

# Martin Dittus @dekstop

EngD student since Jan 2013

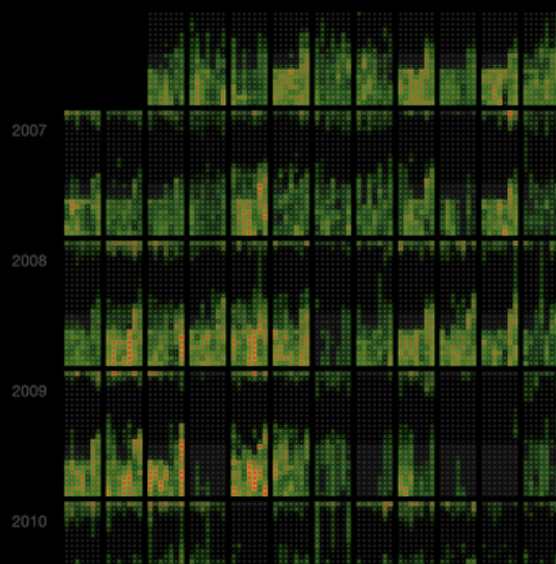
Researching data-gathering communities:  
OpenStreetMap, Cosm, ...



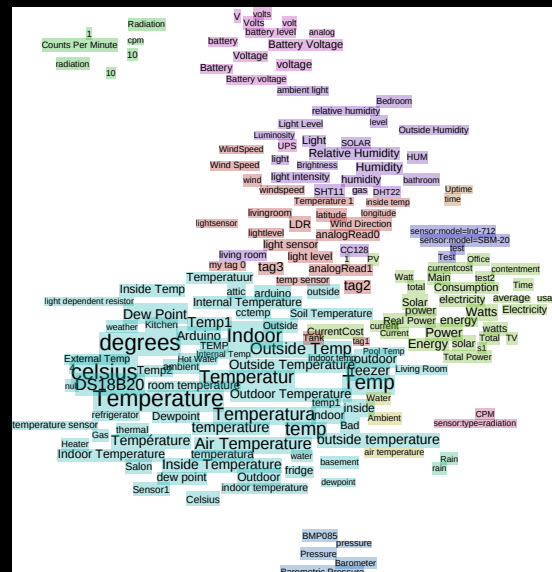
I am particularly interested in community processes:

## How can we coordinate large numbers of contributors?

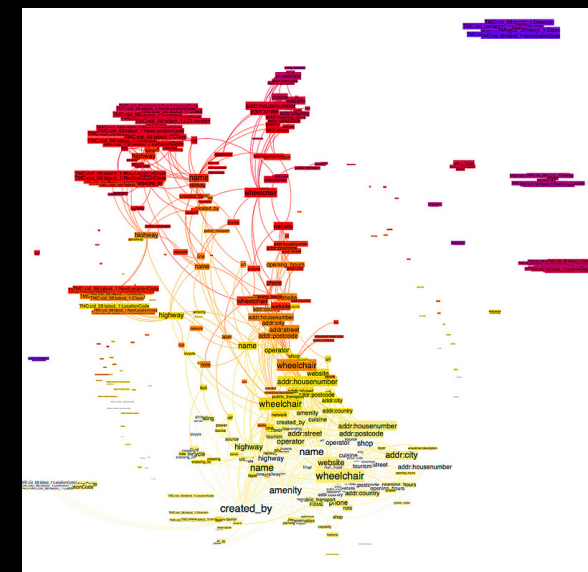
- What interests and motivations can bring people together?
- What tools and processes do they need to make it work?
- How can we assess and assert data quality?



Music listening habits



Sensor data annotations



Map editing patterns

# Collaborative Editing Practice on OpenStreetMap

Martin Dittus, 2014-03-10

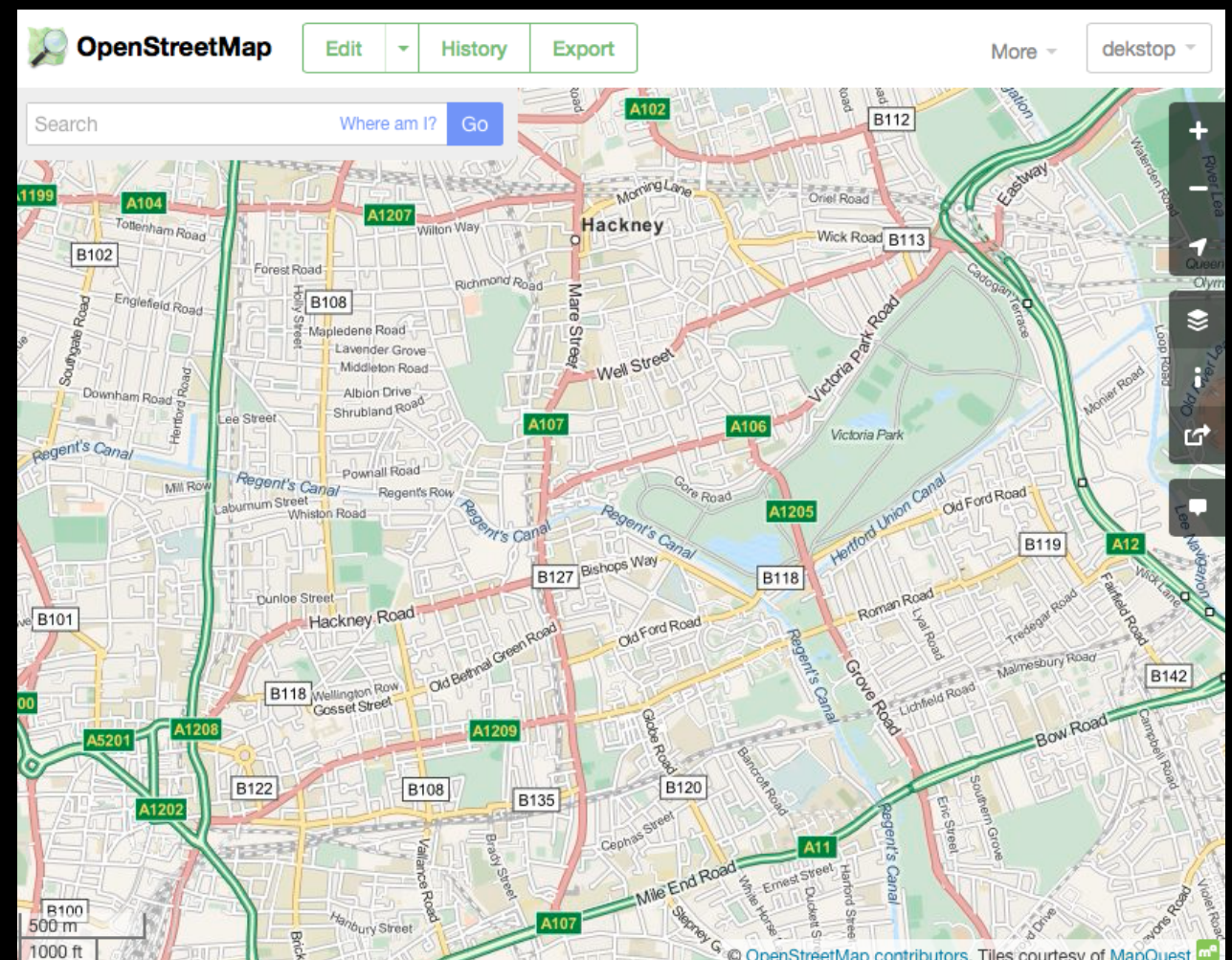
Supervisors: Licia Capra, Yvonne Rogers

# OpenStreetMap: a community-produced map.

Largely collected by volunteers.

How well does it work?

I'm particularly interested  
in the social processes.







OpenStreetMap

Edit



History

Export

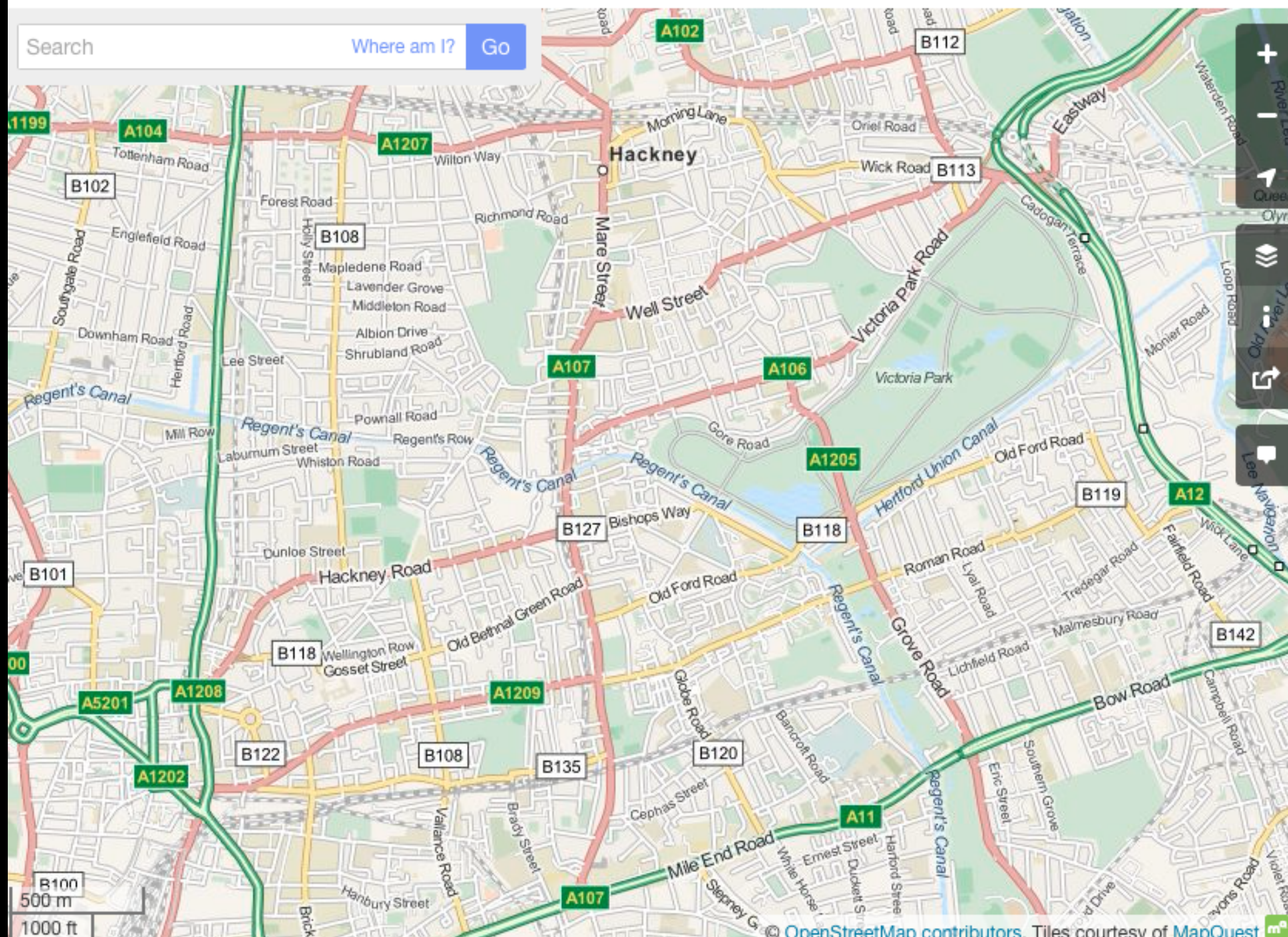
More

dekstop

Search

Where am I?

Go



© OpenStreetMap contributors. Tiles courtesy of MapQuest





## Edit feature



Leisure



Name



London Hackspace



Type



hackerspace

Address



447-453

Hackney Road

City

E2 9DY

Website



<http://london.hackspace.org.uk/>



[View on openstreetmap.org](#)

Point

Line

Area



Billy's Cafe

London Hackspace

Pritchards Road

Marian Street

Treadway Street

bing

© Getmapping plc, © 2012 GeoEye, © 2012 Intermap, © AND



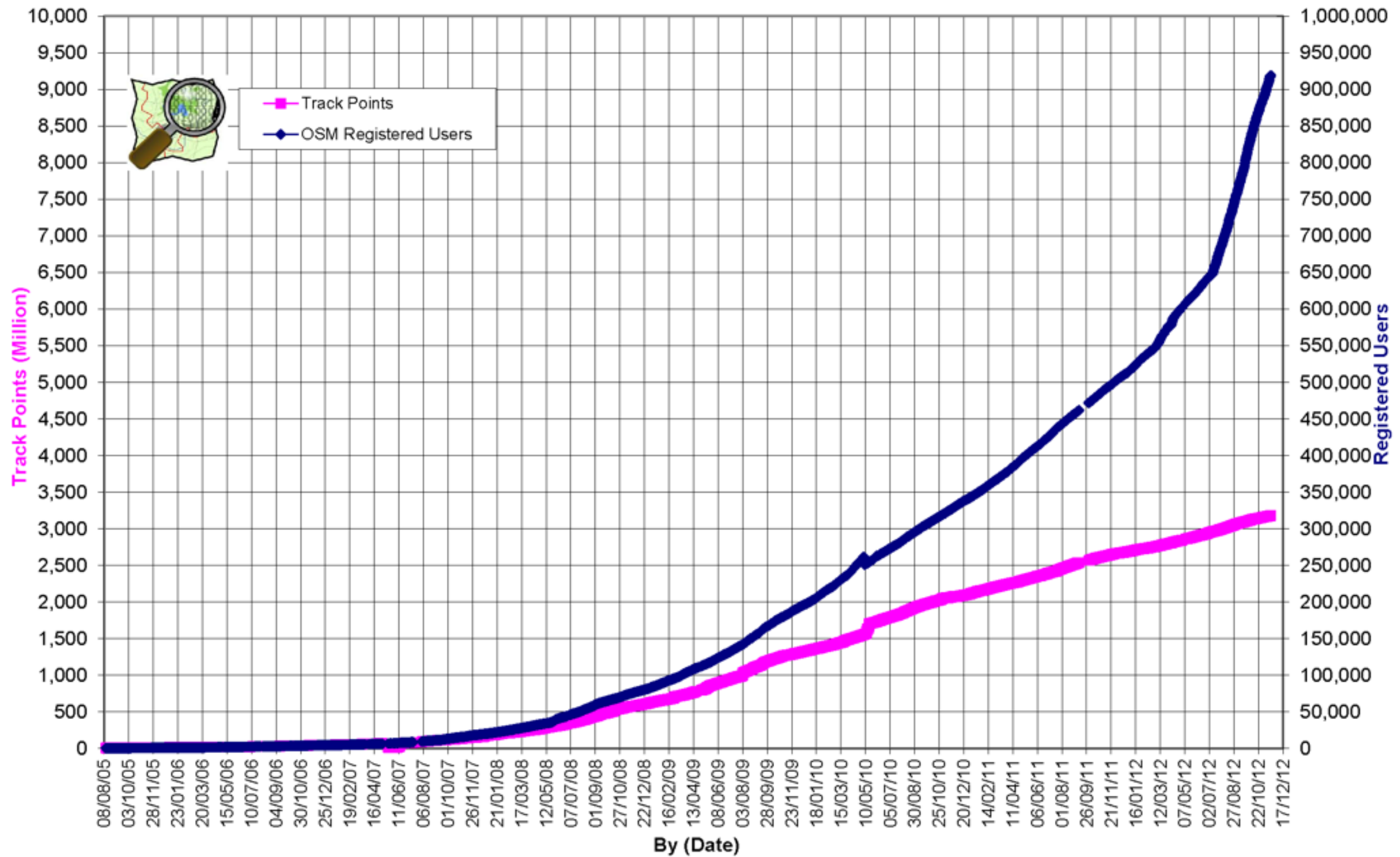
Edits by [woodpeck\\_fixbot](#), [Ed Avis](#), [Russ](#), and [13 others](#)



1.3.7

## OpenStreetMap Database Statistics

### Users and User gpx Uploads (track points)







A world map visualization showing the density of OpenStreetMap contributors. The map uses a color scale from light yellow to dark blue to represent the concentration of contributors. Darker blue areas indicate higher density, while lighter yellow areas indicate lower density. The map shows a high concentration of contributors in North America, Europe, and parts of Asia, with lower concentrations in Africa, South America, and Australia.

## Density of OpenStreepMap contributors

Resolution: 0.5 map units

Martin Dittus Dec 2013  
@dekstop

# No formal peer review on OpenStreetMap.

Expectation should be: map data is of low quality.



Despite this:

**map is of high quality!**

A growing number of studies show that map quality is often comparable to commercial maps, sometimes **even better**. (Based on different measures: POI density, length of road networks, positional accuracy of streets, ...)

# Are there indications of **informal peer reviews**?

Do contributors refine map data provided by others?

## Approach:

- Data mining of full OSM history, billions of edits.
- Looking at “micro-interactions”: points of interest (POI) that were edited by more than one person.
- Under which circumstances does this happen?

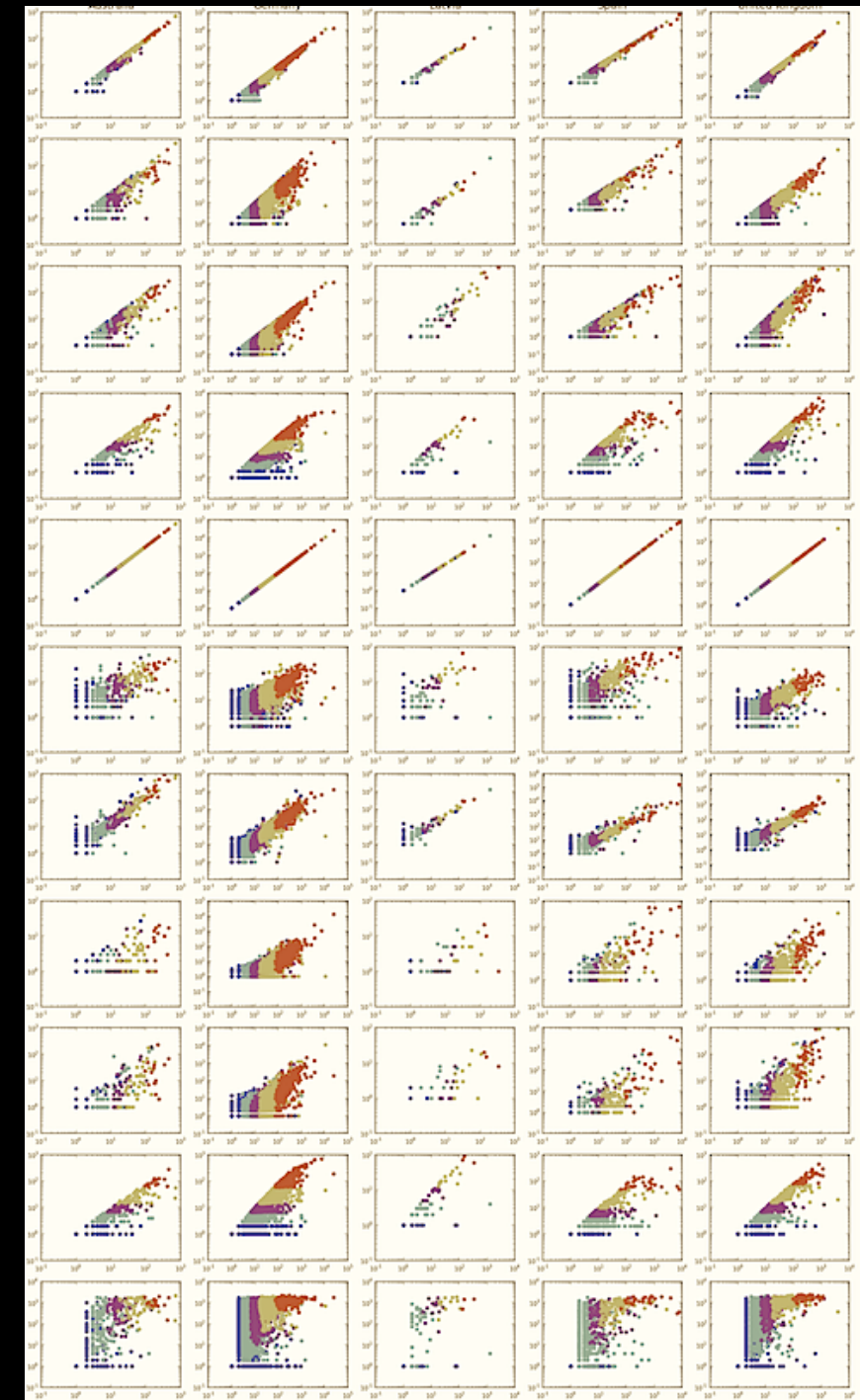
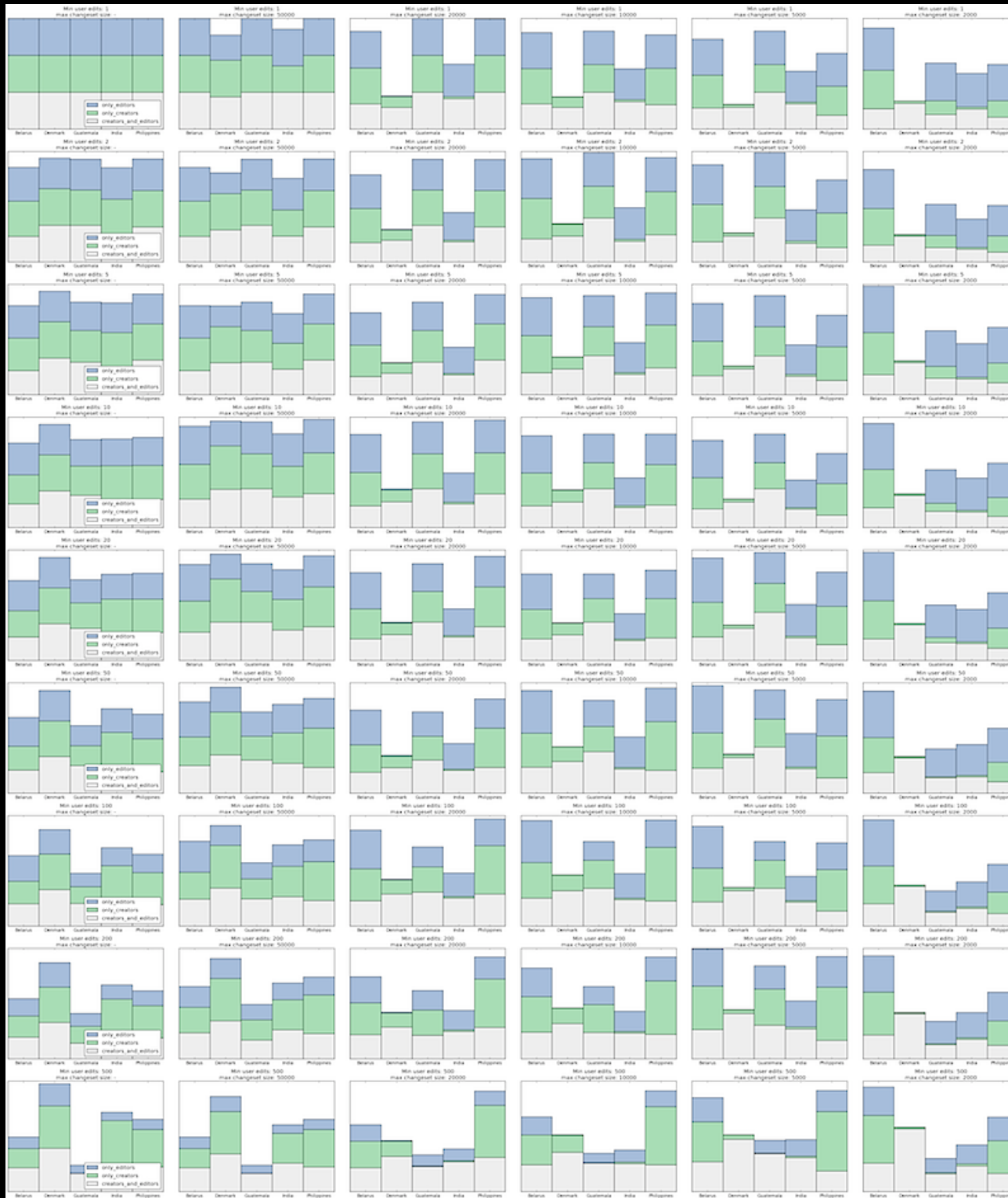


Category	# edits	# POI	# editors	# co-edits	% co-edits
place	14,638	14,052	1,597	2,590	17.69%
building	10,603	10,361	1,833	1,307	12.33%
railway	8,588	8,524	900	931	10.84%
historic	4,613	4,528	1,468	489	10.60%
highway	56,402	56,195	3,637	5,289	9.38%
emergency	4,098	4,076	552	359	8.76%
leisure	4,081	4,040	1,681	350	8.58%
sport	1,386	1,371	688	77	5.56%
shop	17,724	17,468	3,696	940	5.30%
man_made	5,579	5,441	1,206	287	5.14%
amenity	60,555	59,946	7,533	2,708	4.47%
tourism	13,124	12,959	3,049	586	4.47%
barrier	18,299	18,133	3,138	475	2.60%
power	41,306	41,109	863	570	1.38%
natural	26,707	26,667	1,455	283	1.06%
Average					7.04%

Table 4.5: **Edits by POI category** in August 2013, ranked by percentage of collaborative edits, an indicator of the share of collaborative edits within each category. Entries are limited to categories with 500 or more editors.

Country	Add	Update	Remove	Any
United States	10.16%	1.79%	24.21%	36.16%
United Kingdom	11.31%	3.46%	3.15%	17.92%
Germany	10.34%	3.37%	2.96%	16.67%
Russia	9.29%	2.70%	1.66%	13.66%
Poland	7.99%	3.52%	1.86%	13.37%
Austria	6.89%	2.56%	1.52%	10.97%
France	7.45%	1.98%	1.41%	10.84%
Italy	6.16%	2.73%	0.76%	9.64%

Table 4.12: **Share of collaborative edits per country** in August 2013, by type of edit. Percentages are calculated in relation to all edits in the respective country.





Early findings:

Collaboration is widespread (45% of users)  
but limited to a map subset (15% of POI)

Key themes

- **Cleanup** (US Tiger import: remove unused tags)
- **Completeness** (specifying "place": country, state, ...)
- **Thematic shifts** (adding address information)

**Regional culture does play a role** (individualism, power distance; uncertainty avoidance)

Bigger context:

## Do **practices** affect **data quality**?

E.g. is it easy to contribute well?

Many challenges for such research

- Much data. Work is slow.
- Messy data. Not all contributions are human.
- No clear ground truth... map data is subjective.

## Need to develop reliable **data quality measures**!



# Thank You.