Can Volunteered Geographic Information be a participant in eEnvironment and Spatial Data Infrastructures?

Peter Mooney and Padraig Corcoran

National University of Ireland, Maynooth (NUIM)

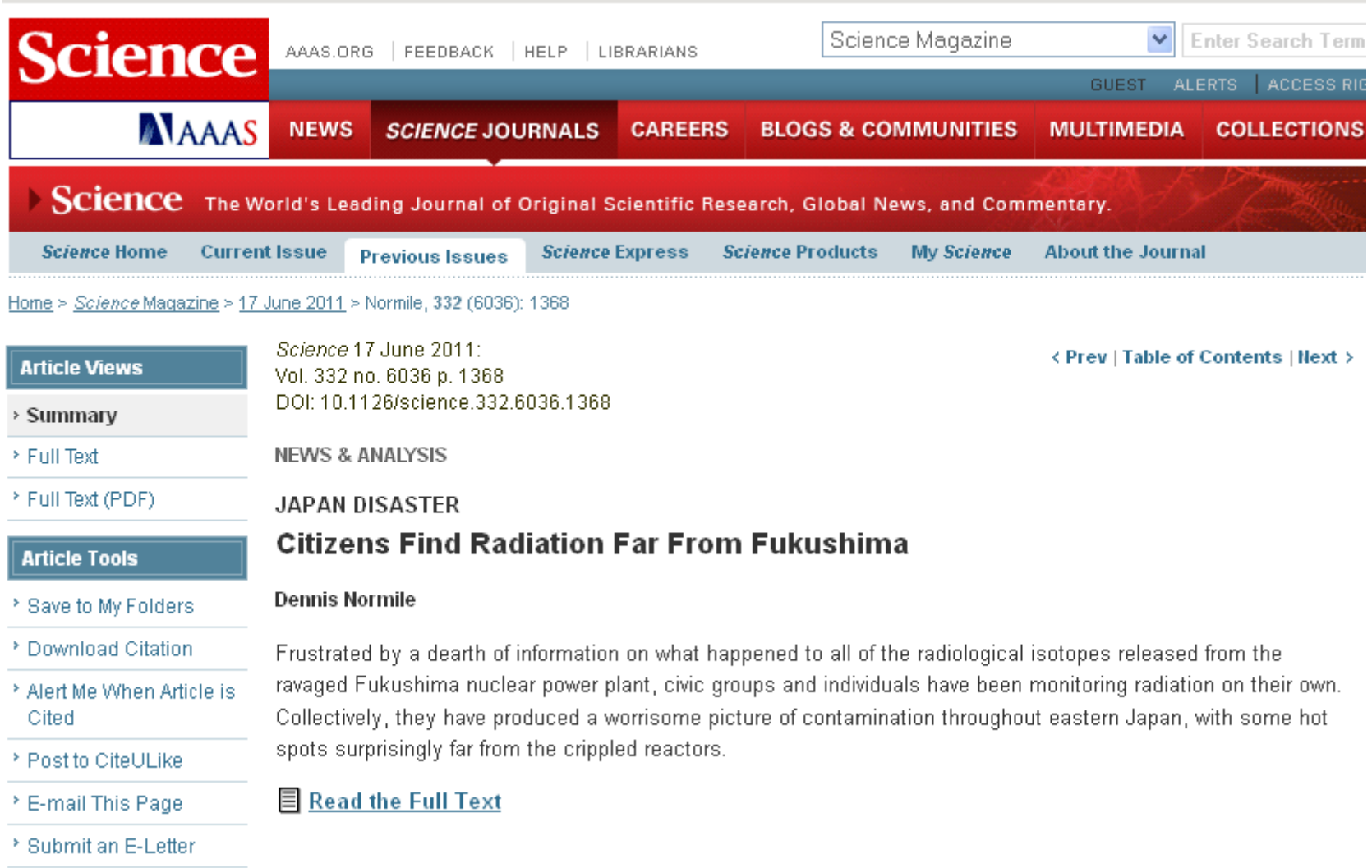
Email: peter.mooney@nuim.ie

Environmental Protection Agency

NUI MAYNOOTH
Ollscoil na hÉireann Má Nuad

http://www.broadsheet.ie/2011/06/21/what-happens-online-in-60-seconds/

“Citizen Sensing” is an emerging technological trend



Science AAAS.ORG | FEEDBACK | HELP | LIBRARIANS

Science Magazine | Enter Search Term

GUEST ALERTS | ACCESS RIG

AAAS NEWS SCIENCE JOURNALS CAREERS BLOGS & COMMUNITIES MULTIMEDIA COLLECTIONS

Science The World's Leading Journal of Original Scientific Research, Global News, and Commentary.

Science Home | Current Issue | Previous Issues | Science Express | Science Products | My Science | About the Journal

Home > Science Magazine > 17 June 2011 > Normile, 332 (6036): 1368

Article Views
- Summary
- Full Text
- Full Text (PDF)

Article Tools
- Save to My Folders
- Download Citation
- Alert Me When Article is Cited
- Post to CiteULike
- E-mail This Page
- Submit an E-Letter

Science 17 June 2011:
Vol. 332 no. 6036 p. 1368
DOI: 10.1126/science.332.6036.1368

‹ Prev | Table of Contents | Next ›

NEWS & ANALYSIS

JAPAN DISASTER

Citizens Find Radiation Far From Fukushima

Dennis Normile

Frustrated by a dearth of information on what happened to all of the radiological isotopes released from the ravaged Fukushima nuclear power plant, civic groups and individuals have been monitoring radiation on their own. Collectively, they have produced a worrisome picture of contamination throughout eastern Japan, with some hot spots surprisingly far from the crippled reactors.

Read the Full Text

There is potential for a “continous stream of data” from citizen sensors

http://www.communitysensing.org/

home technology deployments papers press related people

what if every mobile device had an air quality sensor?

We are developing mobile sensing technologies that help communities gather and analyze environmental data. We hope that this hardware and software will empower everyday citizens to learn more about their environment and influence environmental regulations and policy.

We have developed various research prototypes, which are being used in studies such as a deployment on street sweepers in San Francisco and a deployment of a handheld device in West Oakland. Right now we are focusing our efforts on air quality sensing. Our hope is that our research prototypes will demonstrate the utility of embedding environmental sensors in commercial commodity devices such as mobile phones.

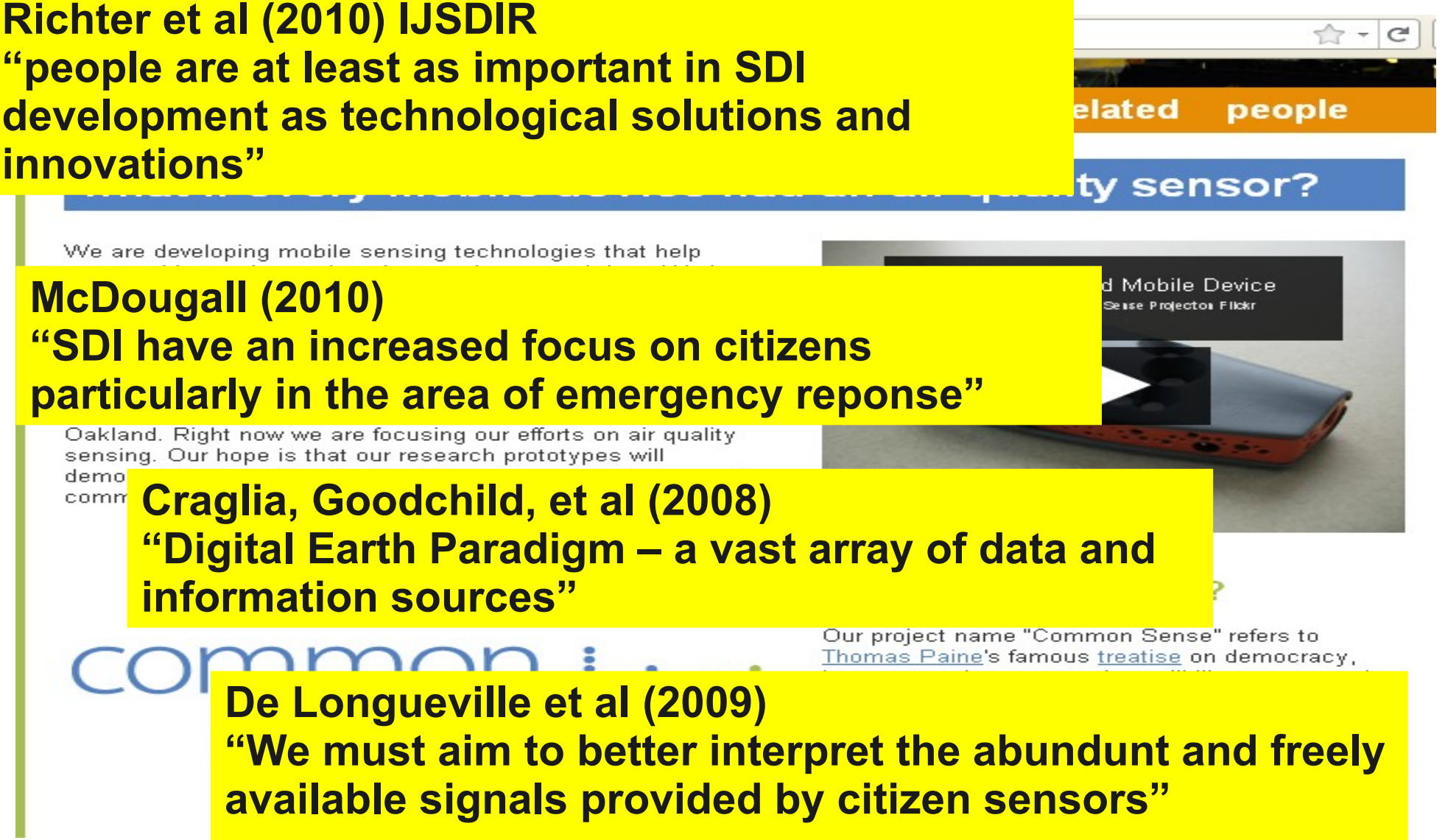
common sense

Why Common Sense?

Our project name "Common Sense" refers to Thomas Paine's famous treatise on democracy, because we hope our project will illustrate ways in which everyday citizens can use sensing technologies to conduct citizen science and participate in the political process. The project name also emphasizes the notion that sensing data can be shared in common by communities, that individuals can join together to collect and analyze environmental data.

There is potential for a “continous stream of data” from citizen sensors

Richter et al (2010) IJSDIR
“people are at least as important in SDI development as technological solutions and innovations”



McDougall (2010)
“SDI have an increased focus on citizens particularly in the area of emergency reponse”

Craglia, Goodchild, et al (2008)
“Digital Earth Paradigm – a vast array of data and information sources”

De Longueville et al (2009)
“We must aim to better interpret the abundunt and freely available signals provided by citizen sensors”

Defining Spatial Data Infrastructures (SDI)

- "***Frameworks of technologies, data, policies, institutional arrangements, and people – aimed at increasing availability, understanding, and accessibility of data for people and policy***" (IJSDIR, 2009)
- National SDI are key building blocks
- INSPIRE – in progress for about 8 years now – potentially another 8 years before implementation [legal] (http://inspire.jrc.ec.europa.eu/index.cfm/pageid/48)
- Technology + spatial data landscape changing very quickly [Positioning technologies, crowdsourcing, on-the-fly capture, sensors, open access, open source]

There are many challenges in the implementation of SDI

- **Commercial "SDI"**
  - Google Maps, Bing maps
    - Critical mass of users, no obligation on quality, limited geo-analysis, varied licenses, driven by "market forces"
- **Public Sector SDI**
  - INSPIRE, GMES, GEOSS, UN, etc
    - Bringing together policy, institutions, people – to provide government – goverment data/information flows
- **VGI-driven SDI**
  - OpenStreetMap
    - Critical mass of users, organisational issues, license issues
    - Centralised geodatabase (with standards) + established contribution mechanism

The relationship between VGI and SDI is not a simple one

- **Question:** Is VGI: complimentary, supportative, incompatible, a replacement, an opponent, more popularly/understandable by citizens?
- **Spatial Data Infrastructures** – usually deep investment of resources – huge scale of financial and people resources
- **Volunteered Geographic Informatio**n – currently low or no investment costs – supported by voluntary donations/fundraising [wikipedia], on the back of commercial ventures (Flickr)

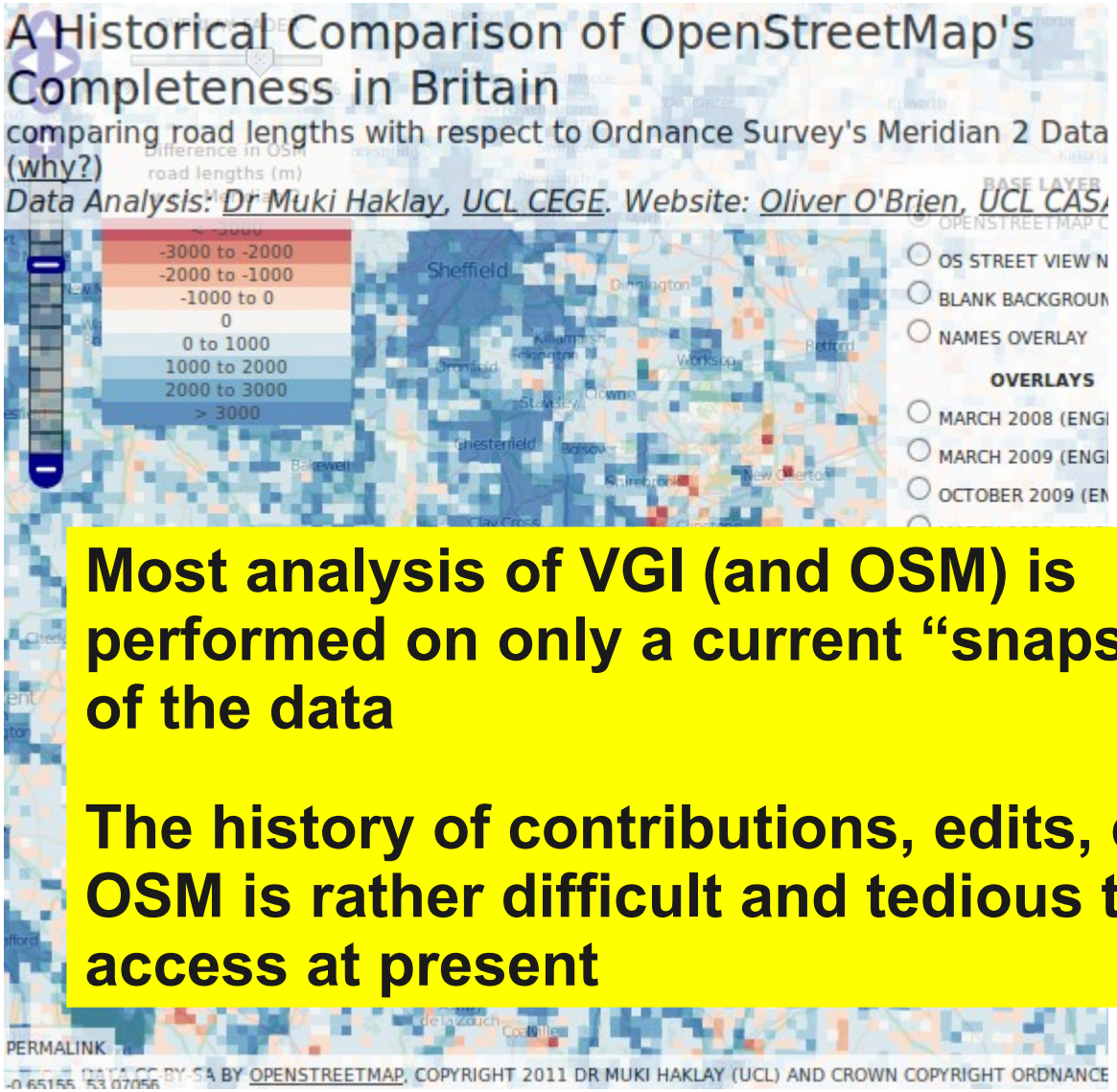
Comparison of VGI and SDI

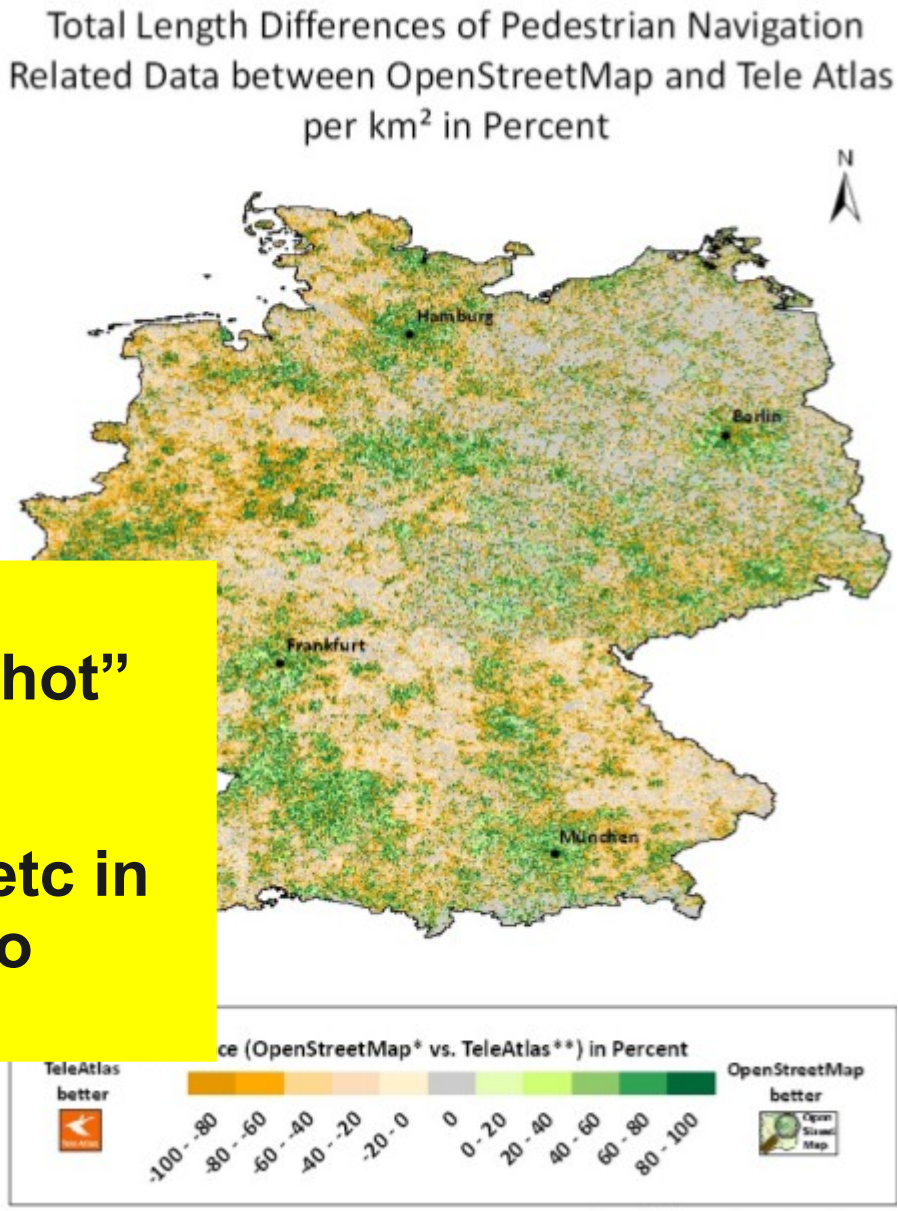
	Government-centric SDI	User-centric VGI
SDI Structure	Highly structured	Ad-hoc and simplistic
Standards	Close adherence to standards	Loose based on communication standards
Maturity of data holdings	Highly mature	New and current but variable
Spatial Accuracy	Complying with mapping standards	Variable
Metadata	Contain detailed metadata	Few standards – ICT based
Openness	Highly controlled	Often new data sets
Data Update	Often slow and overly bureaucratic	Fast and flexible
Potential data maintenance and collection base	Limited to the budget and staffing	Potentially a huge user and contributor base
Adaptability	Low – retrained by mandate, resources and bureaucracy	High

http://www.gsdi.org/gsdiconf/gsdi12/slides/7.3a.pdf

Historical Analysis of Contributions to OpenStreetMap towards understanding how VGI could fit into Spatial Data Infrastructures

"*Those who cannot learn from history are doomed to repeat it*".
- George Santayana, Life of Reason 1905





Most analysis of VGI (and OSM) is performed on only a current "snapshot" of the data

The history of contributions, edits, etc in OSM is rather difficult and tedious to access at present

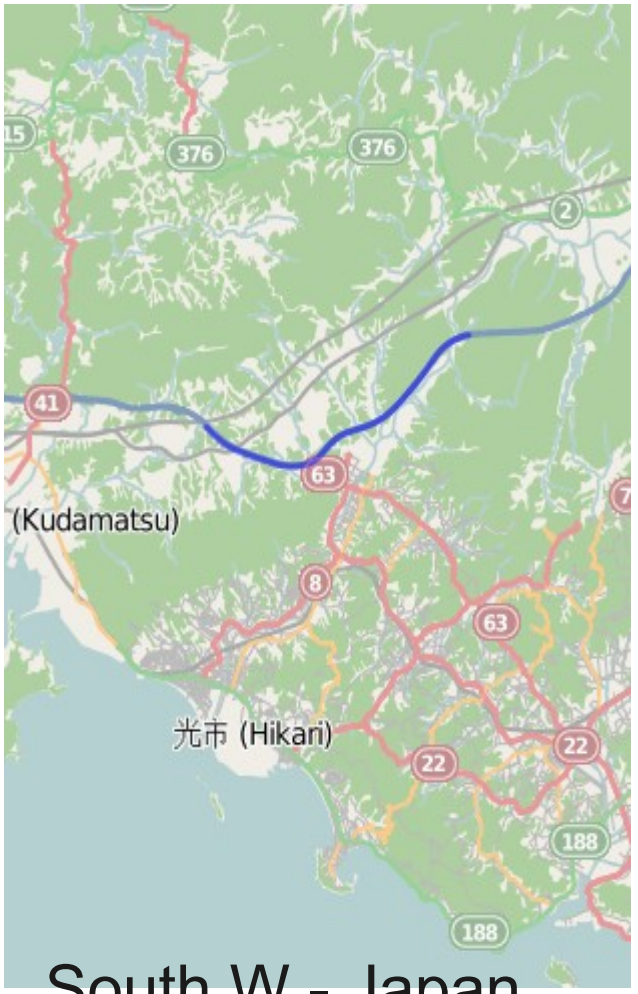
http://orca.casa.ucl.ac.uk/~ollie/osmcompare/

Zielstra, D. and Zipf, A. (2010):
A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany.
AGILE 2010. The 13th AGILE Guimarães, Portugal.

Most Heavily Edited Features on OSM

From very recent “Planet.OSM”

South W - JAPAN
ID = 24770043
Versions = 400



South W - Japan
ID = 24761888 ,
Versions = 399

Lausanne,
Switzerland
ID = 23381485,
Versions = V323

Most Edited: Highway - Japan

24770043

- 400 VERSIONS
- 3 editors
- Start: 8-6-2008
- End: 5-1-2011
- One editor exclusively for first 398 versions!
- 10 Changesets
- Nodes: (Start, Max, Final Version)
- (278, 629,26)

2nd Most Edited: Highway - Japan

- 399 Versions
- 1 editor
- Start: 08-06-2008
- End: 12-10-2009
- One editor for all 399 versions
- 7 Changesets
- Nodes: (444, 652,204)

24761888

3rd Most Edited – Pier - Lausanne (Shows Most Diversity in Editing)

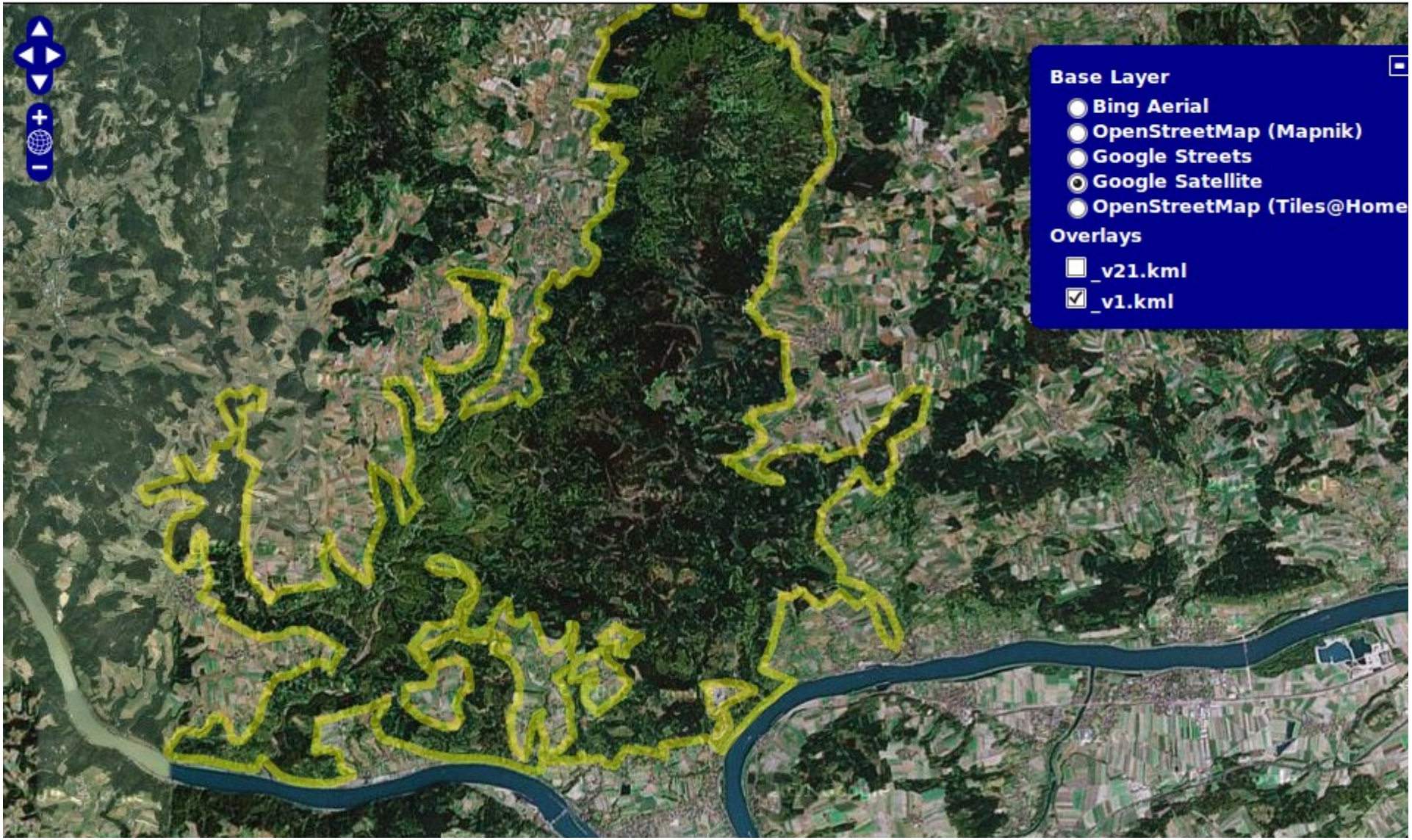
- 323 versions
- 37 editors
- Start: 21-03-2008
- End: 19-10-2010
- One editor created 153 versions
- 149 Changesets
- Nodes: (444, 652,204)

24770043

SOTM-EU 2011 -
“Volatile Volunteered Geographic Information”

Austria: Landuse Polygon (OSM-ID = 13864876)

KML layer overlayed



MARBACH AN DER DONAU
Version 1 ("2007-11-23 20:59:12")

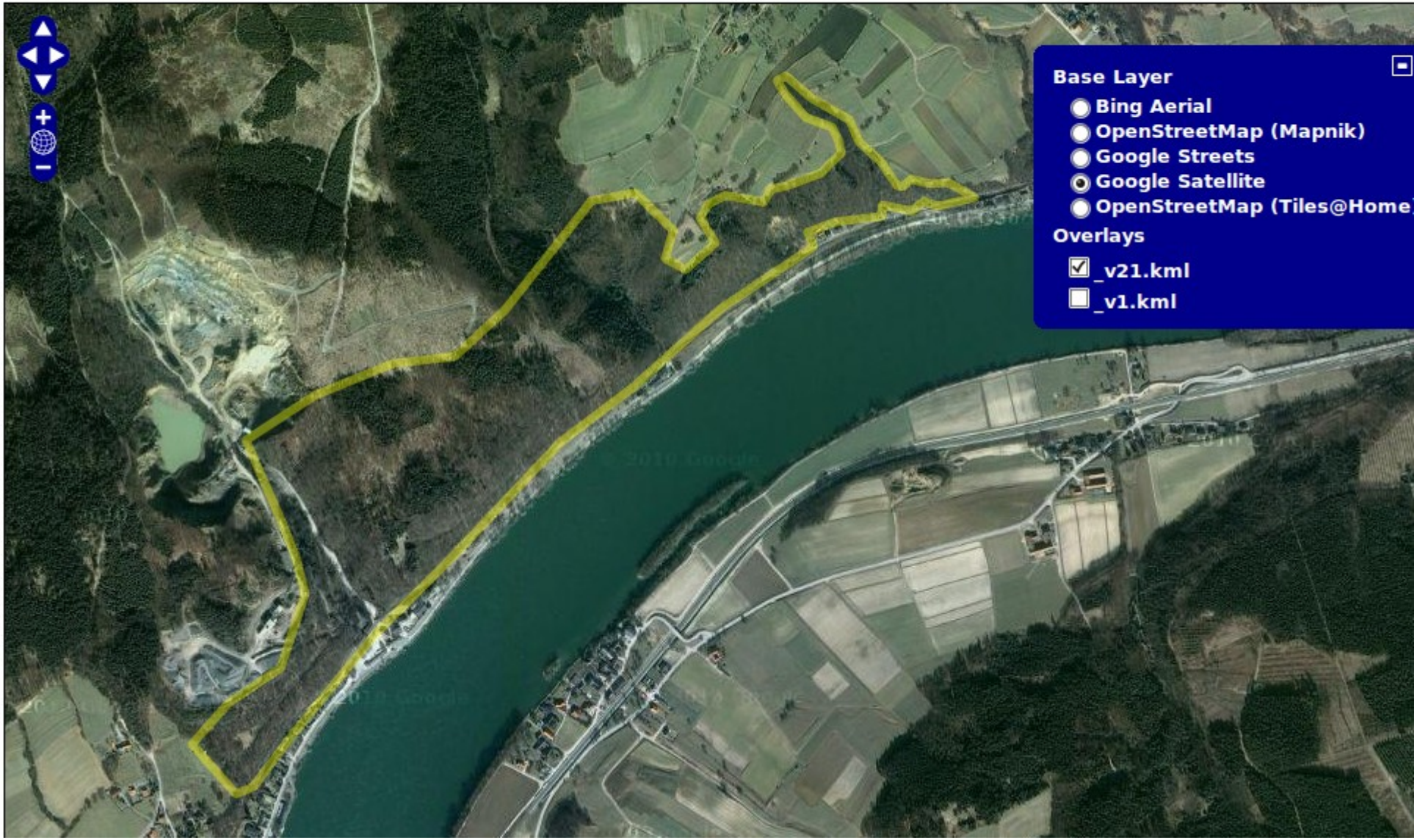
Austria: Landuse Polygon (OSM-ID = 13864876)

KML layer overlayed

MARBACH AN DER DONAU
Version 21 ("2011-03-09 21:08:11" – 9 Unique Users, 1 Tag (landuse = forest))

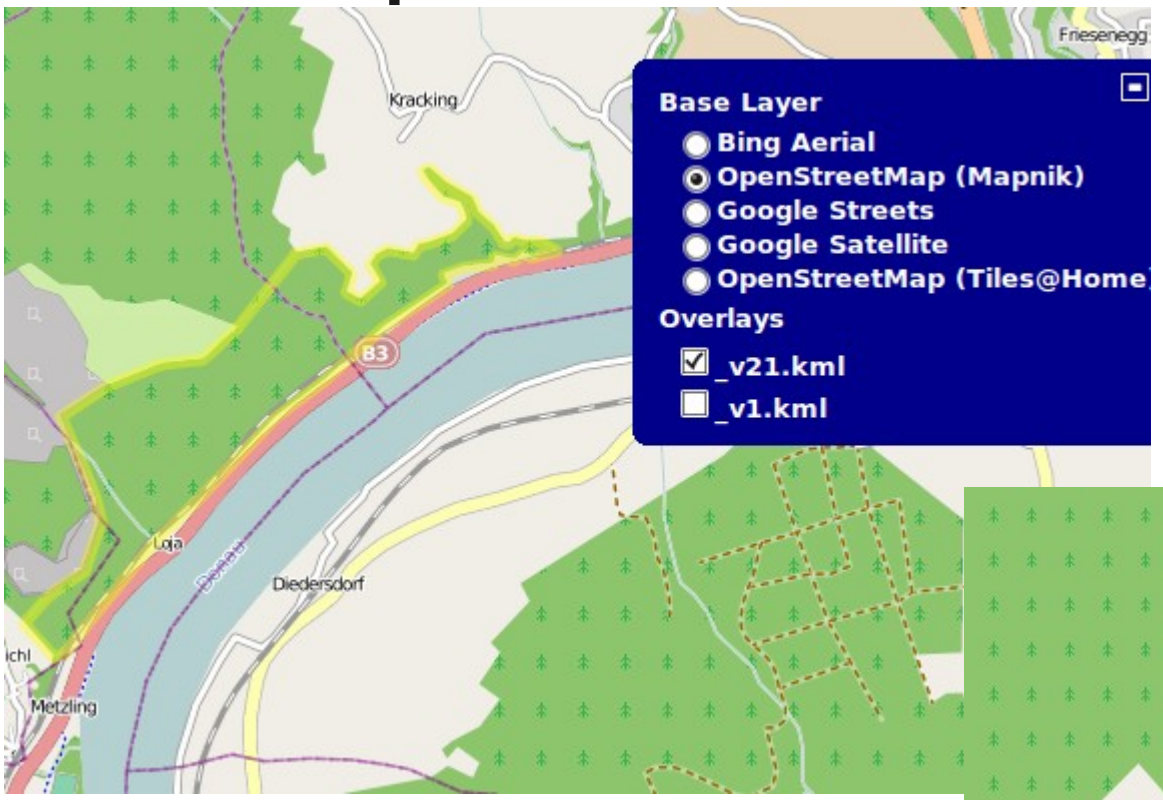
Austria: Landuse Polygon (OSM-ID = 13864876)

KML layer overlayed



MARBACH AN DER DONAU
Version 21 ("2011-03-09 21:08:11" – 9 Unique Users, 1 Tag (landuse = forest))

Further editing of 13864876 has taken place – breaking the polygon

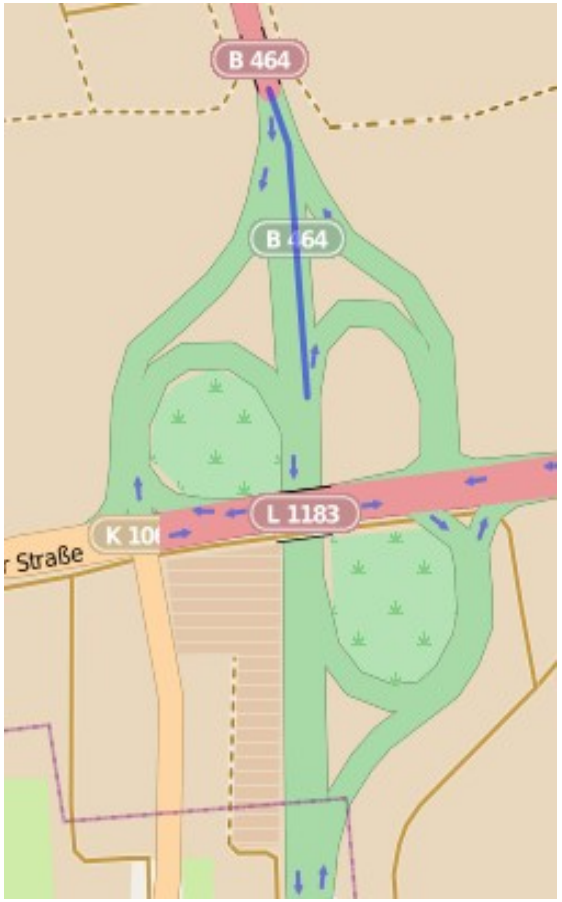


Wed, 06 Apr 2011
20:16:38
Tags deleted (0 tags)
Version 23

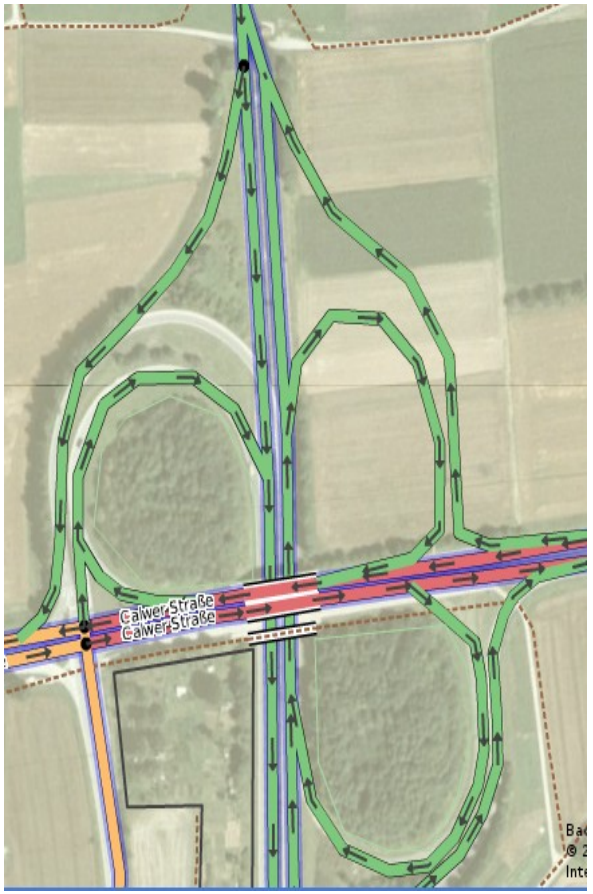
This is an invalid polygon

It cannot be cartographically rendered
Without any tags.

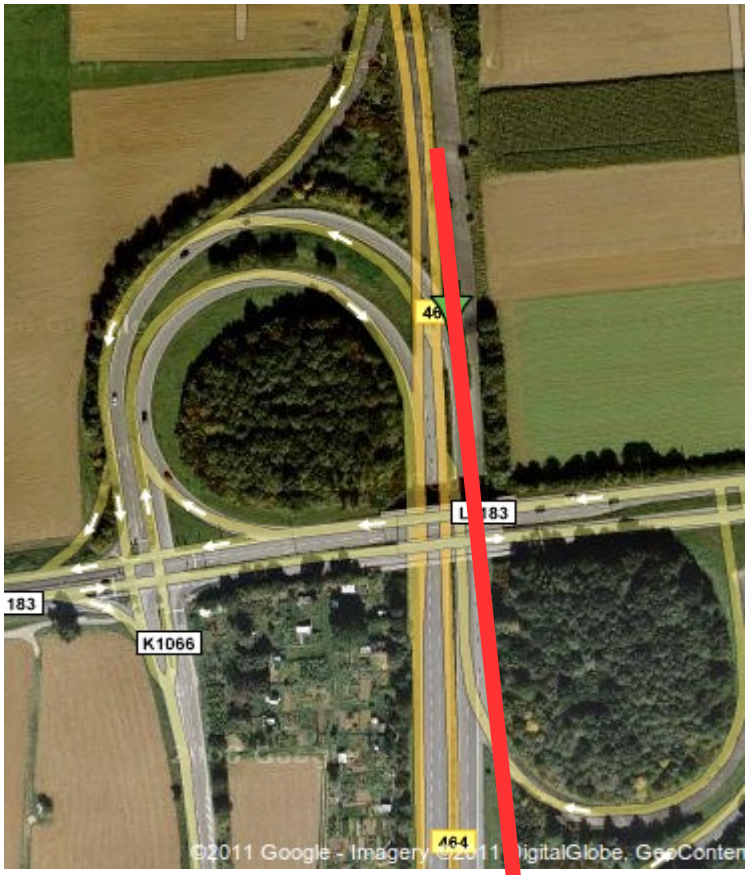
Böblingen,Baden-Württemberg



OSM TILES



POTLATCH 2



GOOGLE AERIAL

http://www.openstreetmap.org/?way=23704273

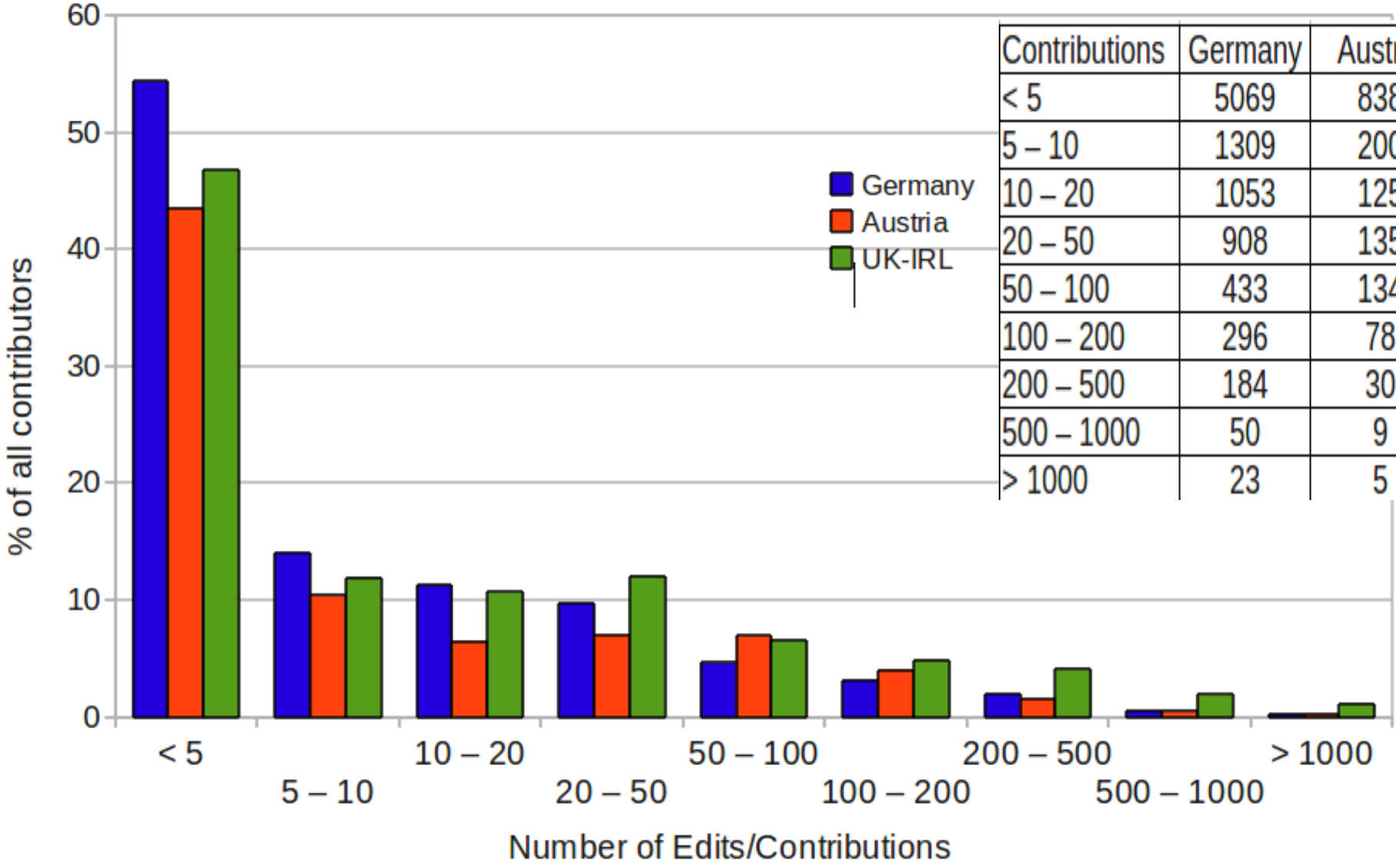
Overall summary of our study of OpenStreetMap Ireland and UK

- Total of **220,779** edits
- England – **171,907**
- Scotland – **21,931**
- Wales – **13,825**
- Ireland – **13,116**
- There are **3084** unique contributors
- **25 users** edited in all 4 countries
- **56 users** edited all 5 chosen feature types
- Earliest contributions begin in **May 2007**

Germany – OSM Database

- We biased our selection of features toward LANDUSE
- Total **10603** features
- Amenity **510**, Highway **684**, Landuse **6998**, Natural **590**, and Waterways **1821**
- Total Objects = 259,049, Consecutive Versions = 248,426
- Unique Contributors = **9325**
- **234 (2.5%)** Contributors edited ALL feature types

Contributor effort in HIGH edit



Contributions	Germany	Austria	UK-IRL
< 5	5069	838	1442
5 – 10	1309	200	368
10 – 20	1053	125	329
20 – 50	908	135	372
50 – 100	433	134	202
100 – 200	296	78	150
200 – 500	184	30	126
500 – 1000	50	9	59
> 1000	23	5	36

AUSTRIA (1926 unique editors)

UK & Irl (3084 unique contributors)
Germany (9325 unique contributors)

Contributions are mostly restricted to very low volume edits

- Contributors who performed **10 or Less EDITS**
- **Ireland** (66% - from 355)
- **England** (58% from 2505)
- **Scotland** (61% from 379)
- **Wales** (68% from 316)
- **Austria** (74% from 1926)
- **Germany** (70% from 9325)

What do “one time” contributors do?

- **UK & Ireland: 1008 users** created polygons and polylines (**44 users** – one and only edit)
- **Germany: 944 users** created polygons and polylines (**62 users** – one and only edit)
- **Austria: 342 users** created polygons and polylines (**20 users** – one and only edit)
-
- These results indicate that this “high-edit” subset of OSM exhibits very similar characteristics to the entire global OSM database

Analysis of the time between consecutive edits to objects

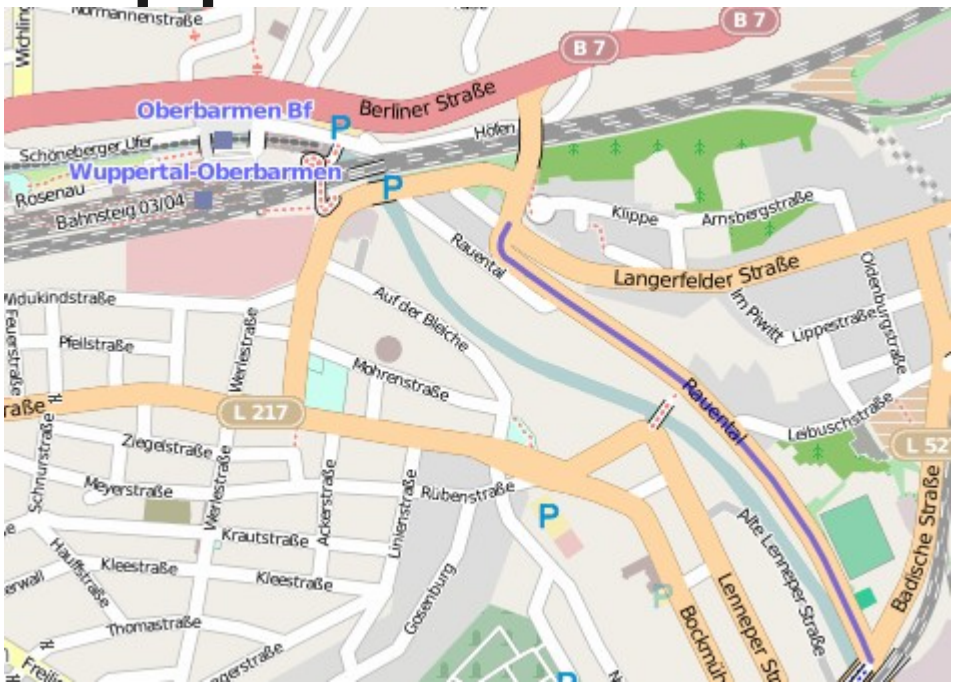
Time Between Versions	Austria (%)	UK/Irl (%)	Germany (%)
<= 5 minutes	22.80	41.25	25.98
5 mins to 30 mins	7.03	8.25	9.42
30 mins to 1 hour	2.01	1.81	2.48
1 hour to 2 hours	1.42	1.47	1.88
2 hours to 12 hours	3.47	2.94	4.28
12 hours to 24 hours	3.05	2.62	3.93
24 hours to 1 week	10.57	8.67	13.26
1 week to month	13.42	9.30	13.94
> 1 month	36.00	23.63	24.84

Austria (3367 Objects – 45,369 consecutive versions)
UK + Ireland (10693 Objects – 206,843 consecutive versions)
Germany (10603 Objects – 248,426 consecutive versions)

User behavior analysis? Length of editing sessions . . . return to edit (delay... loss of interest)

Tagging spatial objects

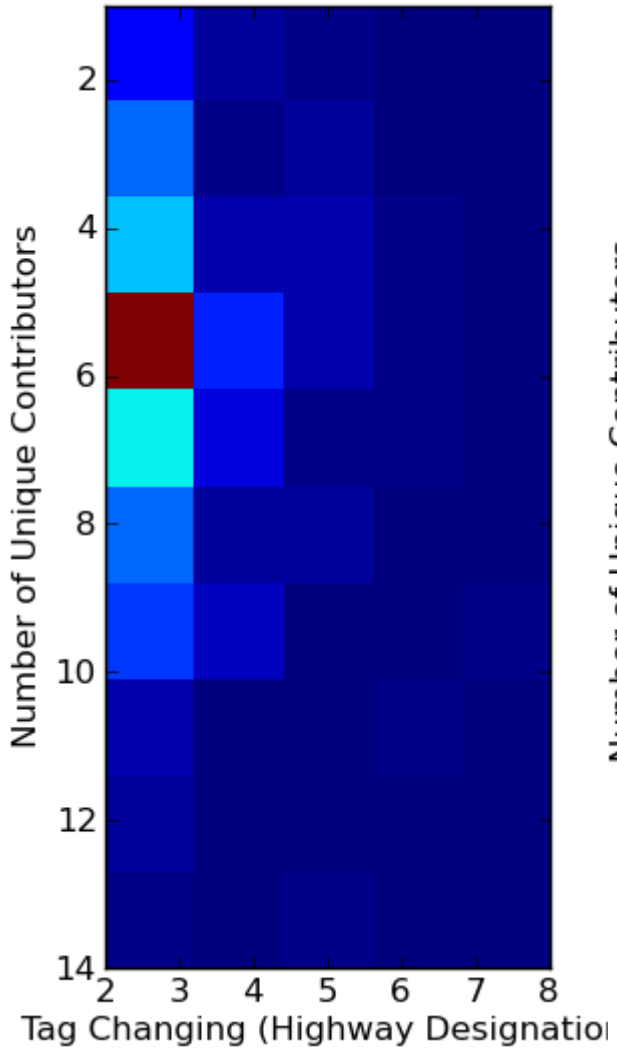
Wuppertal Oberbarmen



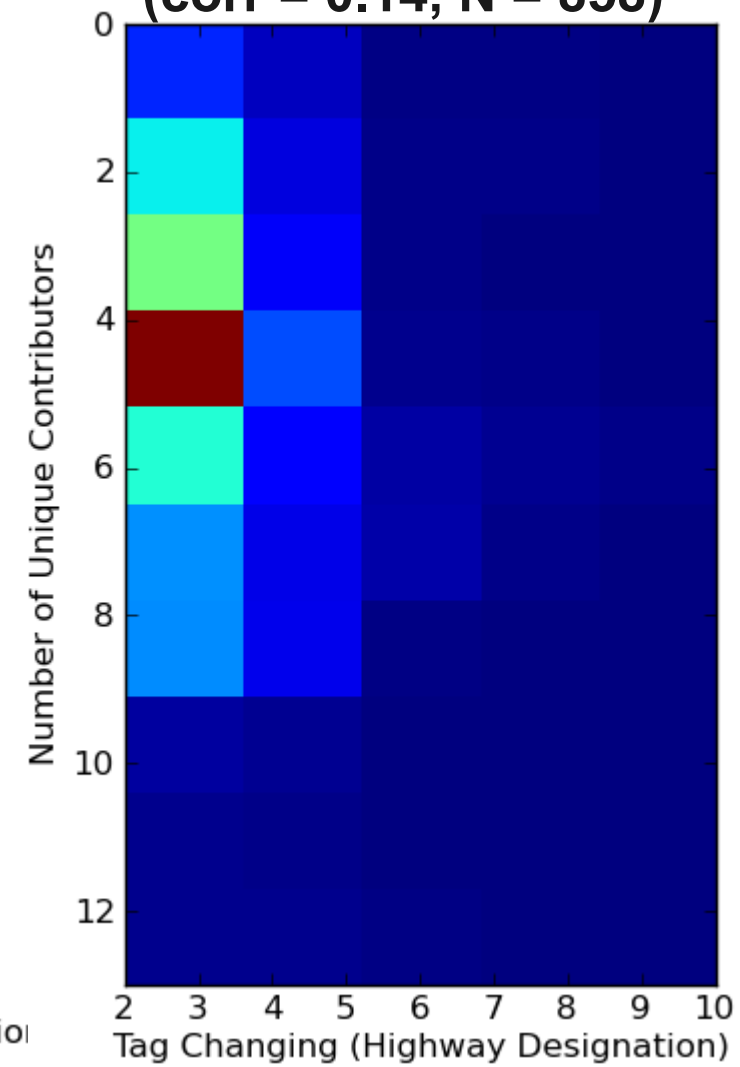
Version	Edit Date	#Users	UserID	KEY: highway	#Tags	#Nodes
1 (First)	19-10-2007 (20:21)	1	15843	No highway tag specified	1	9
2	19-10-2007 (20:21)	1	15843	unclassified	2	9
4	16-05-2008 (14:38)	3	36745	primary	4	10
6	16-05-2008 (14:48)	3	36745	secondary	4	10
7	16-05-2008 (14:48)	3	36745	primary	4	10
8	31-05-2008 (10:05)	4	41974	secondary	4	10
11	05-06-2008 (11:51)	4	36745	primary	4	10
16	19-07-2008 (16:42)	5	41974	secondary	4	10
17	25-07-2008 (11:08)	6	39350	primary	4	10
18	22-09-2008 (18:59)	7	54643	secondary	5	10
29 (Final)	31-03-2011 (17:10)	13	177346	secondary	5	22

Tag “flip flopping” - highways

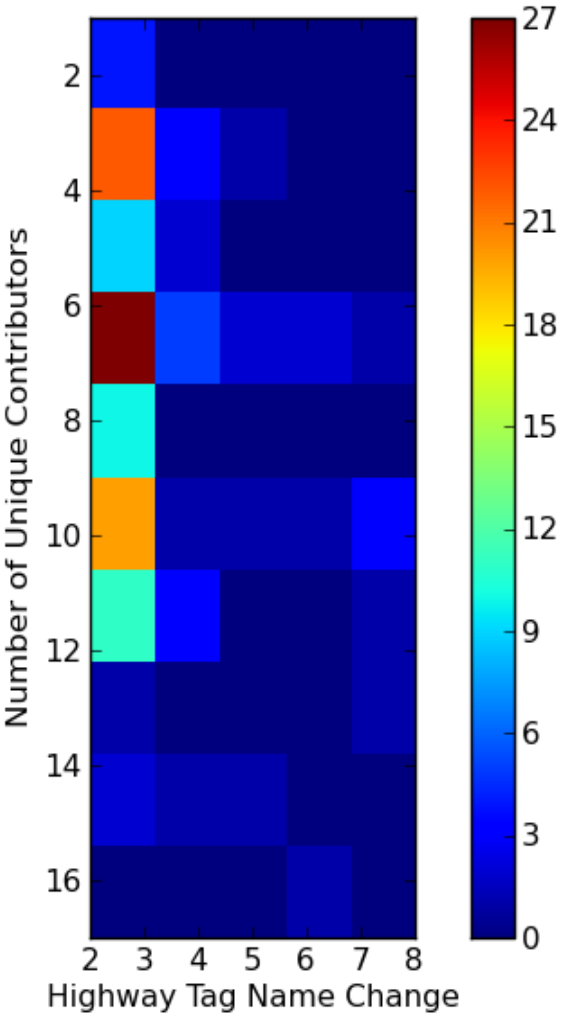
Austria (corr = 0.018, n = 395)



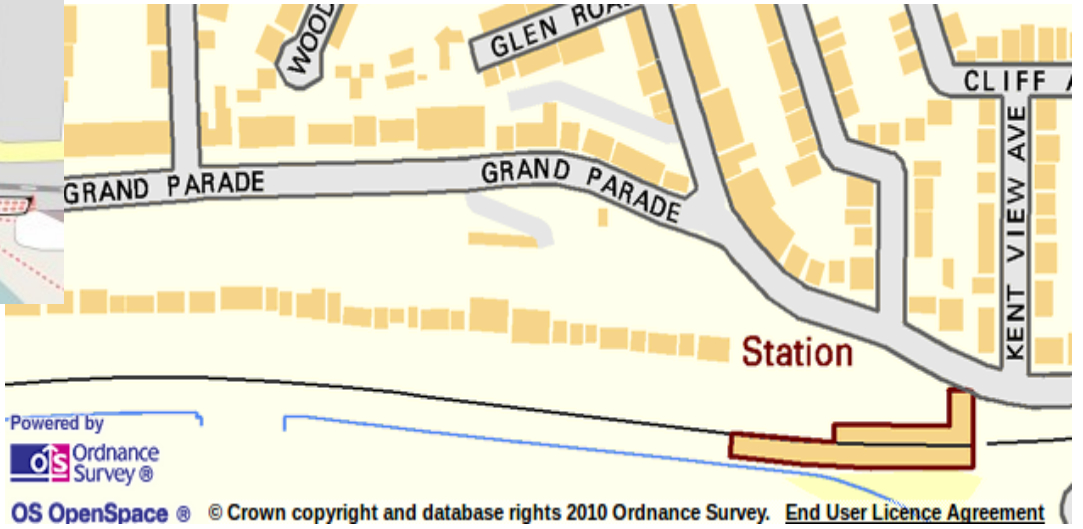
UK and Ireland (corr = 0.14, N = 858)



Germany (corr = 0.29, N = 148)

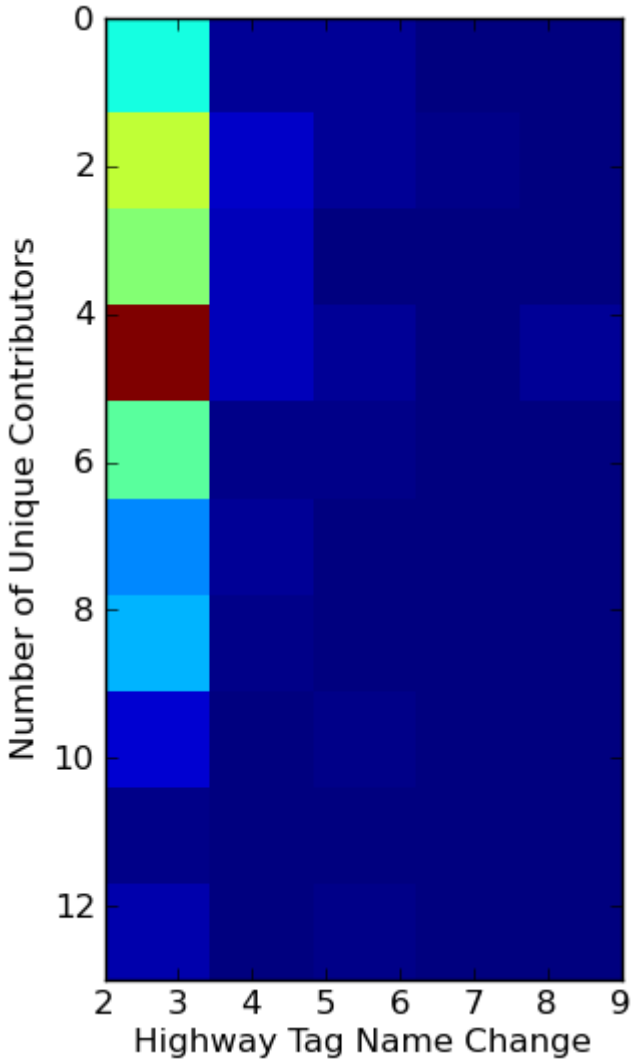


Name Change Example (UK)



OSM_ID =4803031
v1,23/6/2007
v3,name=**Thames Drive**, userid=64941,06/09/2008
v5,name=**Belton Way West**,userid=64941,06/09/2008
v21,name=**Belton Gardens**,userid=20573,28/02/2009
v22,name=**New Road**,userid=20573,28/02/2009
v26,name=**Grande Parade**,userid=320358,24/07/2010
current version

The naming of highway objects (UK)



- UK and Ireland (**corr = -0.13, N = 381**)
- Local variations
- Spelling errors
- Contributor disputes

Crowdsourcing the names of street/road features in OSM

England (24276789, 2 contributors)
7 Changes
"Oakthorp Drive"
"Over Green Drive"
"Oak Thorp Cr"
"Oak Thorp Dr"
"Oak Thorp Dr; Broomcroft Rd"
"Oak Thorp Drive"
"Oak Thorpe Drive"

Austria (4771112, 3 contributors)
5 changes
"Raststätte Kapellerfeld"
"Autobahnraststätte Kapellerfeld (in Bau)"
"Autobahnraststätte Deutsch-Wagram (in Bau)"
"Raststation Deutsch-Wagram (in Bau)"
"Raststation Deutsch-Wagram"

Scotland (23602699, 2 contributors)
5 changes . .
"phenox cres"
"Phenoix cres"
"Phenoix crescent"
"Phenoix Crescent"
"Phoennoix Crescent"
"Phoenix Crescent"

Scotland (4755815, 12 contributors)
5 changes . .
"A199"
"Edinburgh Road"
"Milton Road East"

Conclusions and Close

OSM operates in the classic crowdsourcing model

- **Four fundamental challenges**
  - How to recruit and retain the crowd?
  - What contributions does the crowd make?
  - How are these contributions combined to solve a specific problem?
  - How can we evaluate the crowd and their contributions?

Doan and Ramakrishnan, Communications ACM 54 (4), 2011

Severe lack of metadata in OSM

- Our analysis of OSM history reveals only sporadic inclusion of meaningful metadata
- Authors not documenting changes/sources etc...
- Documentation resolution – at node, polygon, or relation level?
- *There are contributors who are dilligent and consciencious about documenting their contributions but this is not the behaviour of the majority or the ``crowd''* (Mooney and Corcoran, 2011).

VGI needs (1) Strong Use Cases and (2) Influential Users

- "*And it ought to be remembered that there is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things. Because the innovator has for enemies all those who have done well under the old conditions, and lukewarm defenders in those who may do well under the new*".
- The Prince (Chap. IV) Niccolò Machiavelli 1532

Data quality issues in VGI could be addressed under 3 strategies

- **[Enforce QA/QC]** Re-inforce control on the VGI production chaing by establishing a standardised data production metadata and limiting volunteer numbers
- **[Moderators]** Quality control as a “voluntary” process – users then act as filters/gatekeepers in a model similiar to Wikipedia
- **[Acceptance]** Accept the opportunities and abundance of VGI – reliable information/data is extracted – cross validation mechanisms must be developed

Following from Schade et al, 2010 (WebMGS) + De longueville et al (2009)

Overall, VGI presents many issues which must be addressed

Quality and reliability are critical issues – but these are inherently coupled with VGI sustainability

VGI needs champions – and it needs to be used in more than "peripheral" applications (Keßler et al, 2009 – Over et al, 2011)

Technology appears to be causing a rapid convergence of VGI (people) and SDI (policy - government)

THANK YOU FOR YOUR ATTENTION!

Email: peter.mooney@nuim.ie