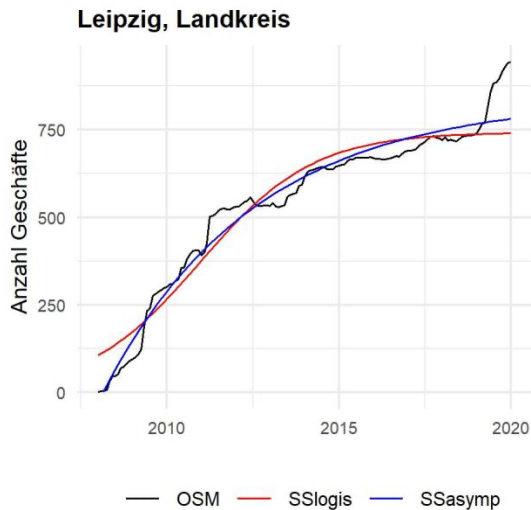
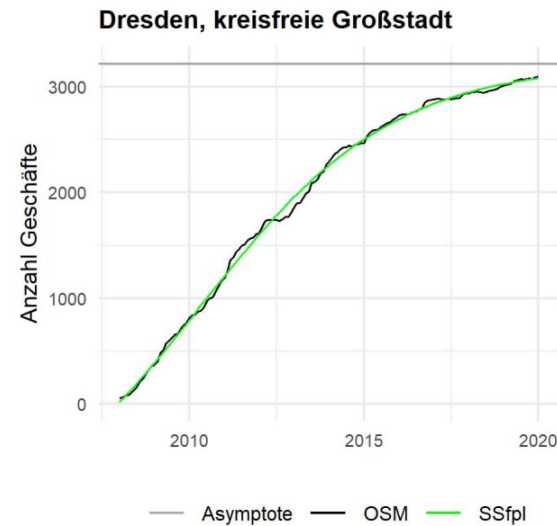
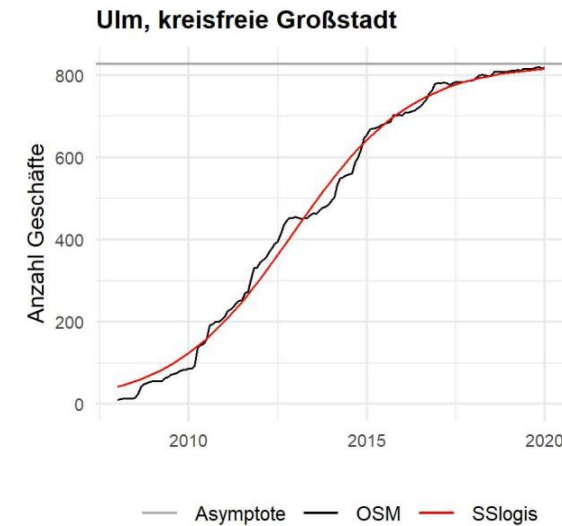
# Nepal earthquake response '15



Intrinsic quality indicators are used to investigate the value of OSM mapping activities for disaster routing

Auer, M.; Eckle, M.; Fendrich, S.; Griesbaum, L.; Kowatsch, F.; Marx, S.; Raifer, M.; Schott, M.; Troilo, R.; Zipf, A. (2018): [Towards Using the Potential of OpenStreetMap History for Disaster Activation Monitoring](#). ISCRAM 2018. USA.

# Fitting different type of saturation curves



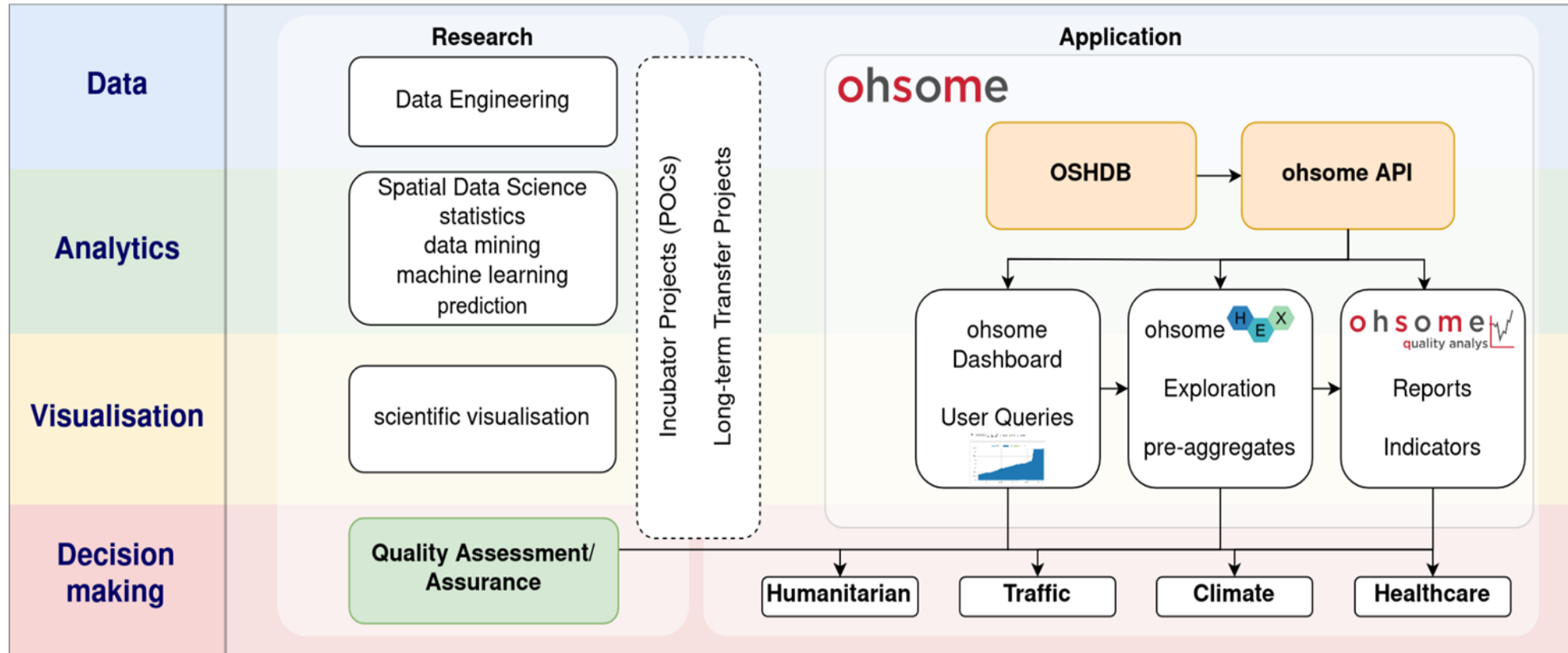
Modell	Verlaufsform	Funktion
SSlogis	beschränktes, logistisches Wachstum	$f(input) = \frac{Asym}{1 + e^{\frac{x_{mid}-input}{scal}}}$
SSfpl	beschränktes, logistisches Wachstum	$f(input) = \frac{A + (B - A)}{1 + e^{\frac{x_{mid}-input}{scal}}}$
SSmicmen	beschränktes Wachstum (Michaelis Menten)	$f(input) = \frac{v_m \times input}{(K + input)}$
SSasymp	beschränktes Wachstum	$f(input) = Asym + (R_o - Asym) \times e^{(-exp(trc) \times input)}$

	SSlogis	SSasymp
<b>Residualstandardfehler, relativ</b>	0,06	0,04
<b>Steigung (Abweichung) [Geschäfte/Jahr]</b>	3	8
<b>Asymptote</b>	585	718
<b>Vollständigkeit [%]</b>	93	76

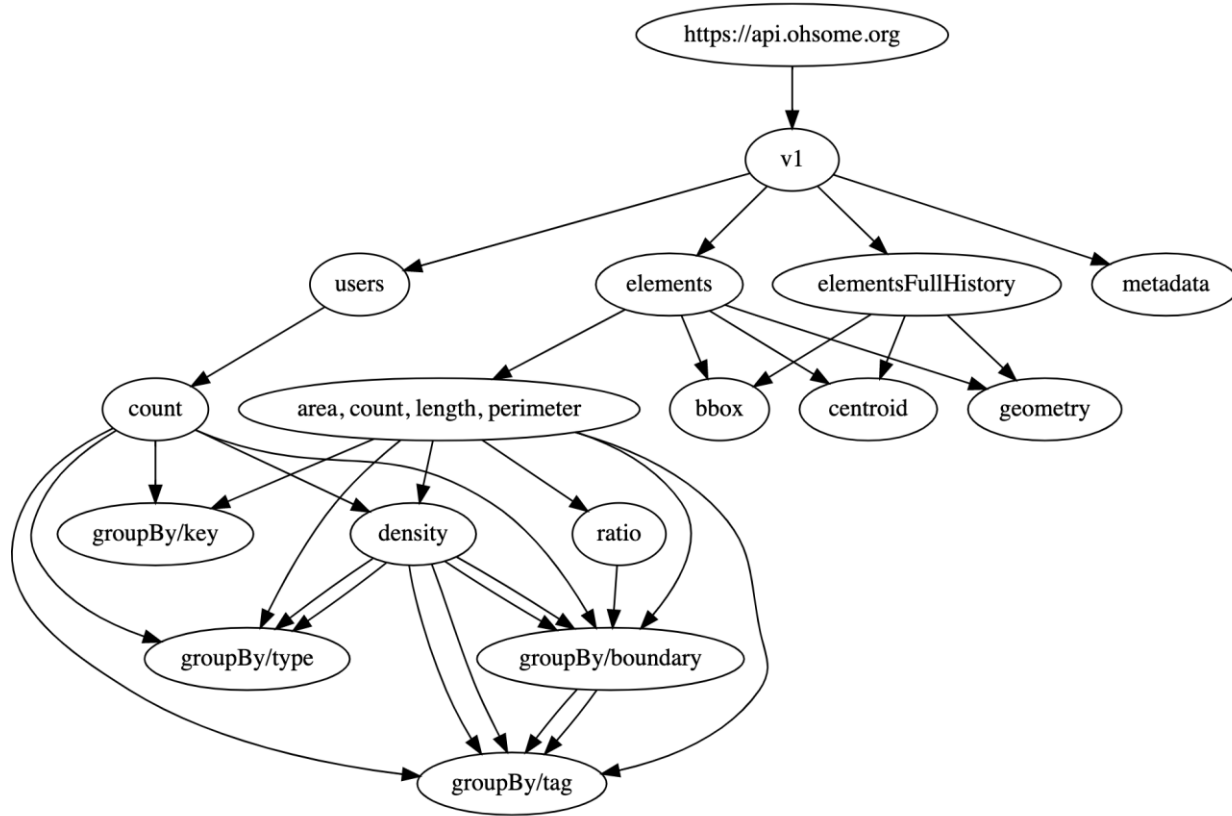
Brückner, J., Schott, M., Zipf, A., Lautenbach, S. (2021): [“Assessing shop completeness in OpenStreetMap for two federal states in Germany.”](#), AGILE GIScience Series. 2(20): 1-7. <https://doi.org/10.5194/agile-giss-2-20-2021>, 2021.

# Our approach

*“empower society to take spatially informed decisions”*







## [api.ohsome.org](https://api.ohsome.org)

docs.ohsome.org/ohsome-api/v1/endpoints.html#aggregation-endpoints

### ohsome API

v1

Search docs

#### Aggregation Endpoints

**POST /elements/(aggregation)**

Get **aggregation** of OSM elements.

- aggregation type: one of **area**, **count**, **length**, **perimeter**

**Query Parameters:**

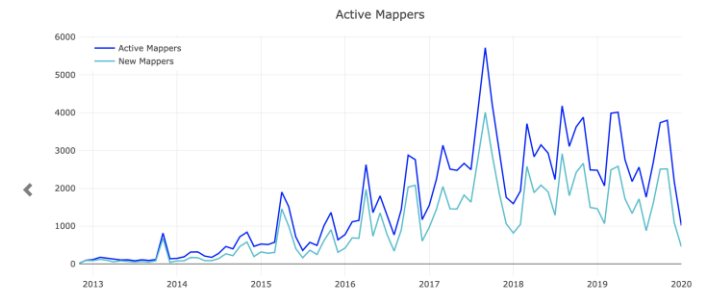
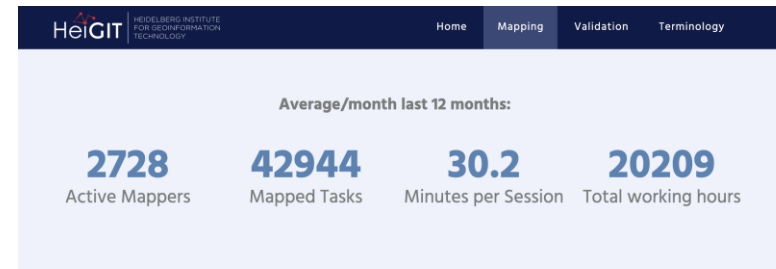
- <boundary>** - One of these boundary parameters: **bboxes**, **bcircles**, **bpolys**. See **boundaries**
- time** - ISO-8601 conform timestring(s); default: latest timestamp in the OSHDB, see **time**
- filter** - combines several attributive filters: OSM type, geometry (simple feature) type, as well as the OSM tag; See **filter**
- format** - 'json' or 'csv'; default: 'json'
- showMetadata** - add additional metadata information to the response: 'true', 'false', 'yes', 'no'; default: 'false'
- timeout** - custom timeout to limit the processing time in seconds; default: dependent on server settings, retrievable via the /metadata request
- types** - Deprecated! Use **filter** parameter instead! Old parameter which allowed to specify OSM type(s) 'node' and/or 'way' and/or 'relation' OR simple feature type(s) 'point' and/or 'line' and/or 'polygon' and/or 'other'; default: all three OSM types
- keys** - Deprecated! Use **filter** parameter instead! Old parameter which allowed to specify OSM key(s) given as a list and combined with the 'AND' operator; default: empty
- values** - Deprecated! Use **filter** parameter instead! Old parameter which allowed to specify OSM value(s) given as a list and combined with the 'AND' operator; values(n) MUST fit to keys(n); default: empty

**CONTENTS:**

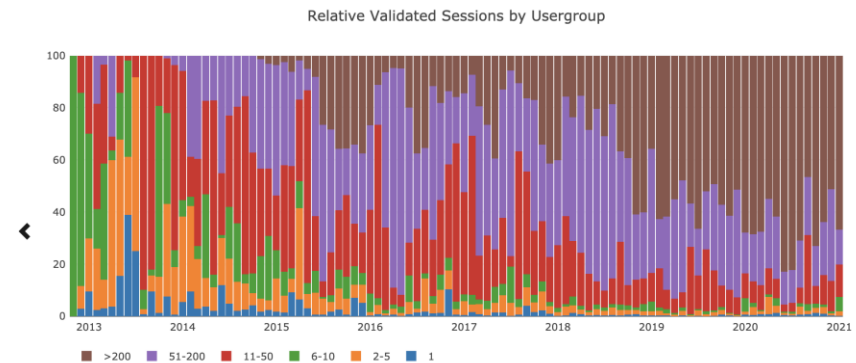
- API Endpoints
  - Aggregation Endpoints
  - Users Aggregation Endpoints
  - Extraction Endpoints
  - Contribution Endpoints
  - Metadata Endpoint
- API Endpoints Visualisations
  - Boundaries
  - Grouping
  - Time
  - Filter
  - Response Parameters
  - HTTP Response Status

# humstats.heigit.org

- Effectiveness of humanitarian mapping campaigns
- Tasking Manager Analytics
- e.g. development of
  - # contributors
  - TM sessions
  - first time contributors
  - hours spend
  - ratio % mapped / % validated
  - survival rate of users
  - Validated session by user group (relative / absolute)
  - leaderboard
  - ...



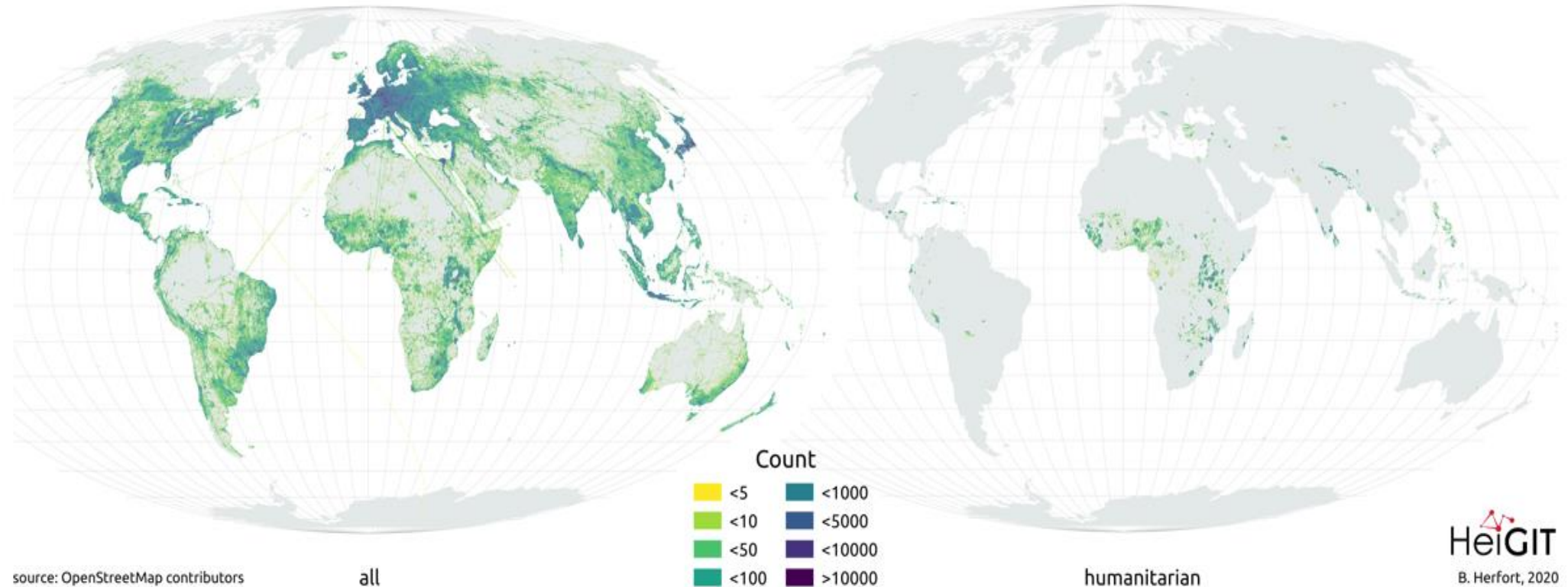
While there are more regularly active Mappers than in the early 2010s the Mapper Community does not seem to be growing. While there are regularly more than 2500 new mappers per month only around 1000 longer term mappers are active every month.



To accentuate the impact of the different usergroups this plot shows how much of the total progress in each month was done by each usergroup.

# OSM humanitarian mapping: spatial distribution & bias

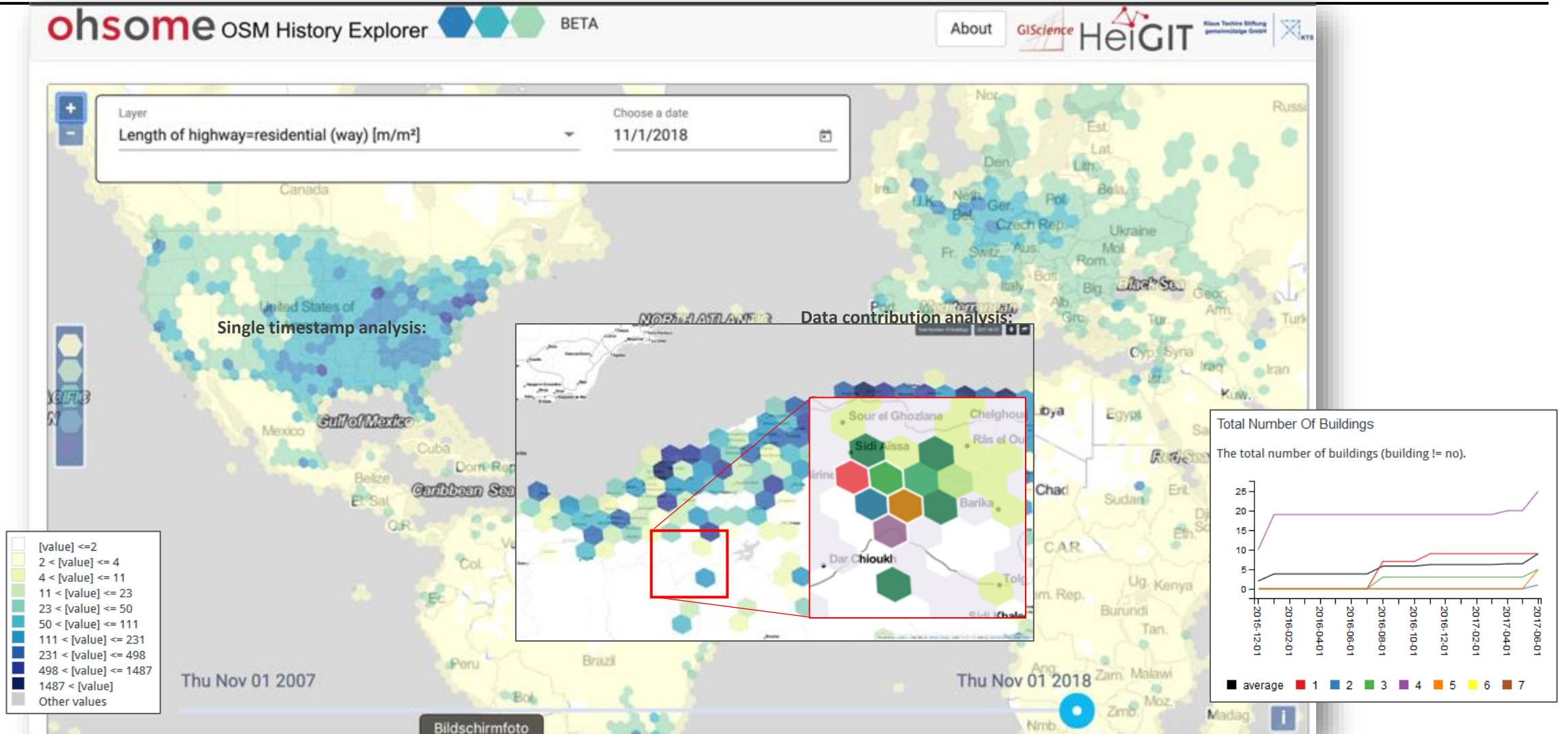
Highway Contributions (creation) in OpenStreetMap since 2008-01-01



Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., Zipf, A. (2021): [The evolution of humanitarian mapping within the OpenStreetMap community](https://doi.org/10.1038/s41598-021-82404-z). *Scientific Reports* 11, 3037 (2021). DOI: 10.1038/s41598-021-82404-z

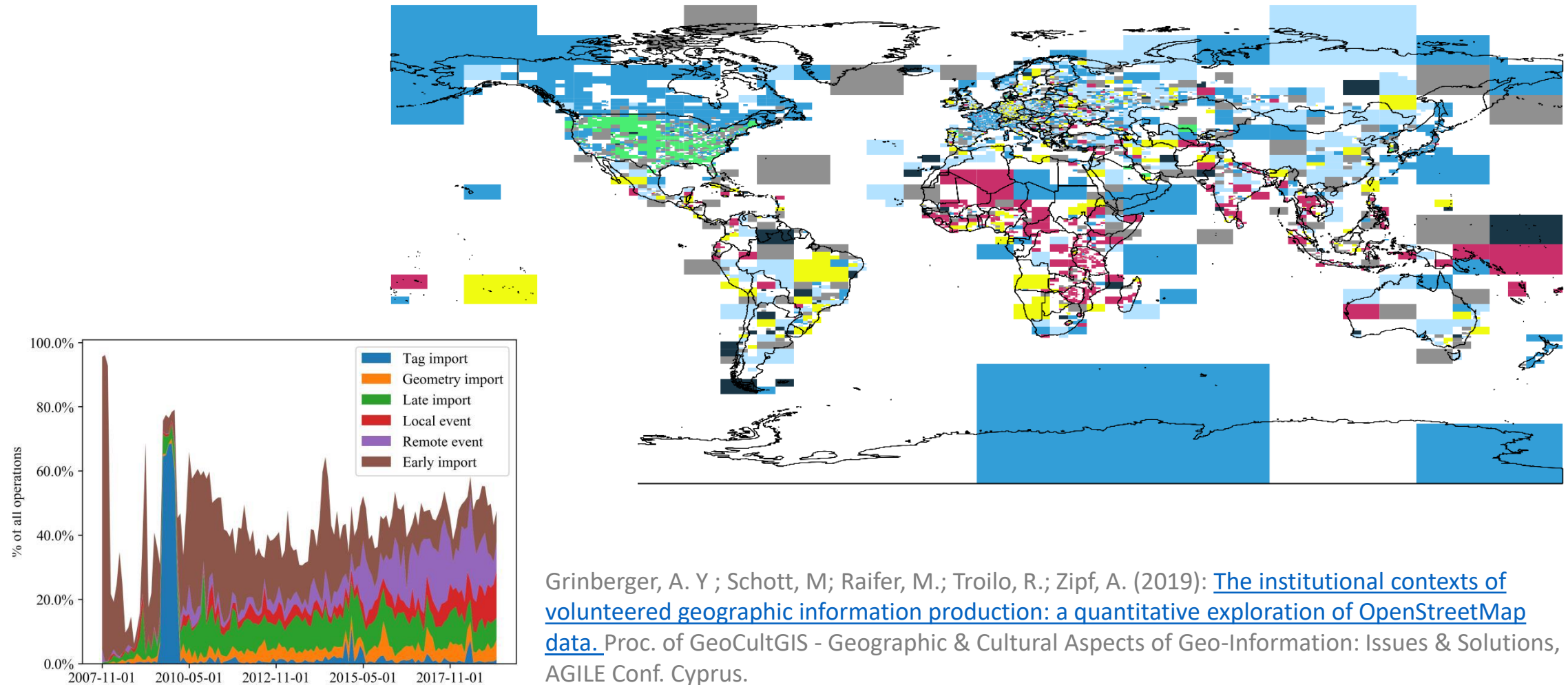
[nature.com/articles/s41598-021-82404-z](https://nature.com/articles/s41598-021-82404-z)

# ohsomeHeX: ohsome OSM History Explorer





# Classifying types of events detected in OSM




Grinberger, A. Y ; Schott, M; Raifer, M.; Troilo, R.; Zipf, A. (2019): [The institutional contexts of volunteered geographic information production: a quantitative exploration of OpenStreetMap data](#). Proc. of GeoCultGIS - Geographic & Cultural Aspects of Geo-Information: Issues & Solutions, AGILE Conf. Cyprus.

Figure 4. The share of event contributions out of all contributions over time, by event type

# Ohsome Quality analysT (OQT)

1. Pick an area on the map.



2. Choose data quality topic.

Simple Report


3. Run analysis.

GET QUALITY REPORT

- customizable reports
- combination of specific quality indicators
- web application & command-line

## Results

### Overall Data Quality Report



Report: simple-report

Good Quality

All indicators show a good quality. The data in this regions seems to be completely mapped.

Report description:  
This report shows the quality for two indicators: mapping-saturation and ghspop-comparison. It's main function is to test the interactions between database, api and website.

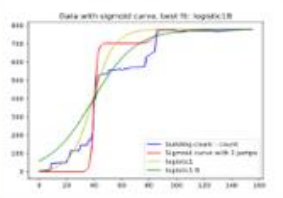
### Data Quality Indicators

#### mapping-saturation for building-count

Good Quality

The saturation for the last 3 years is 1.0. Saturation has been reached. The data in this region seem quite saturated with a growth of data less than 3 % within the last 3 years. This indicates good quality in respect to completeness.

Indicator description:  
Calculate if mapping has saturated.

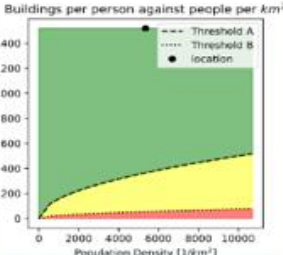


#### ghspop-comparison for building-count

Good Quality

Following the GHS POP dataset, there are 2747 people living in an area of 0.51 sqkm, which results in a population density 5352.69 of people per sqkm. In OSM there are 1519.39 buildings per sqkm mapped. For the given population density, this is a relatively high value and indicates a good data quality in terms of completeness.

Indicator description:  
Comparison between population density and feature density. This can give an estimate if mapping has been completed.



BACK TO TOP

# ohsome quality analyst (OQT)

---

Data quality has to be defined *together with the users* (co-production)

→ **fitness for purpose**

→ [open source: Github.com/GIScience](https://github.com/GIScience)

We develop **indicators** based on the requirements of users:

- define list of OSM features
- discuss thresholds for indicators
- check additional extrinsic/  
reference data sources
- compare to research results

## Growing list of Data Quality Indicators

### *Intrinsic:*

- mapping saturation
- last edit
- tag ratio
- points of interest density
- ...

### *Extrinsic:*

- global human settlement layer comparison for roads & buildings
- global urban footprint comparison for roads & buildings (WIP)
- ...

# ohsome quality analyst (OQT)

---

Data quality **reports** *combine individual indicators*  
(like in a recipe)

- you can combine intrinsic & extrinsic indicators
- custom weighting for individual indicators (*WIP*)
- You can derive reports for custom input geometries (e.g. a list polygons in a geojson file)

## First Data Quality Reports:

- simple report for buildings
- remote mapping report for buildings and highways
- sketch mapping report
- JRC report
- MapAction Country Overview Report
- IDEAL VGI report for land use features (*wip*)
- ...

## Further Plans (WIP) e.g.:

### *Processing:*

- extend coverage
- frequent updates
- integration with ohsomeHex
- ...

### *Indicators & Reports:*

- investigate regional differences
- provide locally adjusted versions of existing indicators
- user experience indicator
- ...

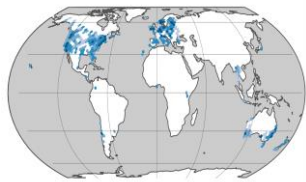


# Public transportation Completeness - Regression Model

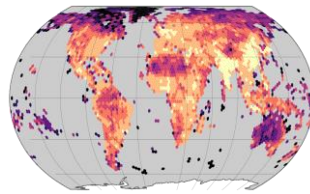
(in ohsomeHEX)

Predict number of bus routes

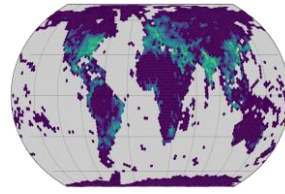
Input feature space includes many datasets, e.g.



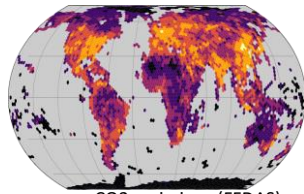
Bus route data  
from publicly available data sets



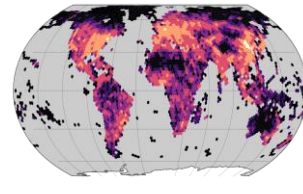
Population (WorldPop)



Night-time light(Black Marble)



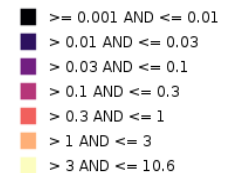
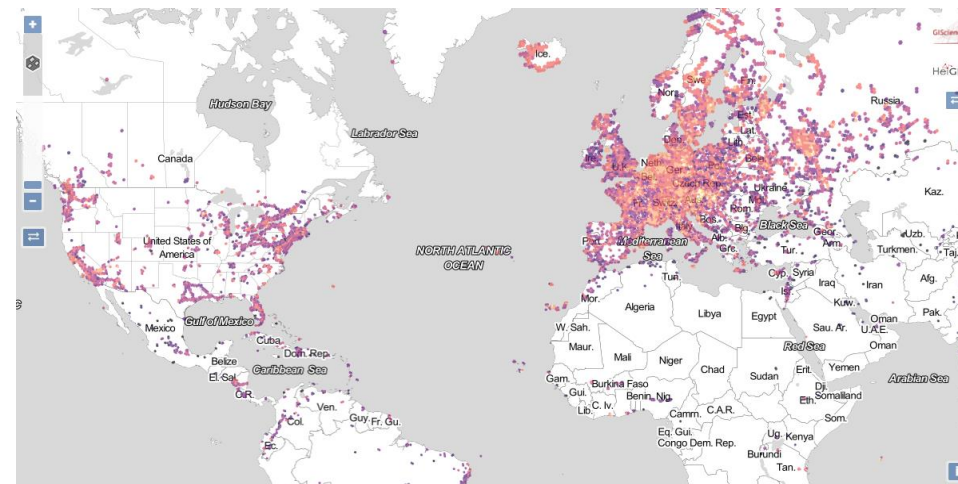
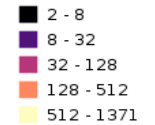
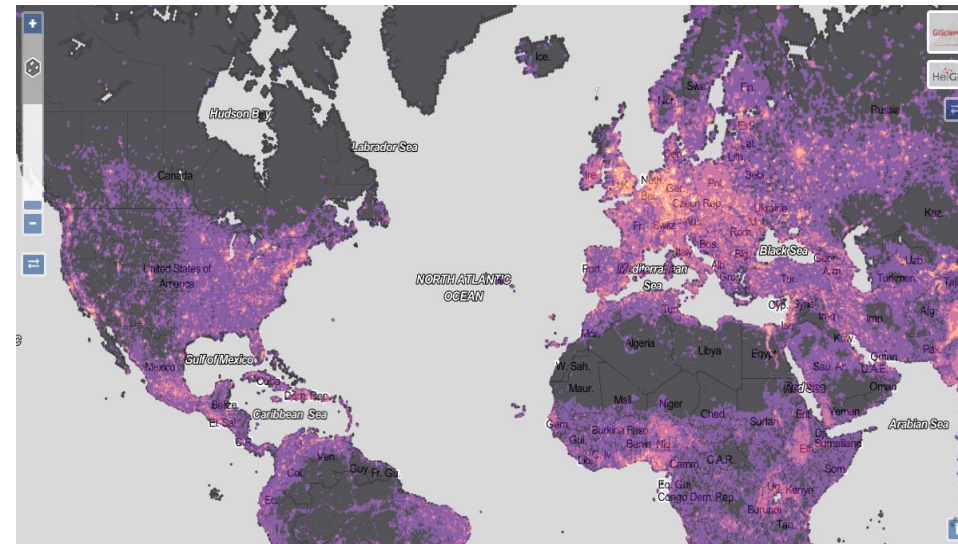
CO2 emissions (FFDAS)



Built-up Density (GHSL)

*Ratio of bus routes in OSM to prediction  
(indicator of feature completeness)*

Fritz, O., Auer, M., Zipf, A.. 2021. Entwicklung eines Regressionsmodells für die Vollständigkeitsanalyse des globalen OpenStreetMap-Datenbestands an Nahverkehrs-Busstrecken. *AGIT – Journal Für Angewandte Geoinformatik*. 7-2021.



# ohsome2label

Provide OSM data as ML  
training samples

## Various image sources

- Google, Bing, Mapbox, Sentinel..

## Supported deep learning models

- object detection
- semantic segmentation
- instance segmentation

## OSM Quality Measurement

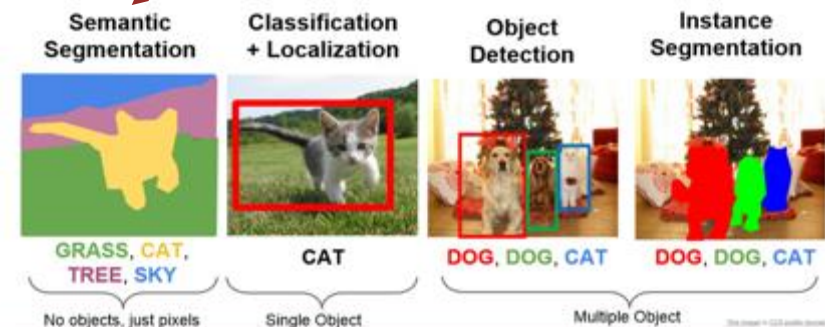
### Result:

- Microsoft.COCO Json

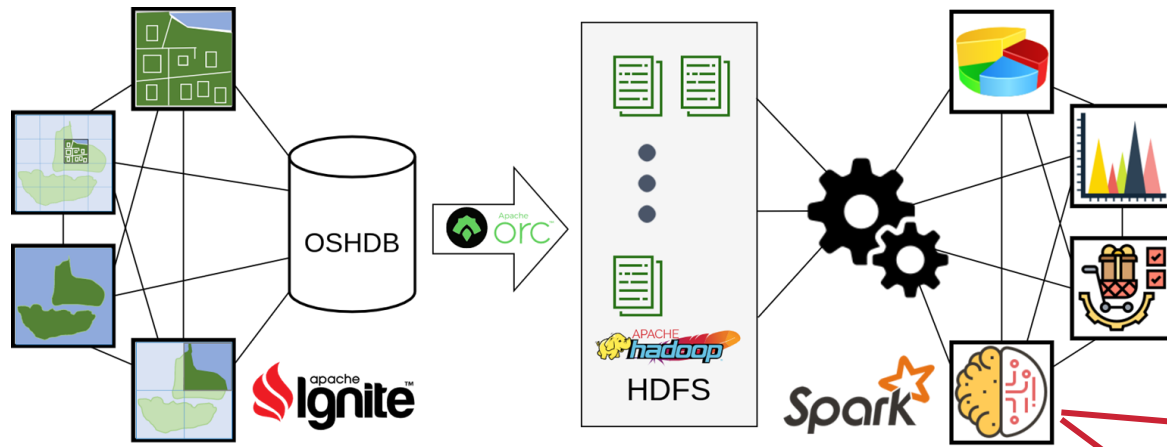
Wu, Z., Li, H., & Zipf, A. (2020): From Historical OpenStreetMap data to customized training samples for geospatial machine learning. Academic Track at the State of the Map 2020. <http://doi.org/10.5281/zenodo.3923040>

heigit.org --- uni-heidelberg.de/gis

Scheme	Parameter	Description
project	name	The title of output folder you want to have.
project	workspace	The relative path your want to store the project.
project	time	The created time of the project.
general	bouding_box	The box boundary of target area, which is given in form <code>[xmin, ymin, xmax, ymin]</code> , x and y refer to longitude and latitude. The default map project is WGS84.
general	key_value_pairs	The target key and values pairs of OSM feature, where <code>name</code> could be defined by yourself. A valid OSM key is necessary. Multiple kv paris should be seperated by <code>:</code> .
general	OSM_timestamp	The timestamp of historical OSM data you want to retrieval. The date should be given in <code>[year-month-day]</code>
general	types	The object types you are aimed at, which could be <code>polygon</code> , <code>line</code> .
label	zoom	The zoom-in level of satelliate imagery. This <code>zoom level</code> would affect the spatial resolution in general.
label	ML_task	The type of machine learning taks you would like to use afterwards. This could be <code>object detection</code> , <code>semantic segementation</code> , <code>instance segementation</code> .
image	source	The satelliate imager service you would like to use. Now <code>bing</code> , <code>mapbox</code> , <code>sentinel</code> are supported.
image	API_token	The API token should be applied individually by usders. Please find the corresponding application pages as follows: <code>bing</code> , <code>mapbox</code> , <code>sentinel</code>



# Logical Consistency: OSM Association Mining



## OpenStreetMap Association Mining *OSMam*

- ML based identification of tag association rules
- Cluster exceptions

antecedent	consequent	confidence	lift
addr:country,addr:postcode,addr:housenumber	addr:street	0.9950408565793181	1.74247955280704
addr:country,addr:postcode,addr:housenumber	building	1	1
addr:country,addr:postcode,addr:housenumber	addr:city	0.9977007607776839	1.84555109498020
addr:postcode,addr:street,addr:housenumber	addr:country	0.7950481905377993	1.92962025847392
addr:postcode,addr:street,addr:housenumber	building	1	1
addr:postcode,addr:street,addr:housenumber	addr:city	0.9947640678622433	1.84011878798117
addr:country,addr:street	addr:housenumber	0.9975924571711047	1.74201460665743
addr:country,addr:street	building	1	1
addr:country,addr:street	addr:postcode	0.9866669887874105	1.90151571658642
addr:country,addr:street	addr:city	0.9964802114103221	1.84329331758845
addr:country,addr:street,building	addr:housenumber	0.9975924571711047	1.74201460665743
addr:country,addr:street,building	addr:postcode	0.9866669887874105	1.90151571658642

Ludwig, C.; Fendrich, S.; Zipf, A. (2020): [Regional variations of context-based association rules in OpenStreetMap](#). Transactions in GIS. Wiley. doi.org/10.1111/tgis.12694



# Preview: Feature Space for Learning OSM Quality

---

- Feature Space for Learning OSM Quality
- Need more training data (manual validation!)
- Tested 20+ first indicators, e.g. (but more to come)
  - Object based                      tags
  - Completeness                      coverage
  - Consistency
  - Context                              population density
  - Temporal accuracy                      last update
  - User based                              user experience
  - Changeset based
  - Source                                  import detector
  - Remote sensing                              RS variance

[ohsome.org](https://ohsome.org)

[github.com/giscience](https://github.com/giscience)

preview, work in progress, unpublished



# Thank you! Questions?

---

**Feel free to share your needs & requirements.**

[ohsome.org](https://ohsome.org)  
[github.com/giscience](https://github.com/giscience)

**Ideas for improvement welcome!**

**We love to collaborate!**

**Thank you!**

Prof. Dr. Alexander Zipf  
[zipf@uni-heidelberg.de](mailto:zipf@uni-heidelberg.de)  
[zipf@heigit.org](mailto:zipf@heigit.org)

