

FOSS4G 2018

Dar es Salaam



POLITECNICO
MILANO 1863

An open source approach for the intrinsic assessment of the temporal accuracy, up-to-dateness and lineage of OpenStreetMap

Marco Minghini^{1,2}, Maria Antonia Brovelli², Francesco Frassinelli²

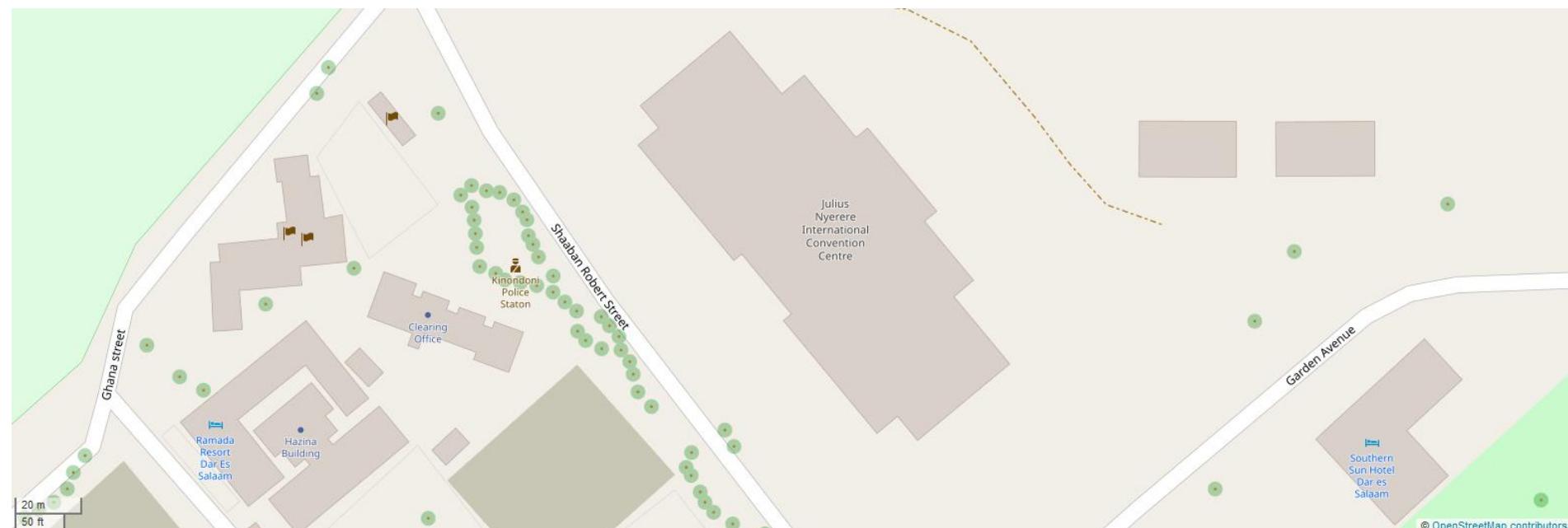
¹ European Commission, Joint Research Center (JRC) - Ispra, Italy

² Dept. of Civil and Environmental Engineering, Politecnico di Milano - Milan, Italy



OpenStreetMap (OSM)

- The most popular [Volunteered Geographic Information \(VGI\)](#) project:
 - started in 2004, currently featuring [4.8 million contributors](#)
 - largest, most detailed, complete & up-to-date global [spatial database](#)
 - available under the [Open Database License \(ODbL\)](#)
 - used by many actors/applications & [studied](#) by researchers



OpenStreetMap (OSM)

- The most popular [Volunteered Geographic Information \(VGI\)](#) project:
 - started in 2004, currently featuring [4.8 million contributors](#)
 - largest, most detailed, complete & up-to-date global [spatial database](#)
 - available under the [Open Database License \(ODbL\)](#)
 - used by many actors/applications & [studied](#) by researchers



OpenStreetMap (OSM) – Data model

- OSM makes use of a [vector data model](#) (geometries + attributes):



POLITECNICO MILANO 1863



OpenStreetMap (OSM) – Data model

- OSM makes use of a **vector data model** (geometries + attributes):
 - geometries
 - **nodes**: single point objects
 - **ways**: ordered lists of nodes (line objects and polygon objects)
 - **relations**: relations between two or more nodes, ways and/or relations



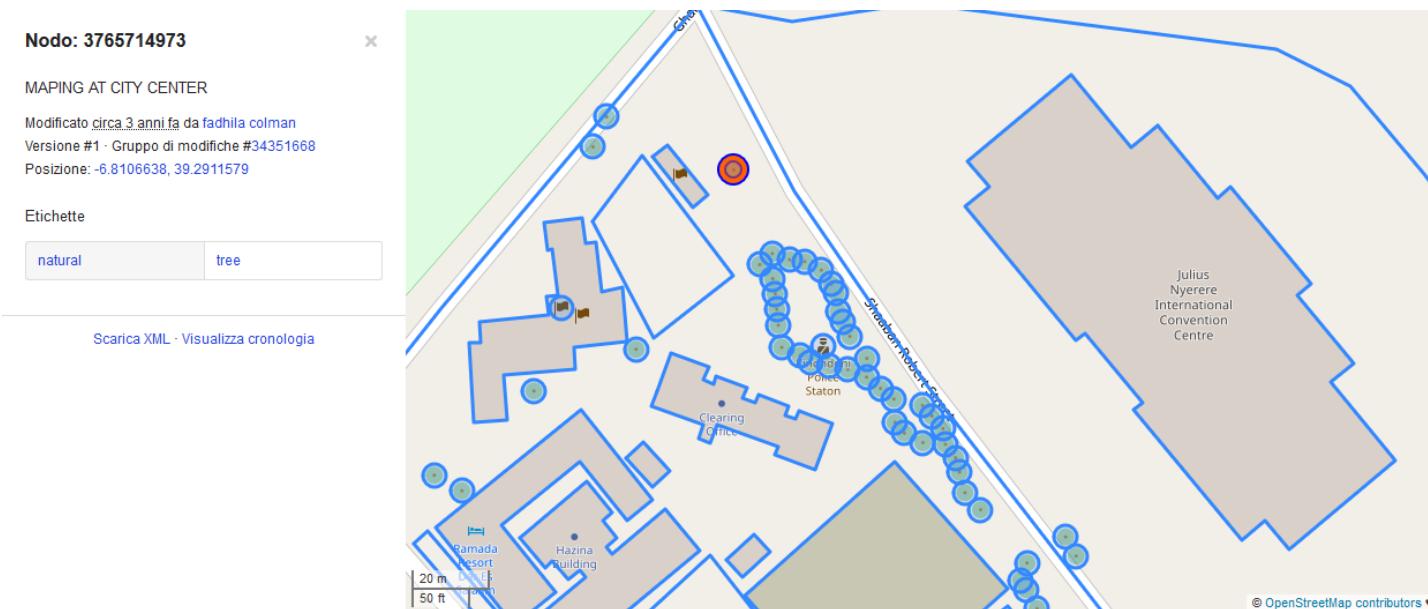
OpenStreetMap (OSM) – Data model

- OSM makes use of a **vector data model** (geometries + attributes):
 - geometries
 - **nodes**: single point objects
 - **ways**: ordered lists of nodes (line objects and polygon objects)
 - **relations**: relations between two or more nodes, ways and/or relations
 - attributes
 - **tags**: key-value pairs



OpenStreetMap (OSM) – Data model

- OSM makes use of a **vector data model** (geometries + attributes):
 - geometries
 - **nodes**: single point objects
 - **ways**: ordered lists of nodes (line objects and polygon objects)
 - **relations**: relations between two or more nodes, ways and/or relations
 - attributes
 - **tags**: key-value pairs



OpenStreetMap (OSM) – Data model

- OSM makes use of a **vector data model** (geometries + attributes):
 - geometries
 - **nodes**: single point objects
 - **ways**: ordered lists of nodes (line objects and polygon objects)
 - **relations**: relations between two or more nodes, ways and/or relations
 - attributes
 - **tags**: key-value pairs



OpenStreetMap (OSM) – Data model

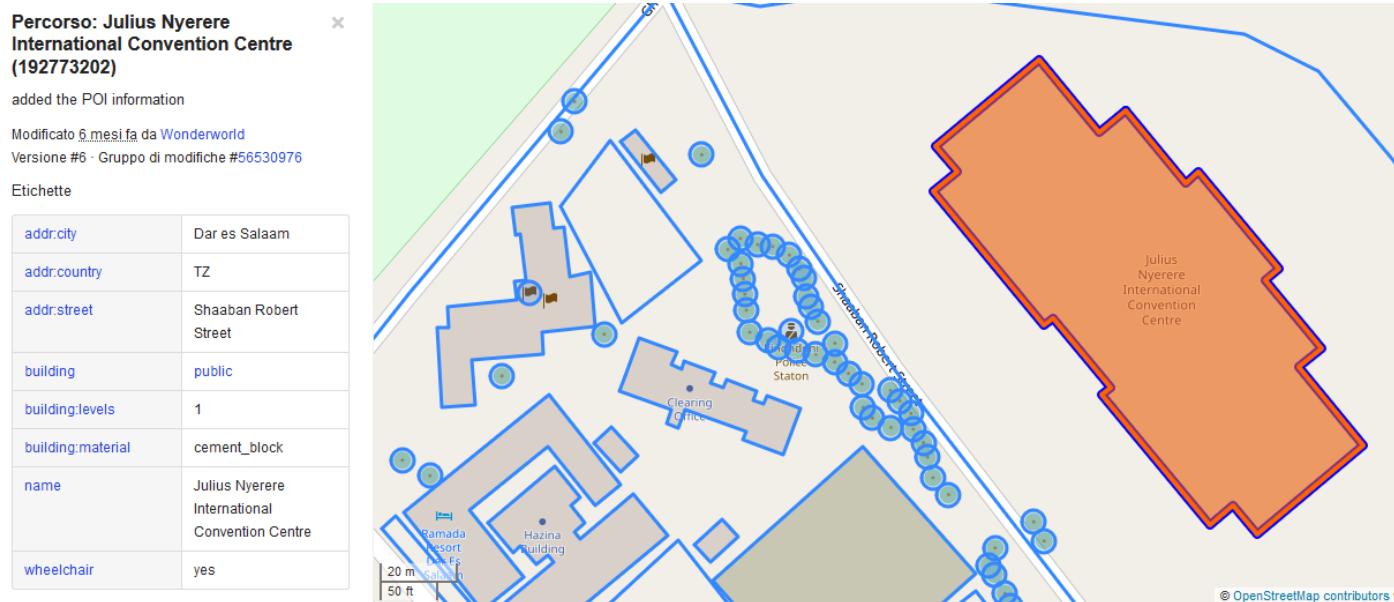
- OSM makes use of a **vector data model** (geometries + attributes):

- **geometries**

- **nodes**: single point objects
 - **ways**: ordered lists of nodes (line objects and polygon objects)
 - **relations**: relations between two or more nodes, ways and/or relations

- **attributes**

- **tags**: key-value pairs



OpenStreetMap quality

- The main **concern** which still limits the widespread use of OSM
- Quality can be assessed through several **parameters**:
 - positional accuracy
 - completeness
 - logical consistency
 - semantic accuracy
 - thematic accuracy
 - temporal accuracy
 - up-to-dateness
 - lineage
 - fitness-for-use & fitness-for-purpose
 - ...



OpenStreetMap quality

- The main **concern** which still limits the widespread use of OSM
- Quality can be assessed through several **parameters**:
 - positional accuracy
 - completeness
 - logical consistency
 - semantic accuracy
 - thematic accuracy
 - temporal accuracy
 - up-to-dateness
 - lineage
 - fitness-for-use & fitness-for-purpose
 - ...
- OSM quality assessment approaches are:
 - **extrinsic**, when OSM is compared to a reference dataset
 - **intrinsic**, when OSM is compared to itself
 - based on **OSM history** (OSM API / Full History Planet File)



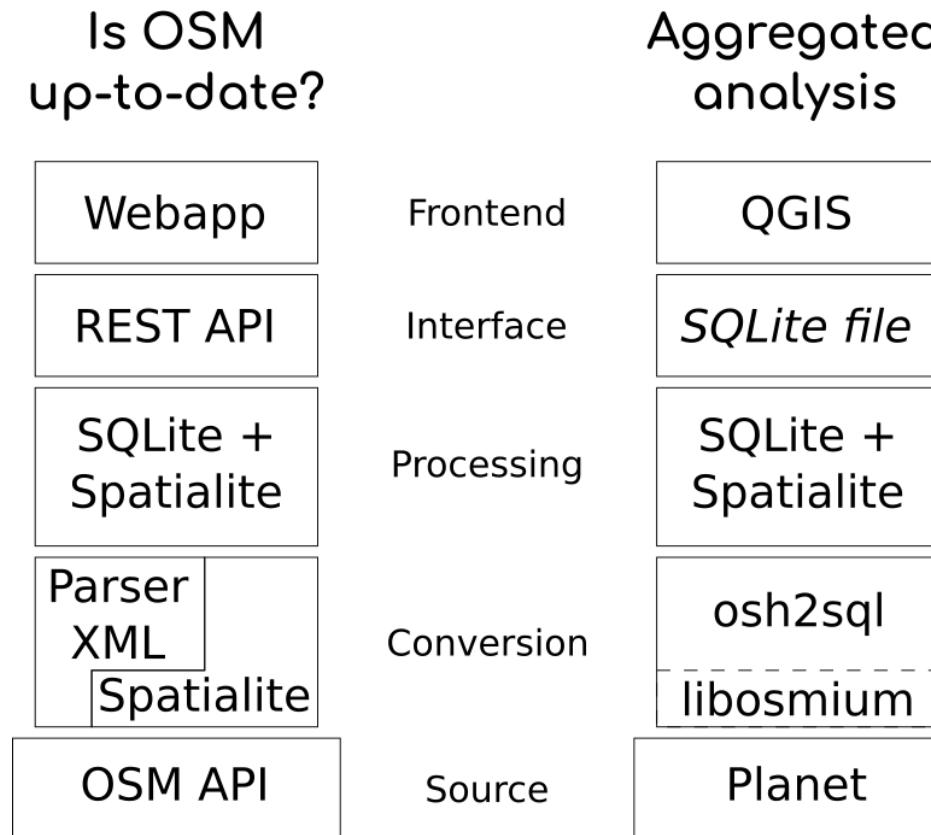
OpenStreetMap quality

- The main **concern** which still limits the widespread use of OSM
- Quality can be assessed through several **parameters**:
 - positional accuracy
 - completeness
 - logical consistency
 - semantic accuracy
 - thematic accuracy
 - temporal accuracy
 - up-to-dateness
 - lineage
 - fitness-for-use & fitness-for-purpose
 - ...
- OSM quality assessment approaches are:
 - **extrinsic**, when OSM is compared to a reference dataset
 - **intrinsic**, when OSM is compared to itself
 - based on **OSM history** (OSM API / Full History Planet File)



Methodology - Architecture & application

- 2 open source software stacks:
 - **web application** to return real-time results for single OSM nodes/ways
 - aggregated analysis on a predefined area



Methodology - Study area & hypotheses

- Study area: Dar es Salaam, Tanzania
 - densely mapped in OSM, mainly thanks to the Dar Ramani Huria project
- Analysis on nodes (POIs) and ways:
 - nodes/ways deleted not considered
 - edits considered are only those with changes in tags
 - changes made in a single changeset count as one single new version
- For the aggregated analysis, data downloaded on May 3, 2018:
 - 129572 nodes and 1156948 ways
 - edits by 1959 different contributors
 - 150716 and 1592221 versions for nodes and ways, respectively



Methodology - Web application

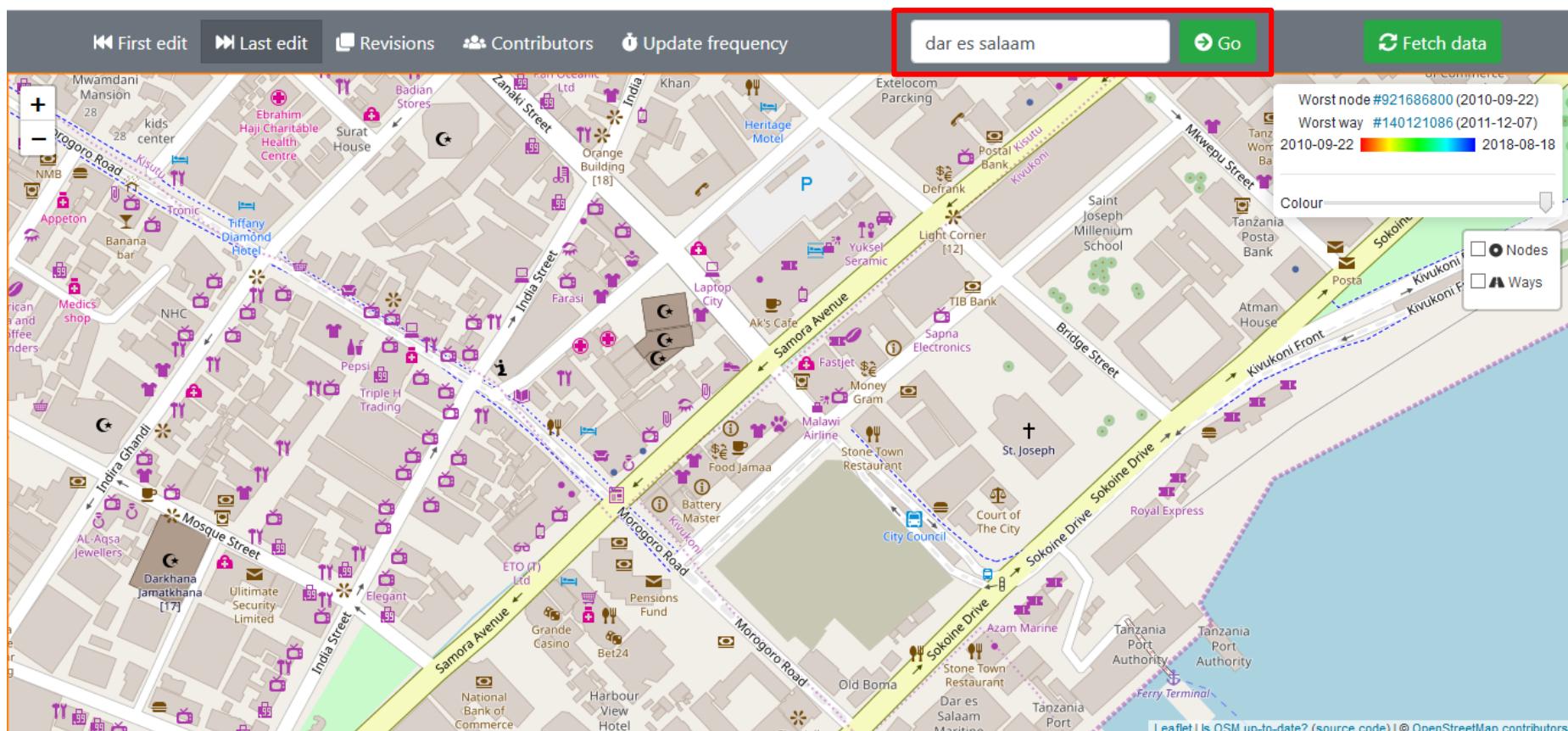


- “Is OSM up-to-date?”: <https://is-osm-updated.frafra.eu>
 - source code (AGPL v3): <https://github.com/frafra/is-osm-updated>
 - description: https://wiki.openstreetmap.org/wiki/Is_OSM_up-to-date



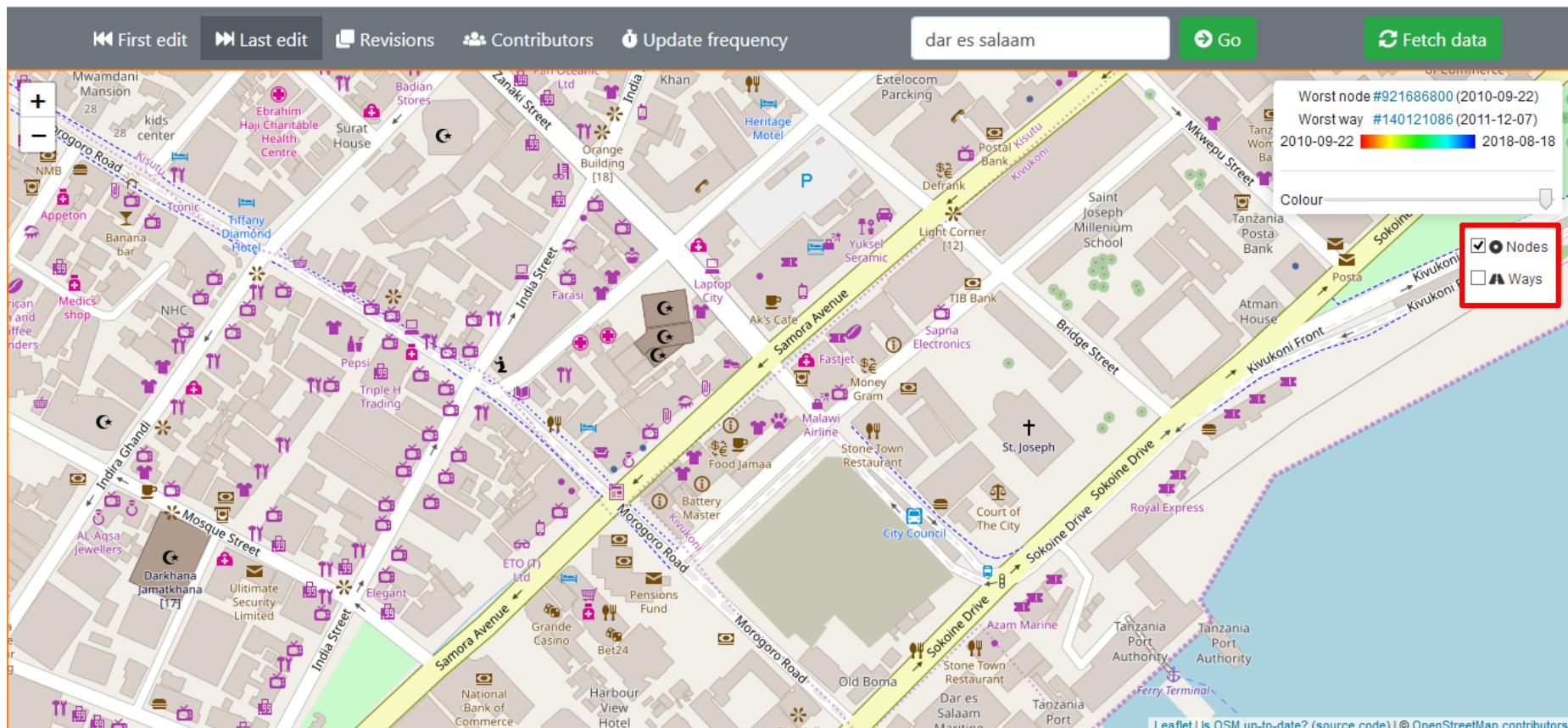
Methodology - Web application

- “Is OSM up-to-date?”: <https://is-osm-updated.frafra.eu>
 - search & zoom the map on a specific location



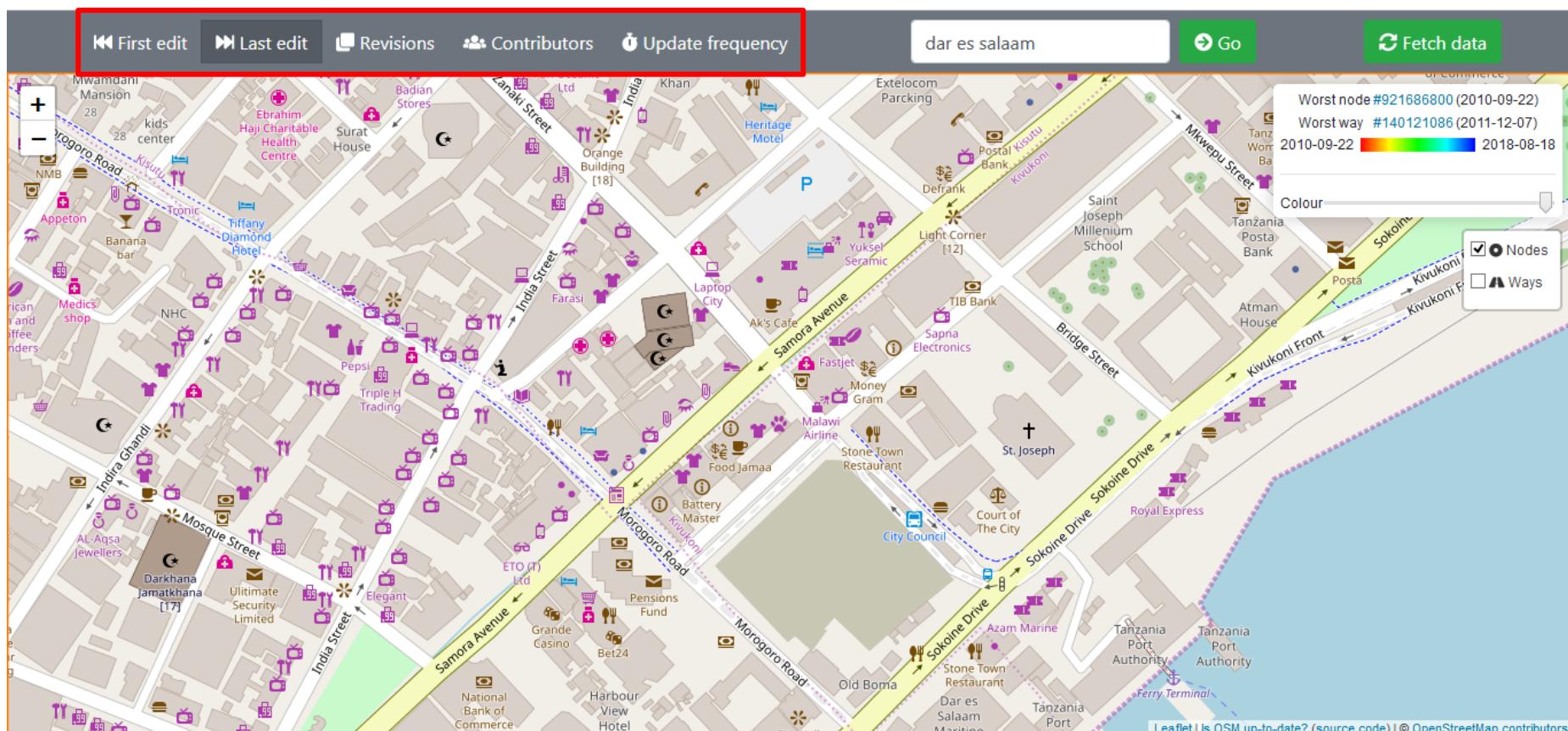
Methodology - Web application

- “Is OSM up-to-date?”: <https://is-osm-updated.frafra.eu>
 - choose whether to analyze OSM nodes, ways, or both



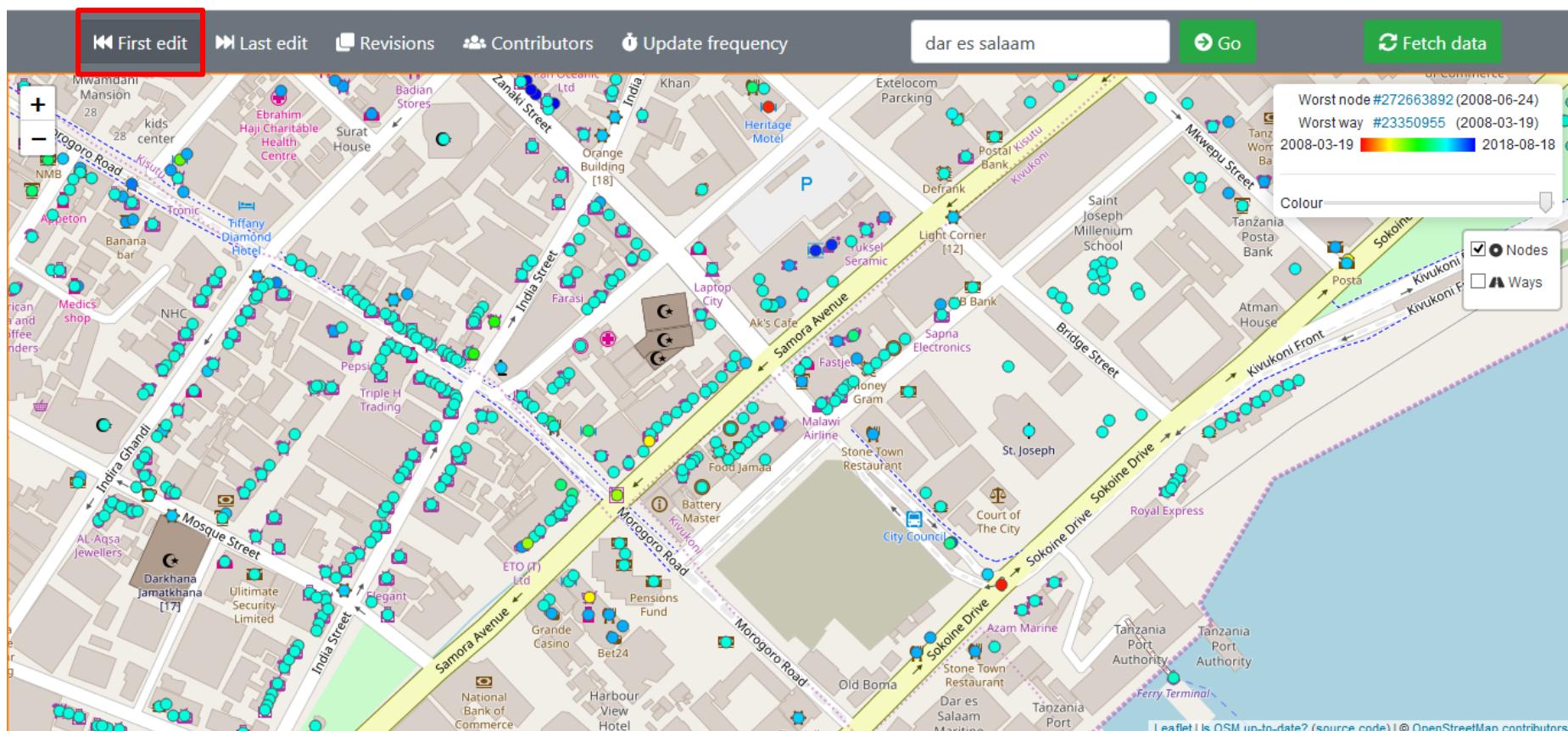
Methodology - Web application

- “Is OSM up-to-date?”: <https://is-osm-updated.frafra.eu>
 - choose the attribute to analyze: date of creation, date of last edit, number of versions, number of different contributors, update frequency



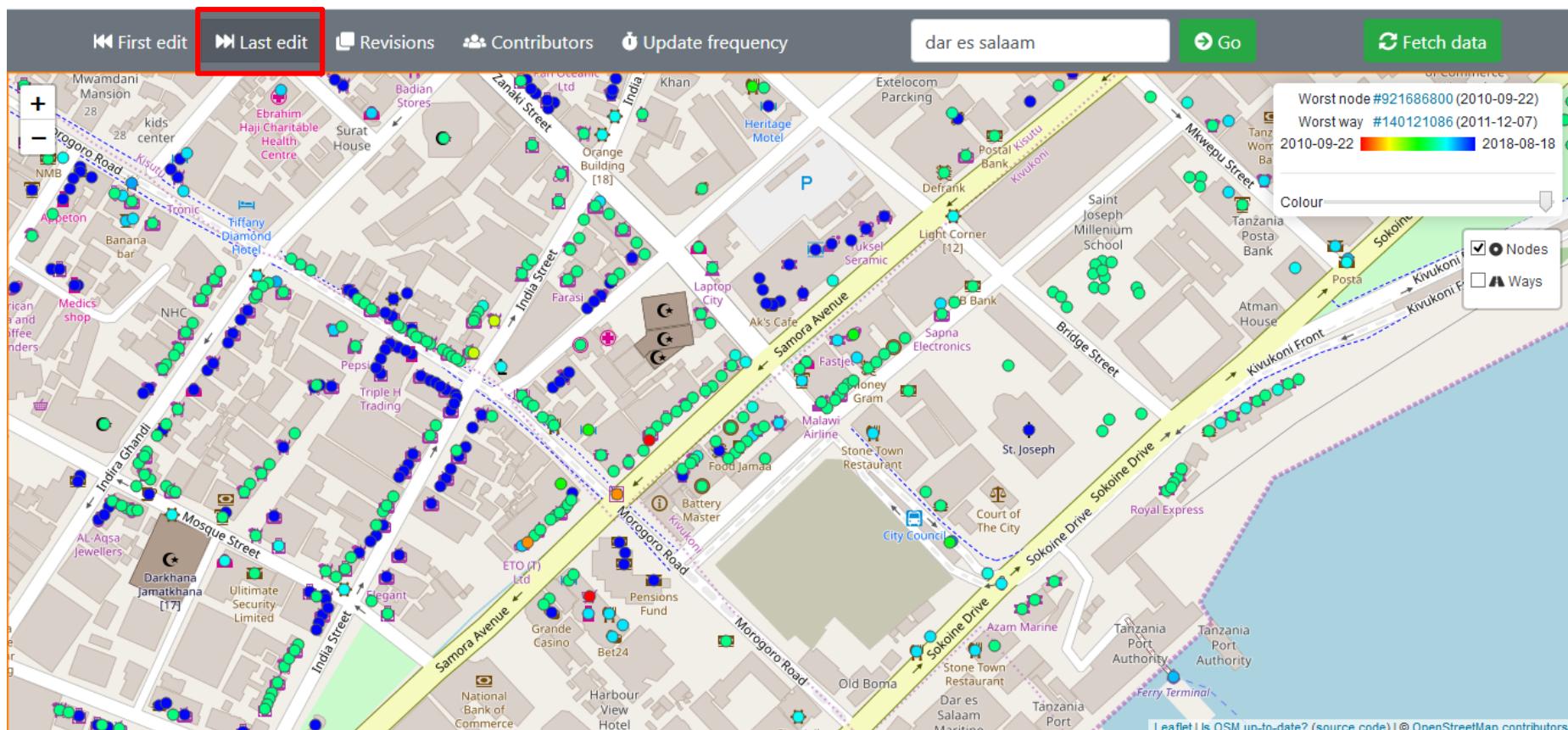
Methodology - Web application

- “Is OSM up-to-date?”: <https://is-osm-updated.frafra.eu>
 - visualize color classifications of OSM nodes/ways based on the attribute
 - date of first edit (i.e. creation) of nodes/ways



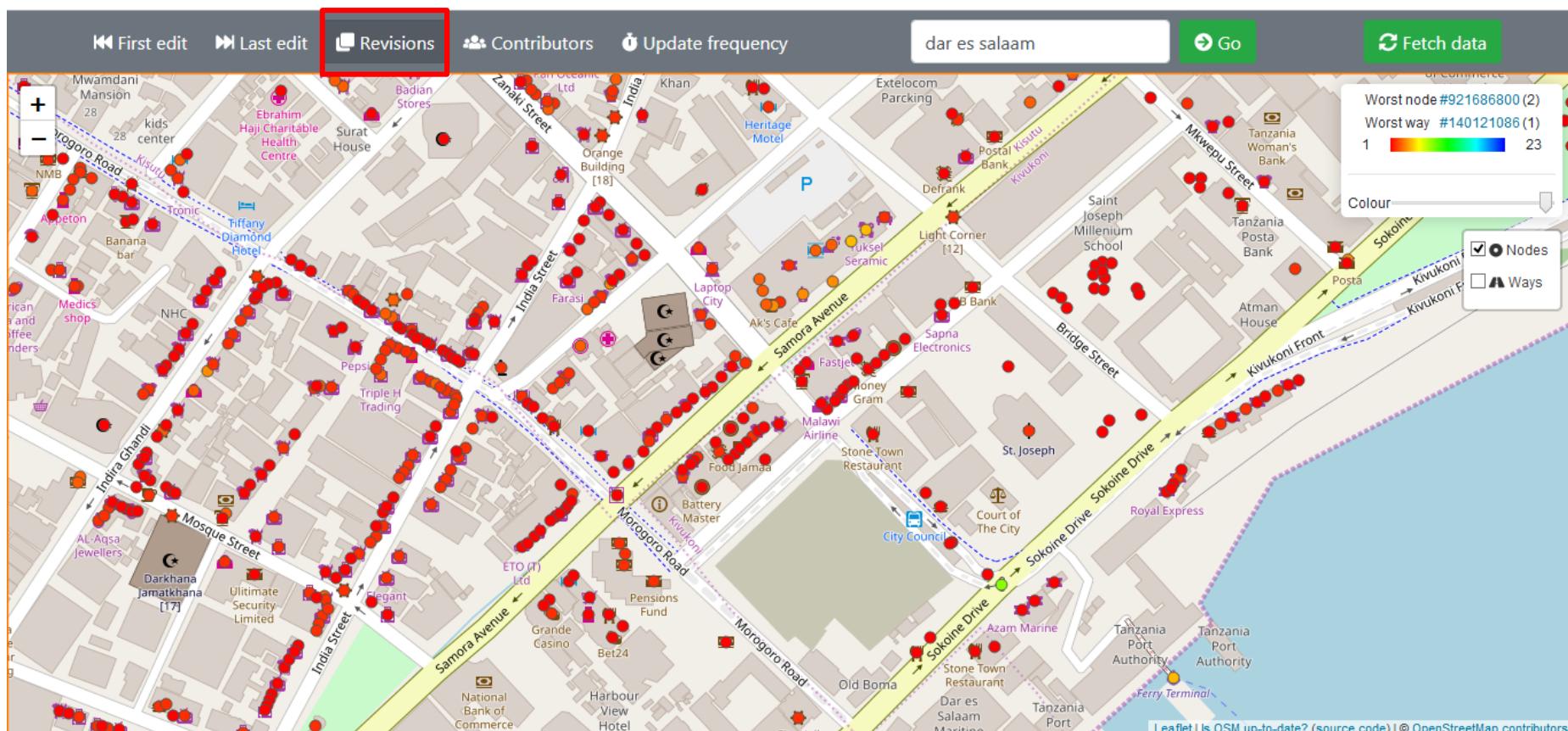
Methodology - Web application

- “Is OSM up-to-date?”: <https://is-osm-updated.frafra.eu>
 - visualize color classifications of OSM nodes/ways based on the attribute
 - date of last edit of nodes/ways



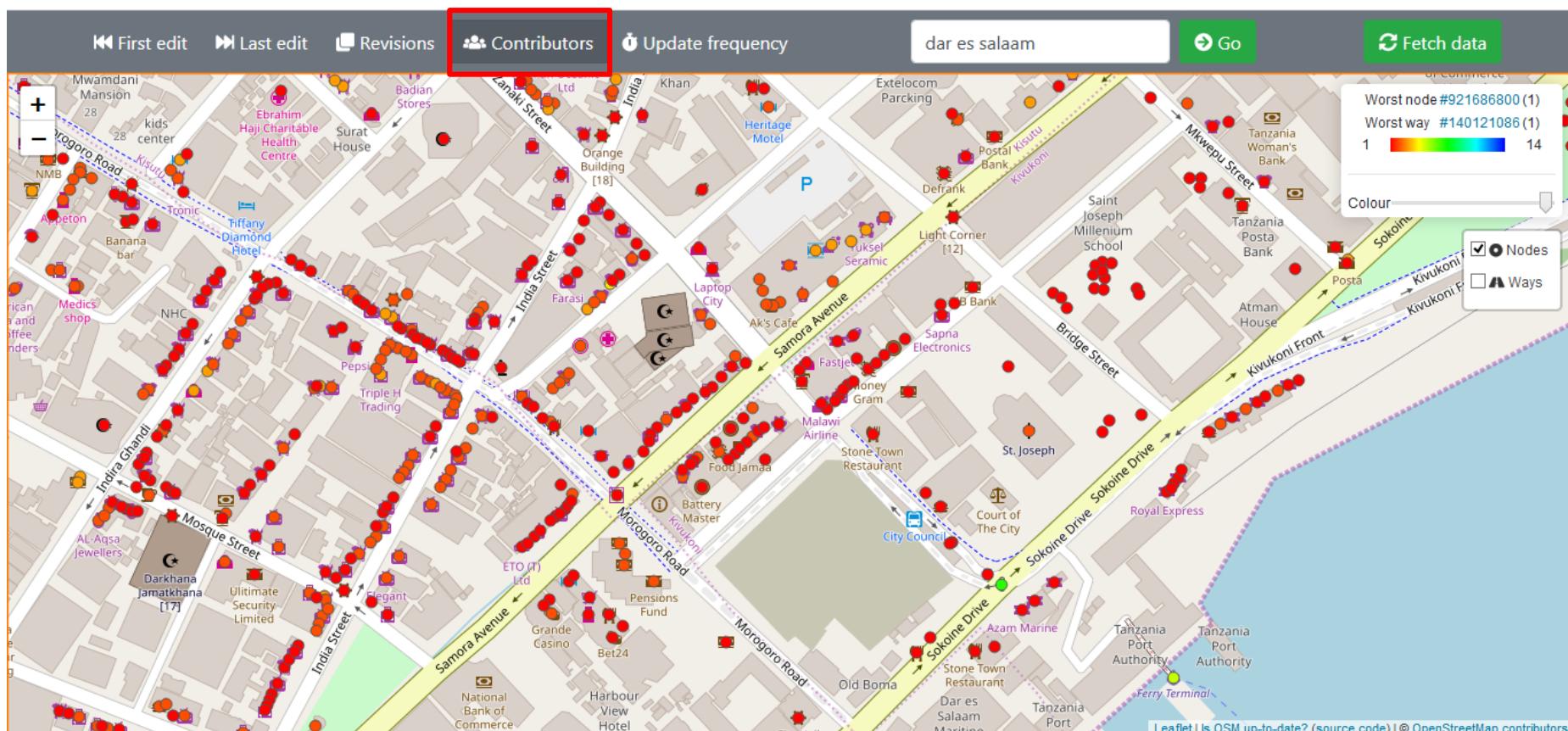
Methodology - Web application

- “Is OSM up-to-date?”: <https://is-osm-updated.frafra.eu>
 - visualize color classifications of OSM nodes/ways based on the attribute
 - number of revisions (i.e. versions) of nodes/ways



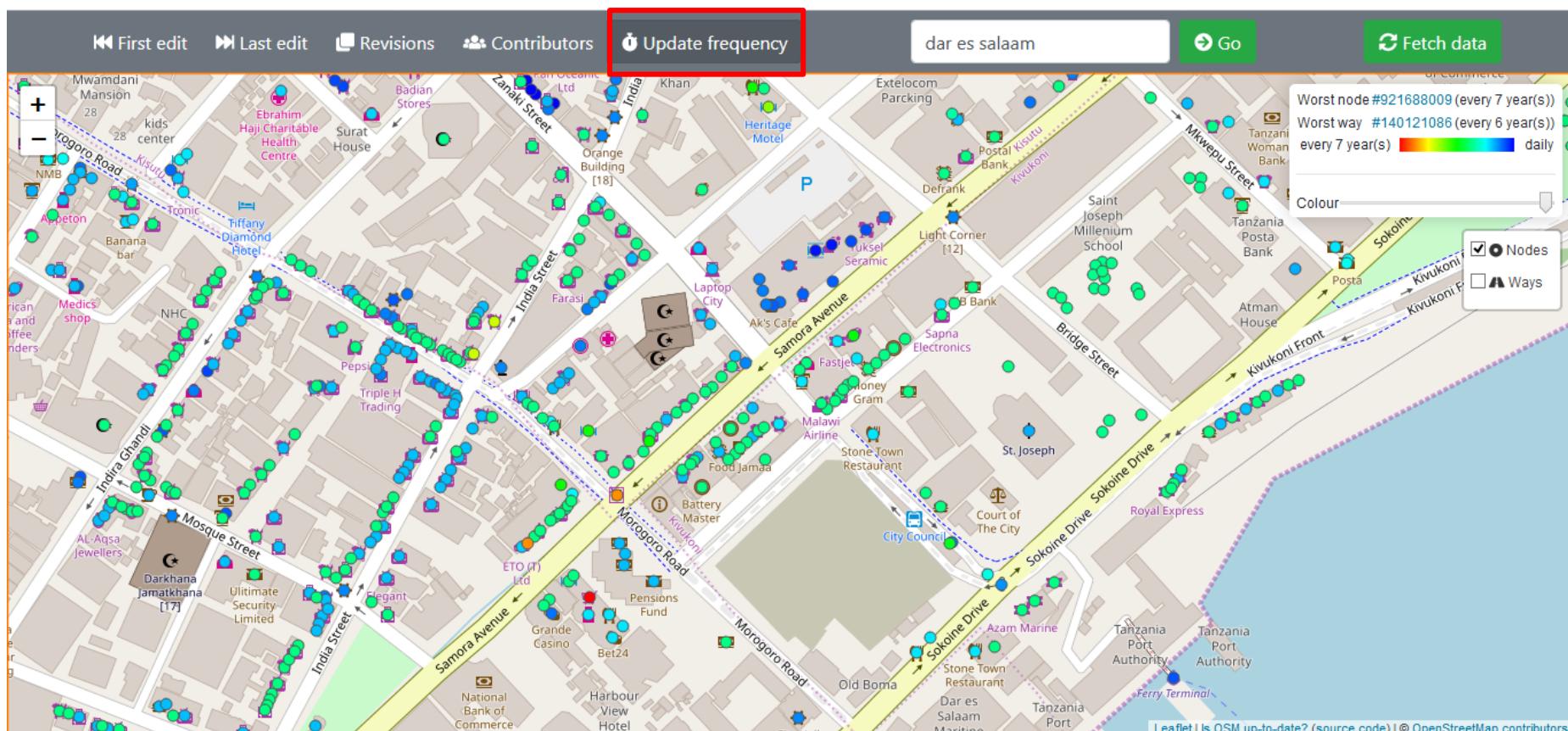
Methodology - Web application

- “Is OSM up-to-date?”: <https://is-osm-updated.frafra.eu>
 - visualize color classifications of OSM nodes/ways based on the attribute
 - number of different contributors who have edited the nodes/ways



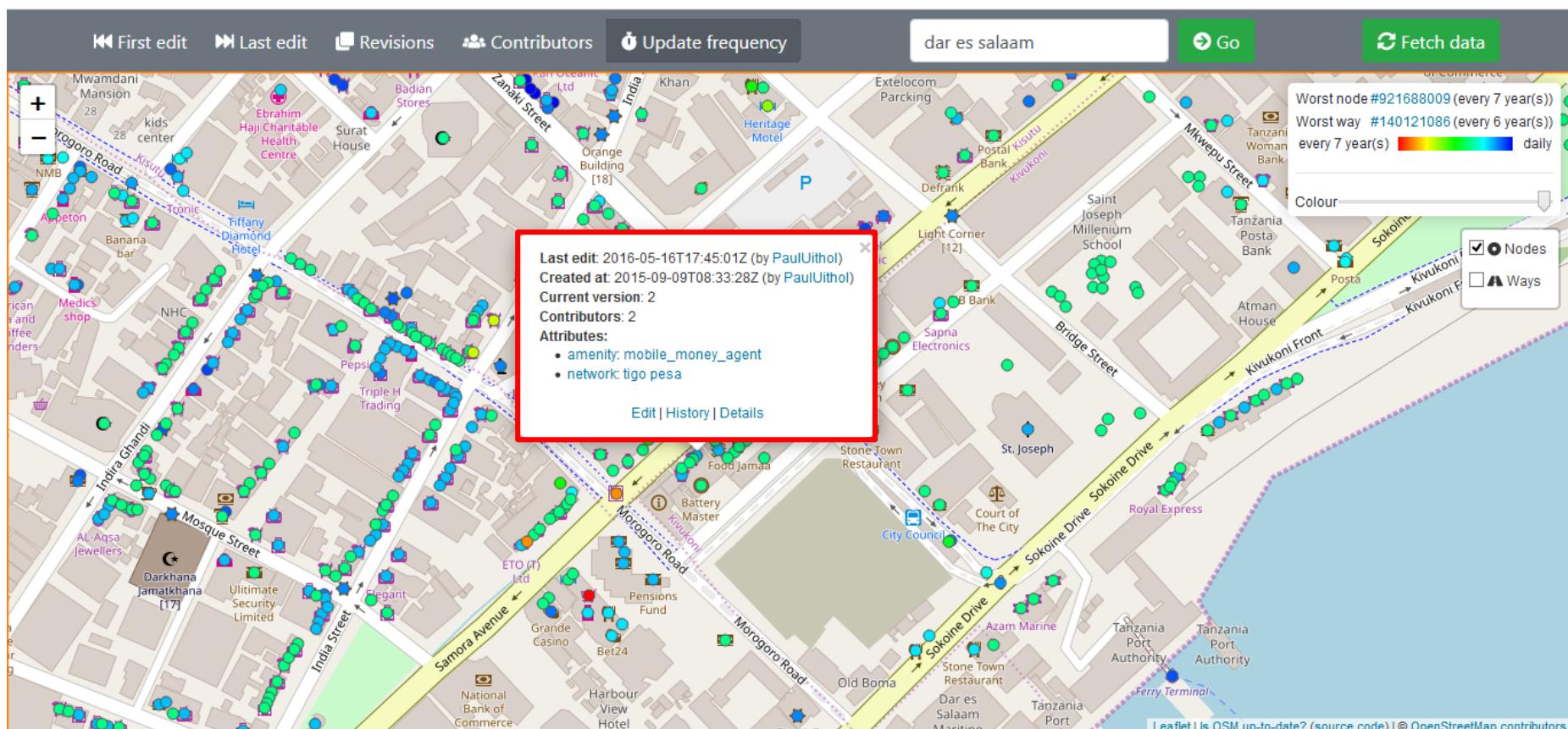
Methodology - Web application

- “Is OSM up-to-date?”: <https://is-osm-updated.frafra.eu>
 - visualize color classifications of OSM nodes/ways based on the attribute
 - frequency of update of nodes/ways



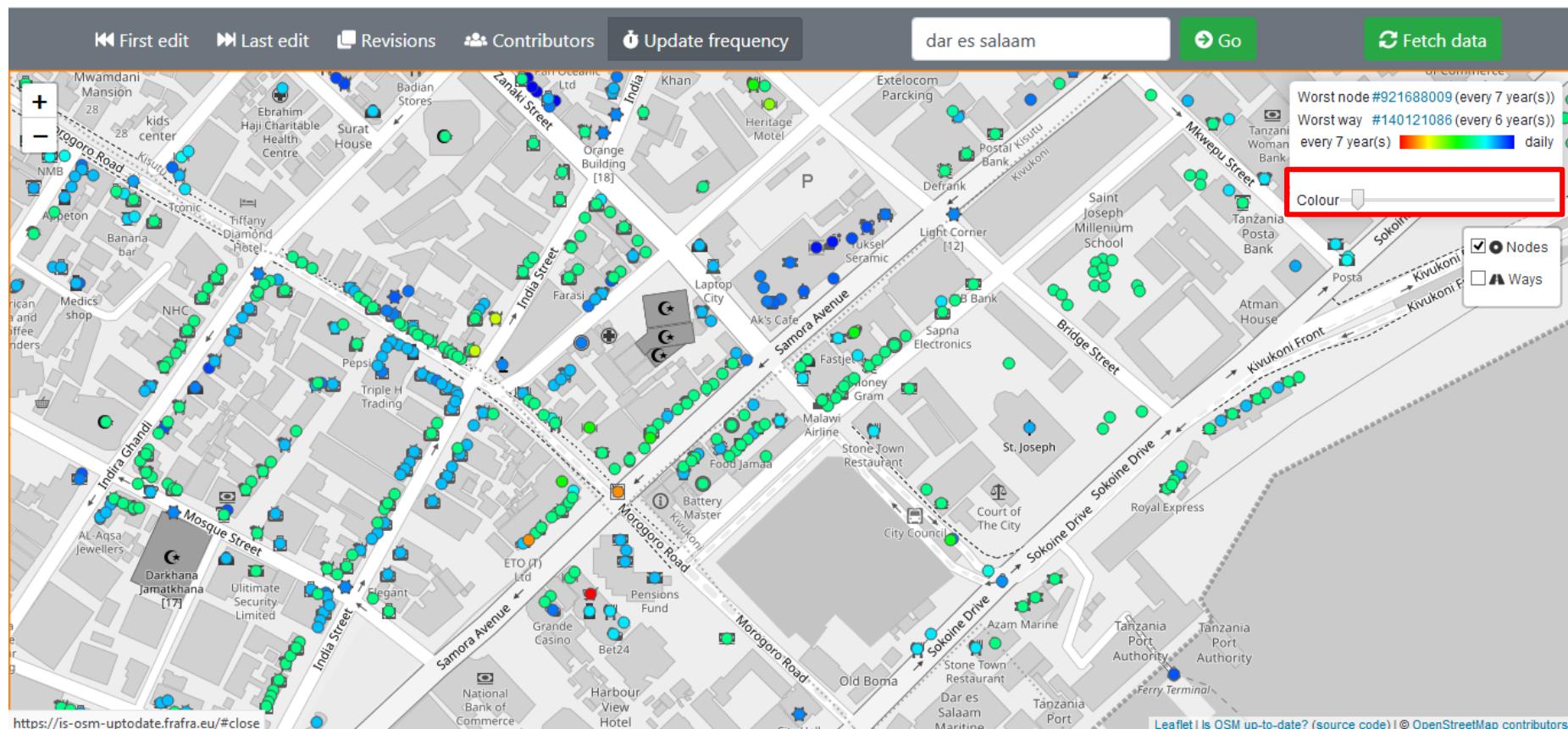
Methodology - Web application

- “Is OSM up-to-date?”: <https://is-osm-updated.frafra.eu>
 - click on an OSM node/way to visualize a **popup** with all the attribute information, tags, and links to visualize/edit the node/way in OSM



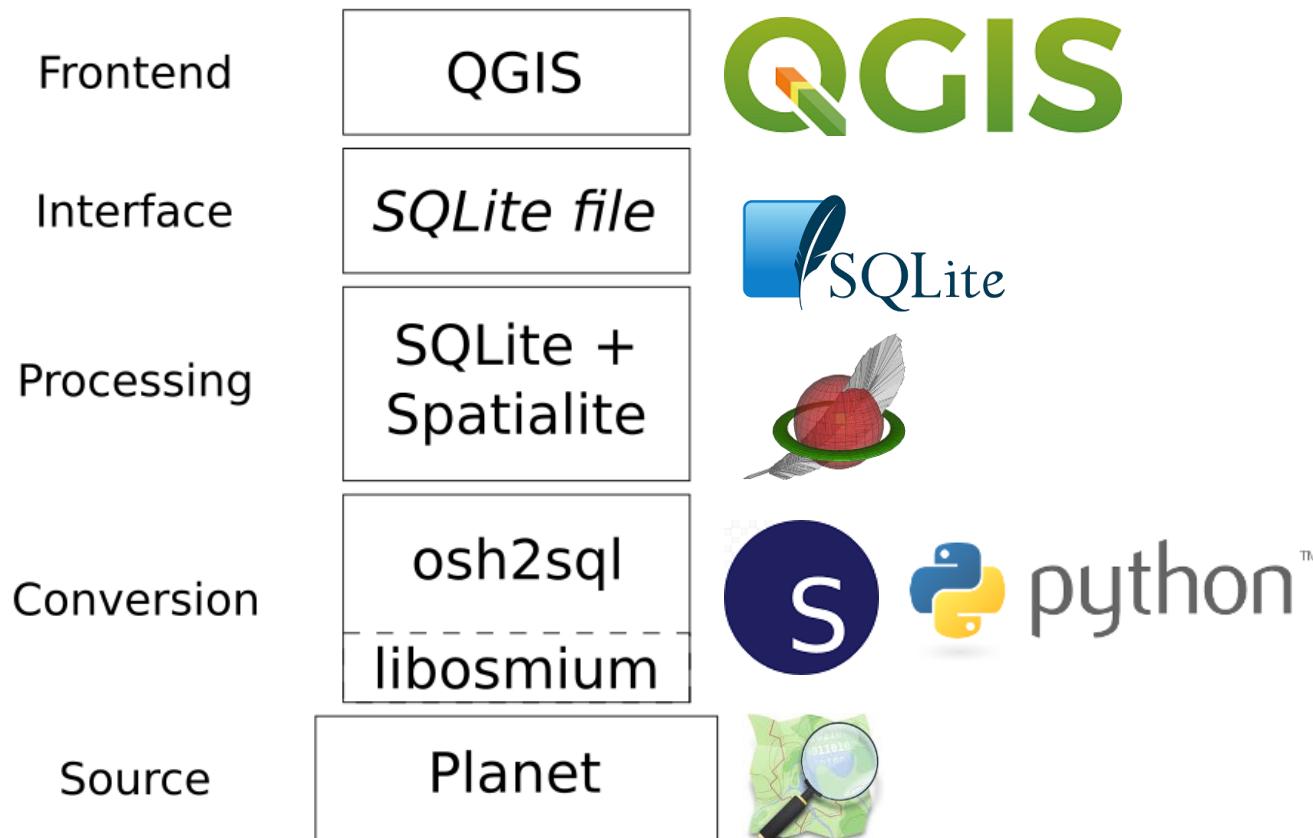
Methodology - Web application

- “Is OSM up-to-date?”: <https://is-osm-updated.frafra.eu>
 - adjust the color of the basemap through a **colorbar** to improve the visualization of OSM nodes/ways



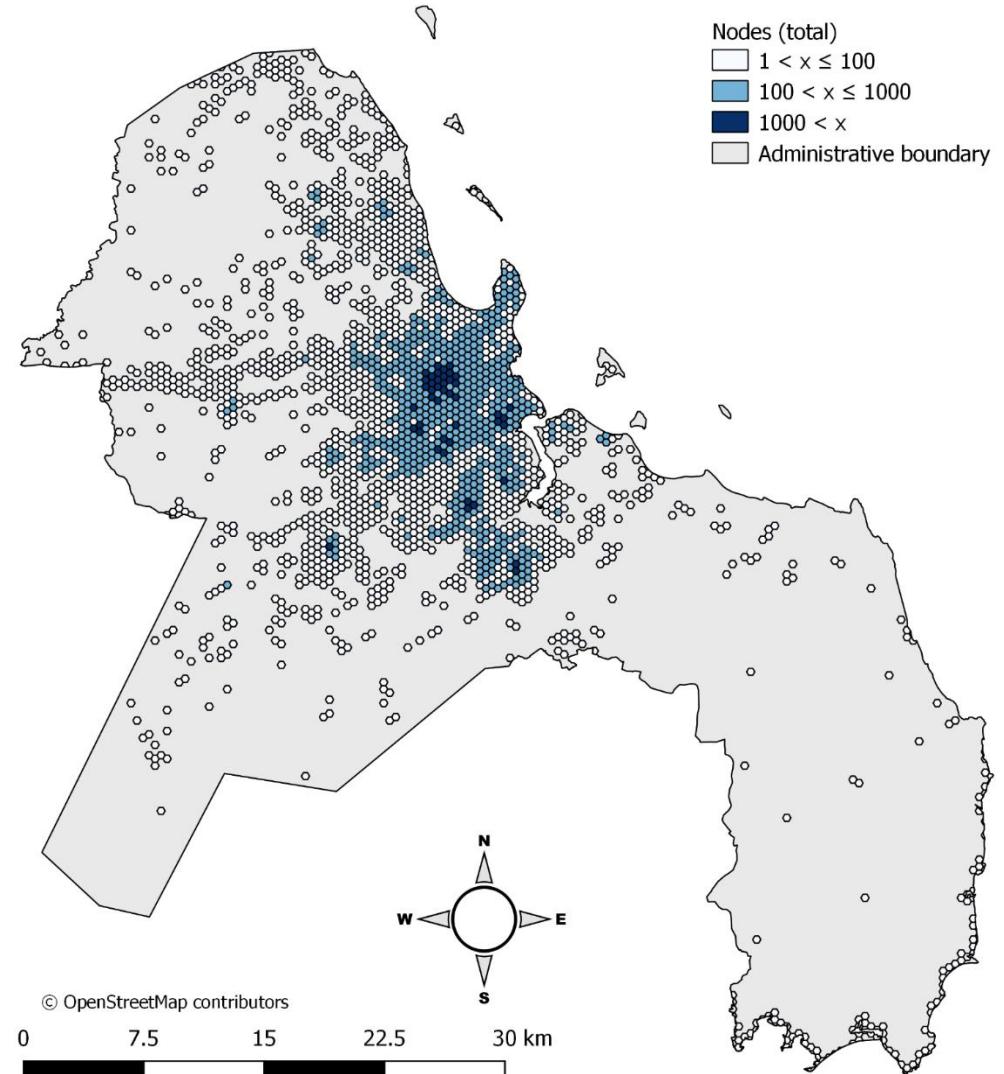
Methodology - Aggregated analysis

- More extensive analysis on a predefined area:
 - aggregate and **store results** in a database
 - suitable for further **GIS processing**



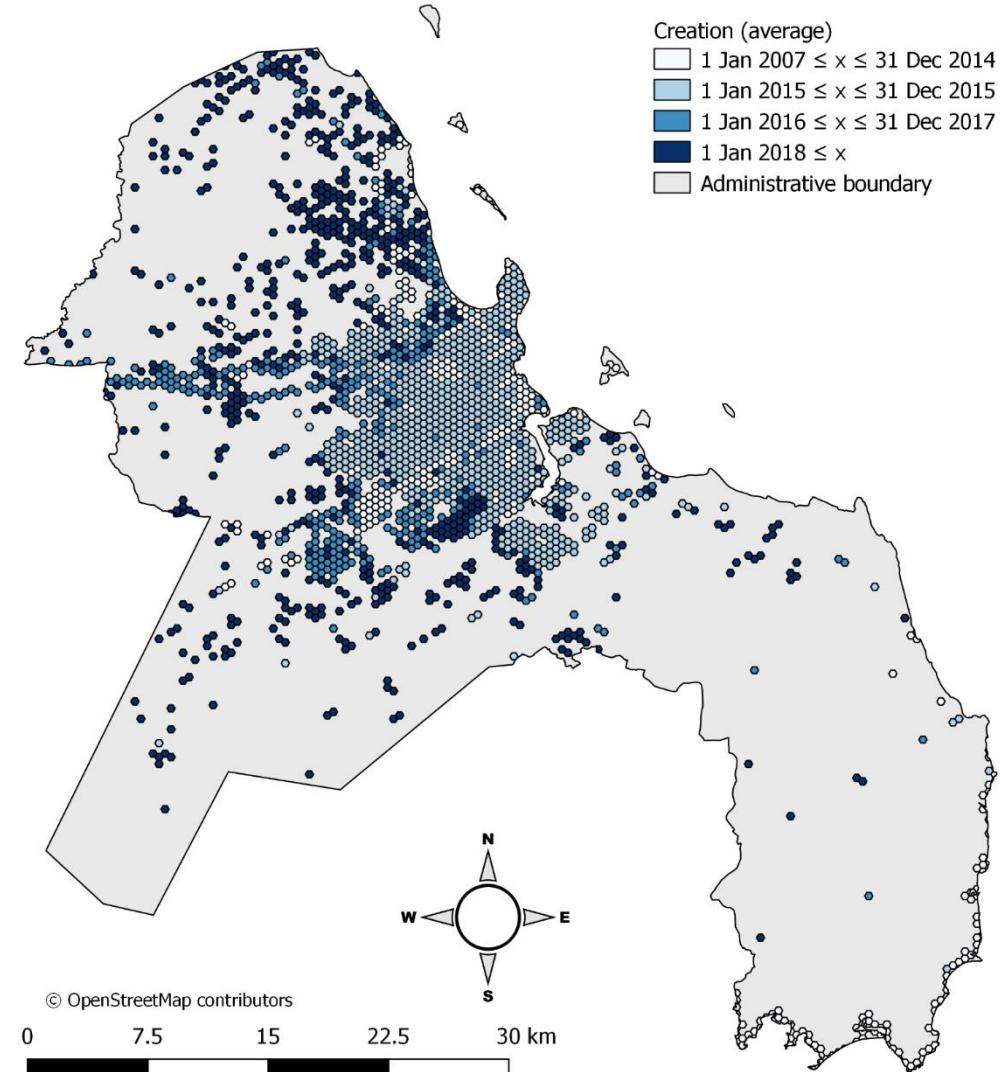
Results - Aggregated analysis

- Total number of OSM nodes:
 - 79% of the total area does not contain any node
 - density of nodes progressively **increasing** from the rural to the most urbanized areas



Results - Aggregated analysis

- Average date of creation of OSM nodes:
 - most of the nodes in the **city center** created in **2015**
 - attention gradually moved to the **peripheral areas** in **2016, 2017 and 2018**

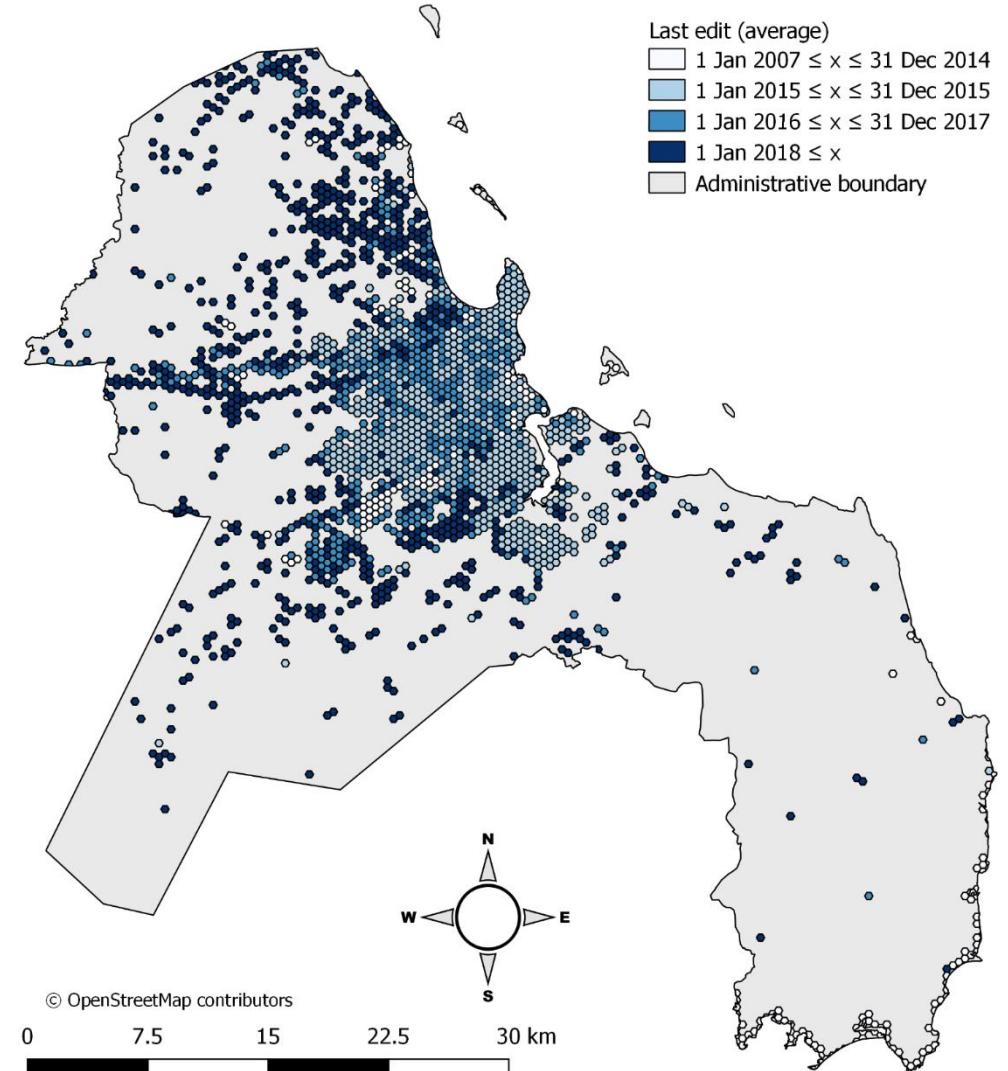


POLITECNICO MILANO 1863



Results - Aggregated analysis

- Average date of last edit of OSM nodes:
 - few of the nodes created in 2014-2015 were later updated
 - mapping in 2018 focused on peripheral areas

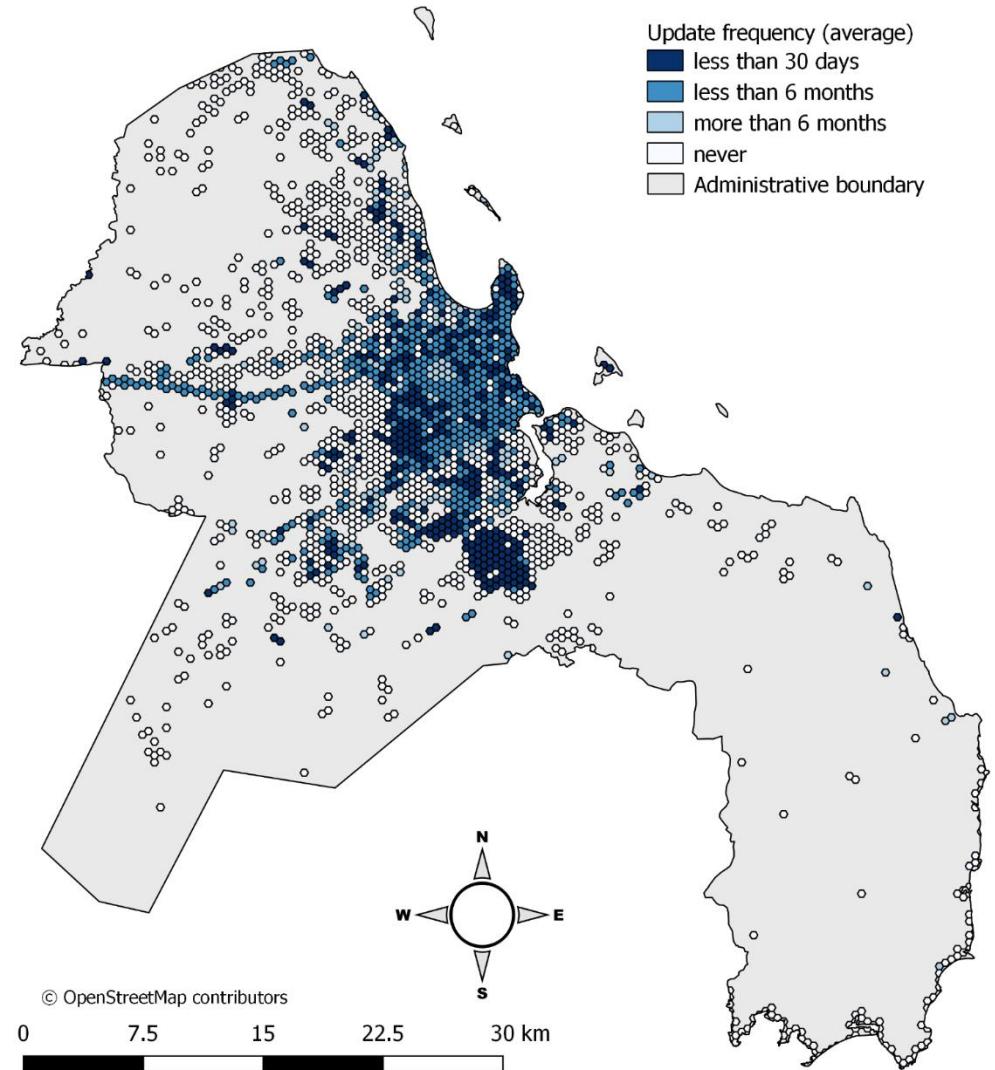


POLITECNICO MILANO 1863



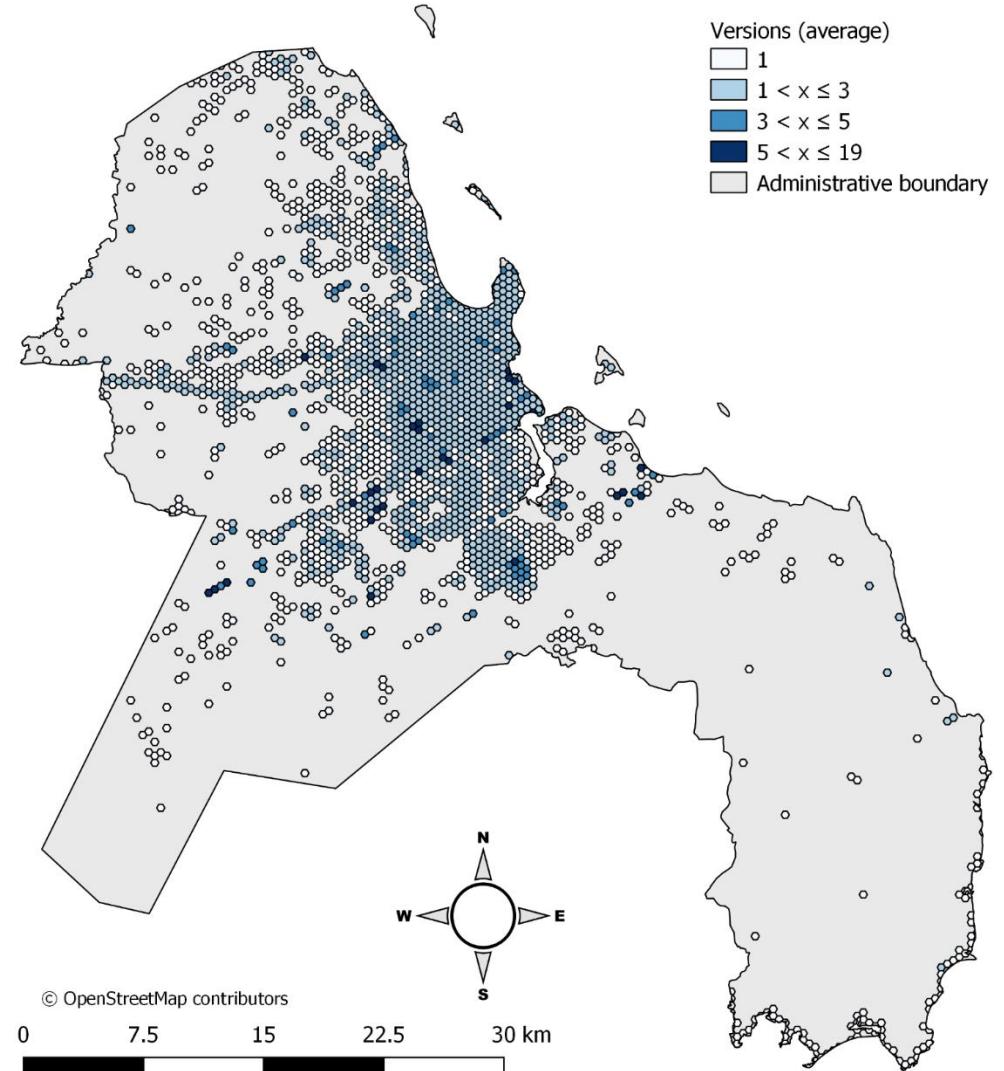
Results - Aggregated analysis

- Average update frequency of OSM nodes:
 - highest update frequencies in the city center
 - most of the nodes created in 2018 have not yet been updated



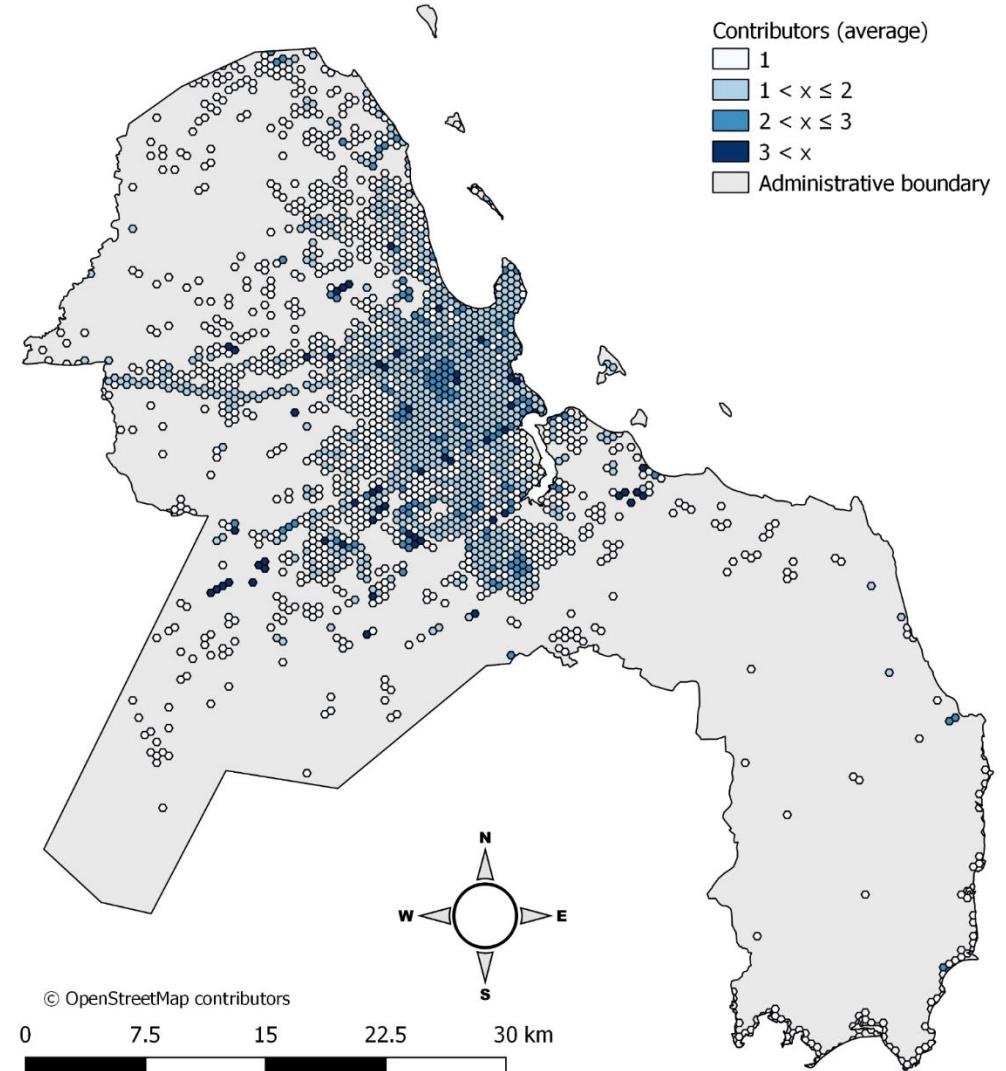
Results - Aggregated analysis

- Average number of versions of OSM nodes:
 - most of recently created nodes not (yet) updated
 - increase in the number of version when moving to the city center



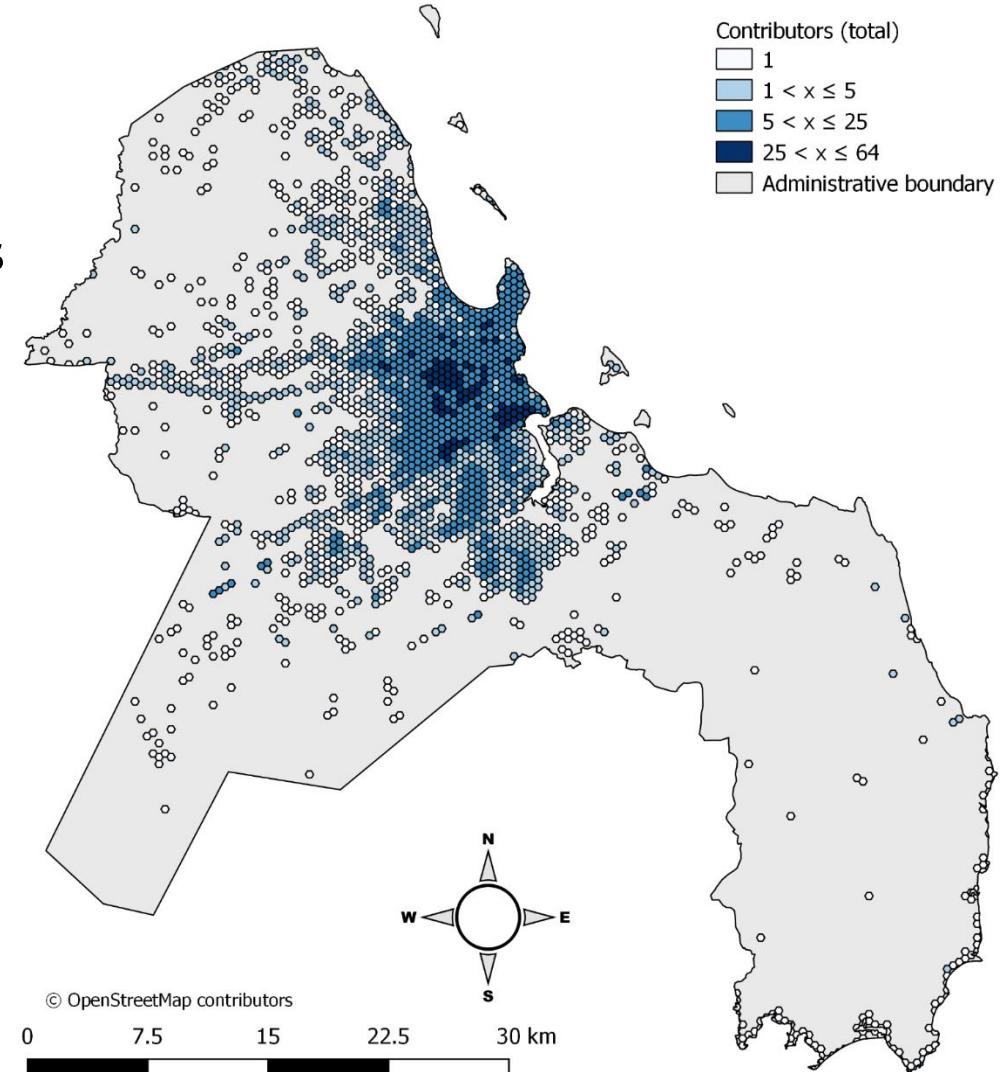
Results - Aggregated analysis

- Average number of different contributors on OSM nodes:
 - equal to 1 for 53% of the cells, mainly in the outskirts
 - increases towards the city center



Results - Aggregated analysis

- Total number of different contributors on OSM nodes:
 - generally equal to 1 in the periphery, increases towards the city center



Conclusions - Quality assessment

- OSM history to unveil the mapping process happened in Dar es Salaam:
 - driven by the [Dar Ramani Huria project](#), started in 2015
 - OSM community was [formed](#)
 - mapping gradually expanded from the city center to the periphery



Conclusions - Quality assessment

- OSM history to unveil the mapping process happened in Dar es Salaam:
 - driven by the [Dar Ramani Huria project](#), started in 2015
 - OSM community was [formed](#)
 - mapping gradually expanded [from the city center to the periphery](#)
- [Quality](#) (from the intrinsic assessment):
 - OSM development still [at a young stage](#)
 - (where OSM is available) [temporal accuracy](#), [up-to-dateness](#) and [lineage](#) reflect the same center/periphery trend



Conclusions - Quality assessment

- OSM history to unveil the mapping process happened in Dar es Salaam:
 - driven by the [Dar Ramani Huria project](#), started in 2015
 - OSM community was [formed](#)
 - mapping gradually expanded [from the city center to the periphery](#)
- [Quality](#) (from the intrinsic assessment):
 - OSM development still [at a young stage](#)
 - (where OSM is available) [temporal accuracy](#), [up-to-dateness](#) and [lineage](#) reflect the same center/periphery trend
 - [completeness](#) decreases from center to periphery
 - [fitness-for-use](#) for the requirements of the Dar Ramani Huria project is high (infrastructure networks)



Conclusions - Future work

- Customize the intrinsic analysis for specific categories of OSM objects:
 - addresses, commercial activities, natural elements, etc. which have very different update cycles
- Combine the history of OSM objects with the history of OSM contributors:
 - data reliability may depend on the contributor's experience
 - a single quality index?
- Correlate quality elements with demographic/territorial parameters:
 - population density, elevation, land cover/land use, etc.



Conclusions - Future work

- Customize the intrinsic analysis for specific categories of OSM objects:
 - addresses, commercial activities, natural elements, etc. which have very different update cycles
- Combine the history of OSM objects with the history of OSM contributors:
 - data reliability may depend on the contributor's experience
 - a single quality index?
- Correlate quality elements with demographic/territorial parameters:
 - population density, elevation, land cover/land use, etc.
- FOSS4G technology was key to achieve this!



References & Acknowledgements

- Reference material:
 - SQLite database table with aggregated OSM data:
<https://frafra.eu/archive/osm/dar-es-salaam.zip>
 - corresponding paper: <https://tinyurl.com/y7ryboqb>
 - this presentation: <https://tinyurl.com/ybmrukwt>
- Special thanks to:



International Journal of
Geo-Information

2018 TRAVEL
AWARDS
FOR POSTDOCS

Thank you!

Marco Minghini - [@MarcoMinghini](mailto:marco.minghini@ec.europa.eu)



POLITECNICO MILANO 1863



A Special Issue of interest



- Special Issue: Open Source Geospatial Software
- Journal: [Open Geospatial Data, Software and Standards \(Springer\)](#)
- Guest Editors:
 - [Marco Minghini](#), European Commission - Joint Research Center, Italy
 - [Amin Mobasher](#), Heidelberg University, Germany
 - [Victoria Rautenbach](#), University of Pretoria, South Africa
 - [Maria Antonia Brovelli](#), Politecnico di Milano, Italy
- Articles accepted:
 - standard [research articles](#)
 - [software articles](#), focused on open source software of broad interest
- APC: [\\$1030](#) - some fee waivers available, please contact us!
- Deadline: [November 15, 2018](#)
- Link: <https://opengeospatialdata.springeropen.com/osgs>



POLITECNICO MILANO 1863

