Geographies of crowdsourced information and their implications

Andrea Ballatore*

Department of Geography
Birkbeck, University of London

Stefano De Sabbata

School of Geography, Geology and the Env.,
University of Leicester

Birkbeck
UNIVERSITY OF LONDON

aballatore.space

@a_ballatore

VGI-Alive, AGILE
June 2018, Lund, Sweden

Outline

1. Where are we?
2. Core research questions
3. Paradigm limitations
4. Beyond the usual suspects
5. Case studies: Search behaviour
6. **[Stefano will do the rest]**

Where are we?

The success of crowdsourcing

https://ideascale.com

"85% of the top global brands have used crowdsourcing in the last ten years. […] According to Gartner, 75% of the world's high performing enterprises will be using crowdsourcing by 2018."

(Deloitte, 2016)

It is when new, successful technologies withdraw into the "woodwork of everyday banality" that their effects become real and profound.

(Vincent Mosco, 2004)

The success of crowdsourcing

https://ideascale.com

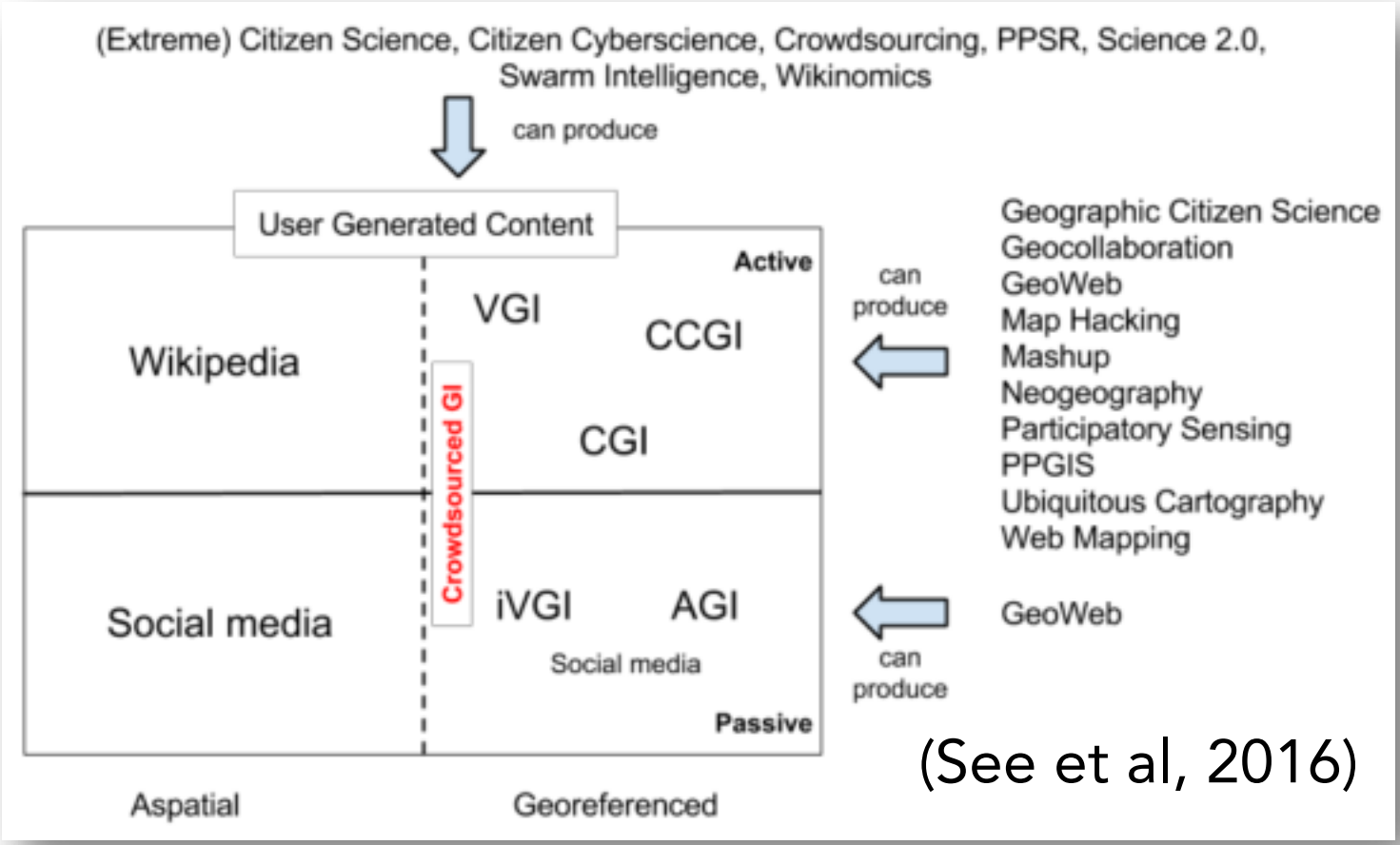
"85% of the top global brands have used crowdsourcing in the last ten years. […] According to Gartner, 75% of the world's high performing enterprises will be using crowdsourcing by 2018."

(Deloitte, 2016)

It is when new, successful technologies withdraw into the "woodwork of everyday banality" that their effects become real and profound.

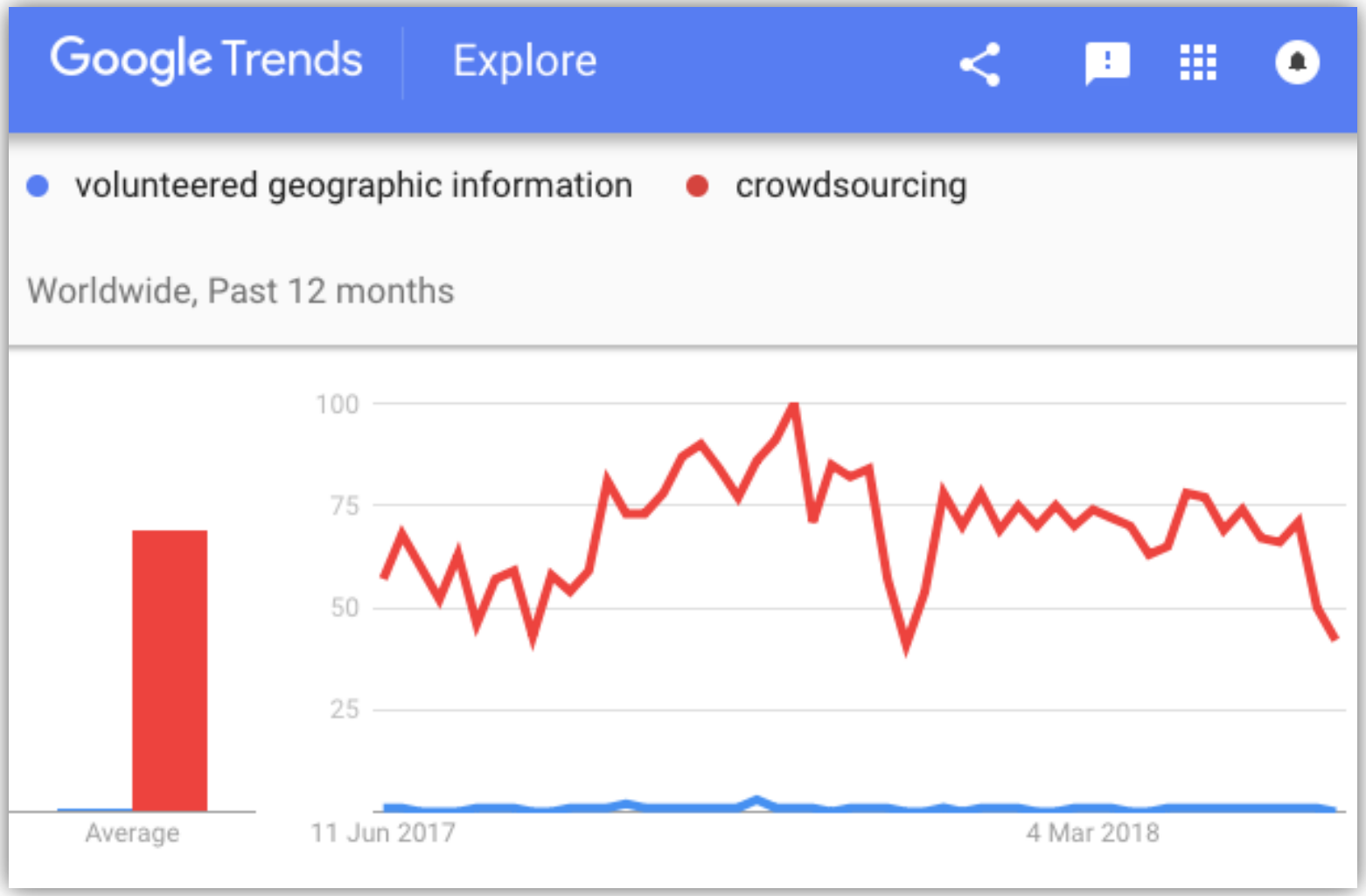
(Vincent Mosco, 2004)

Crowdsourcing + geolocation: A mature field



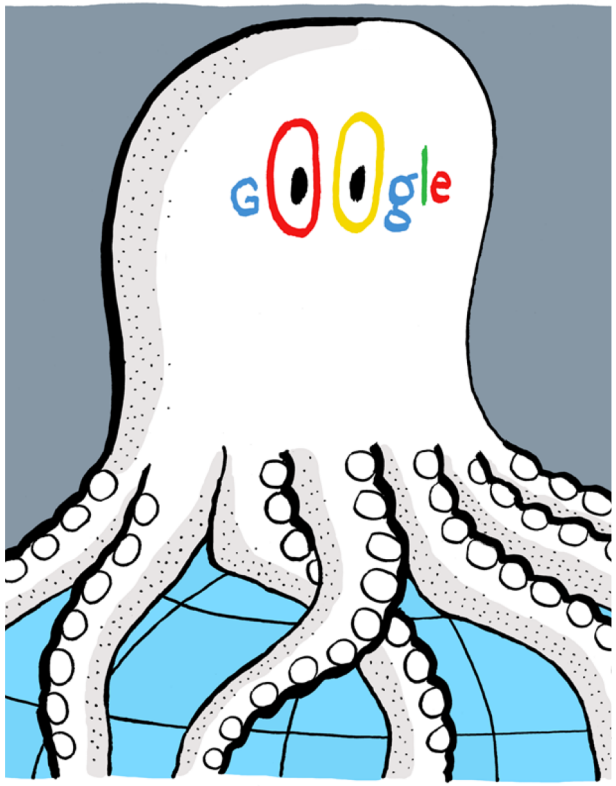
(See et al, 2016)

Volunteered or Crowdsourced GI?



Placing the crowds

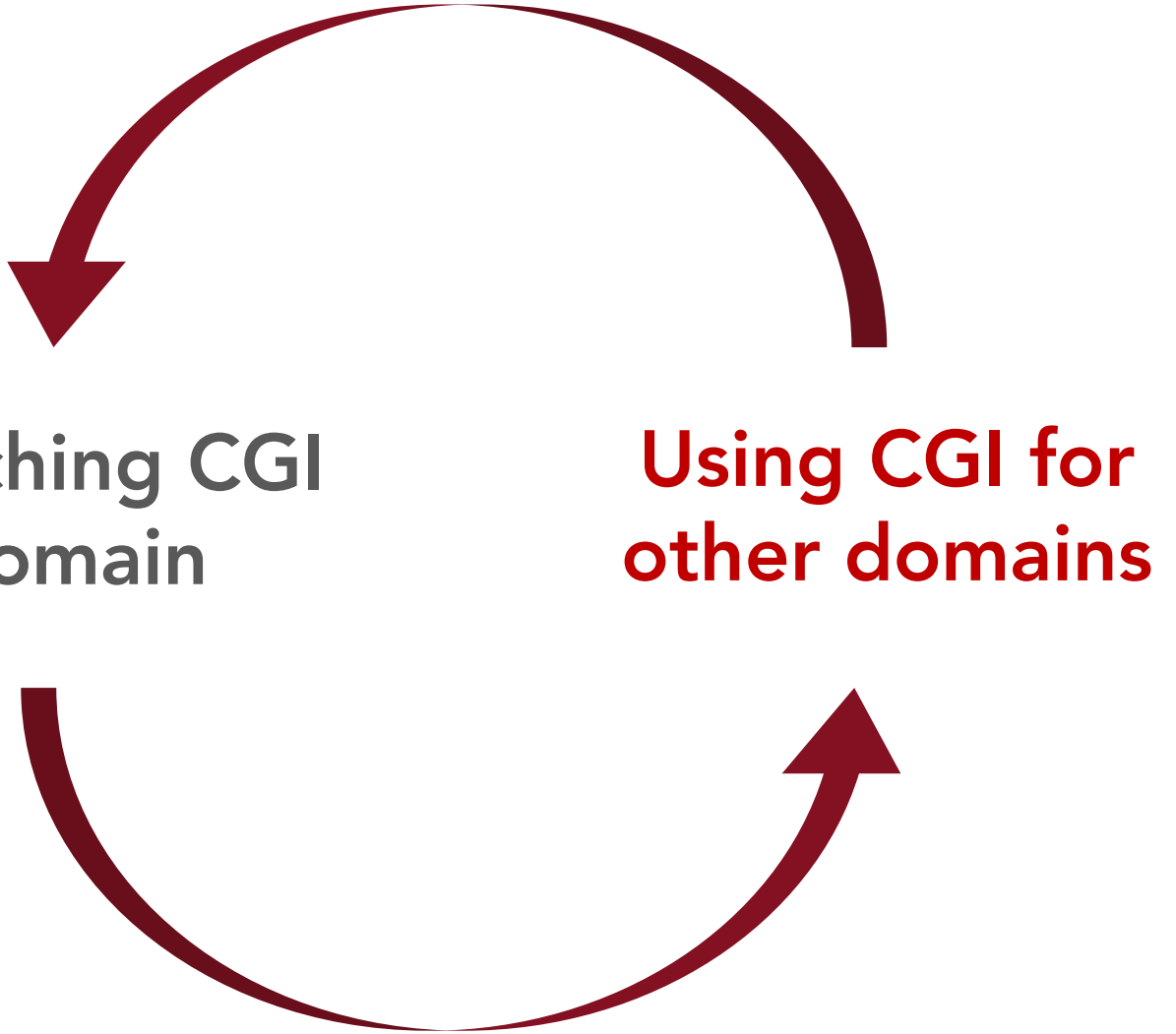
- **Crowdsourced geographic information (CGI)**
- From experimental phase to **oligopoly** (e.g., Google, Facebook)
- From **cyberoptimism** to **cyberpessimism**



https://www.cagle.com

Placing the crowds

- Crowdsourced geographic information (CGI)
- From experimental phase to **oligopoly** (e.g., Google, Facebook)
- From **cyberoptimism** to **cyberpessimism**


Researching CGI as domain
Using CGI for other domains

Core CGI research

1. **Who** are the contributors and **why** do they engage in spatial information production, and what **incentives** work or do not work? How do they **collaborate** and **organise**? How do we include **marginalised** communities?

   (Budhathoki and Haythornthwaite 2013)

2. How can we calculate the **quality** and fitness for purpose of crowdsourced data in a reliable, preferably intrinsic way? (Goodchild and Li 2012)

3. What are the **limitations** of such models and what are their spatial, epistemic and cultural biases? (Dodge and Kitchin 2013)

Core CGI research

1. **Who** are the contributors and **why** do they engage in spatial information production, and what **incentives** work or do not work? How do they **collaborate** and **organise**? How do we include **marginalised** communities?

   (Budhathoki and Haythornthwaite 2013)

2. How can we calculate the **quality** and fitness for purpose of crowdsourced data in a reliable, preferably intrinsic way? (Goodchild and Li 2012)

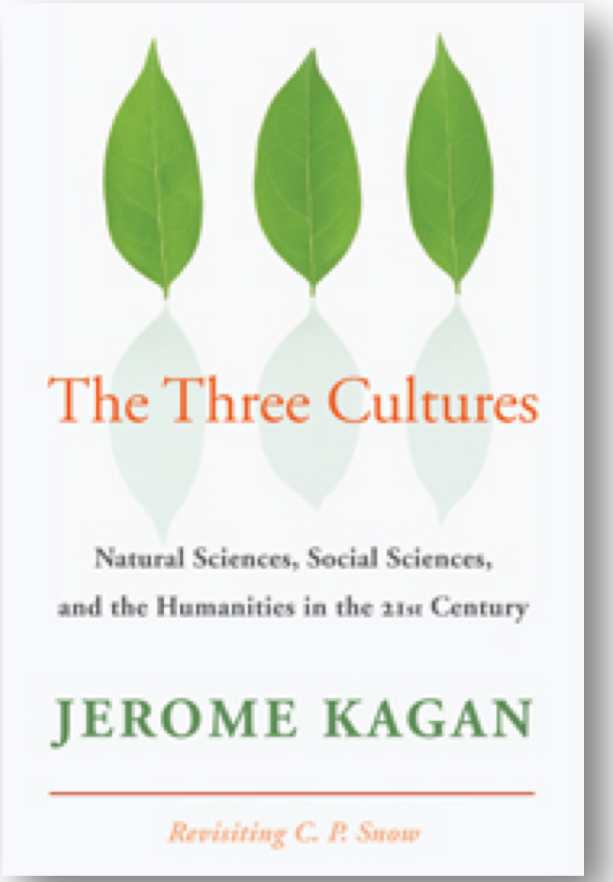
3. What are the **limitations** of such models and what are their spatial, epistemic and cultural biases? (Dodge and Kitchin 2013)

CGI for other domains

Natural sciences: biology, climate science, Earth sensing

Social sciences: urban planning, transportation, public health, economics, human geography

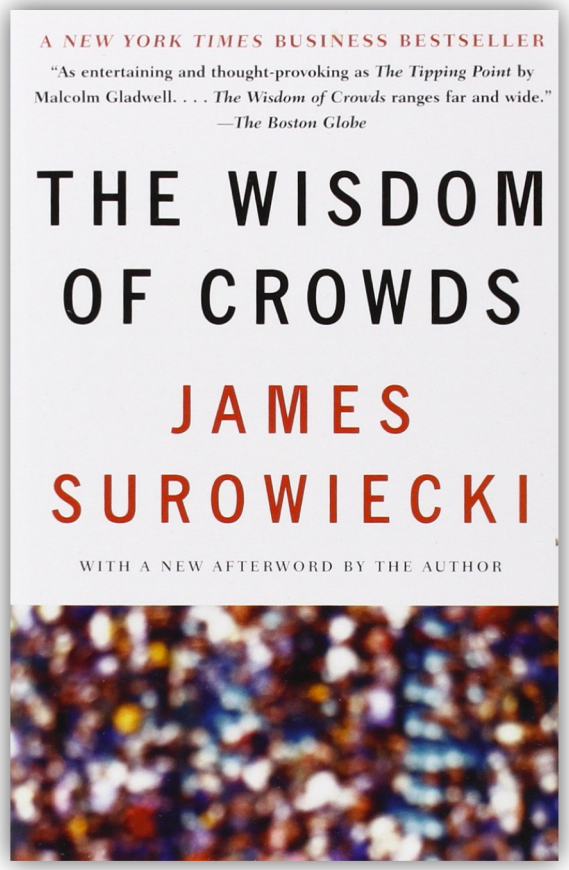
Humanities: digital humanities, history, cultural analytics



Limitations of crowdsourcing

Limitations of the paradigm

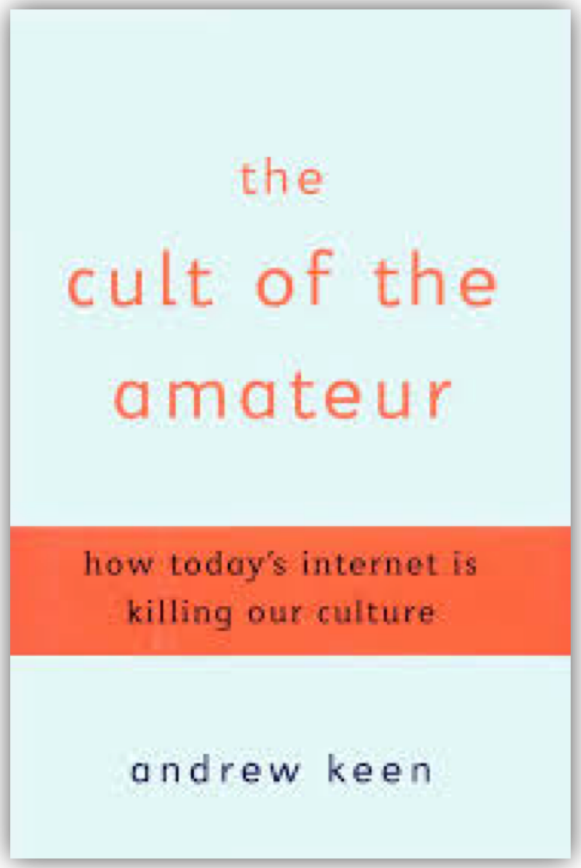
- "Ignorance of crowds" (Carr, 2007)
- Conditions for wisdom
- Menial work, no real innovation/creativity
- Undermining paid work
- Variable quality



2004

Limitations of the paradigm

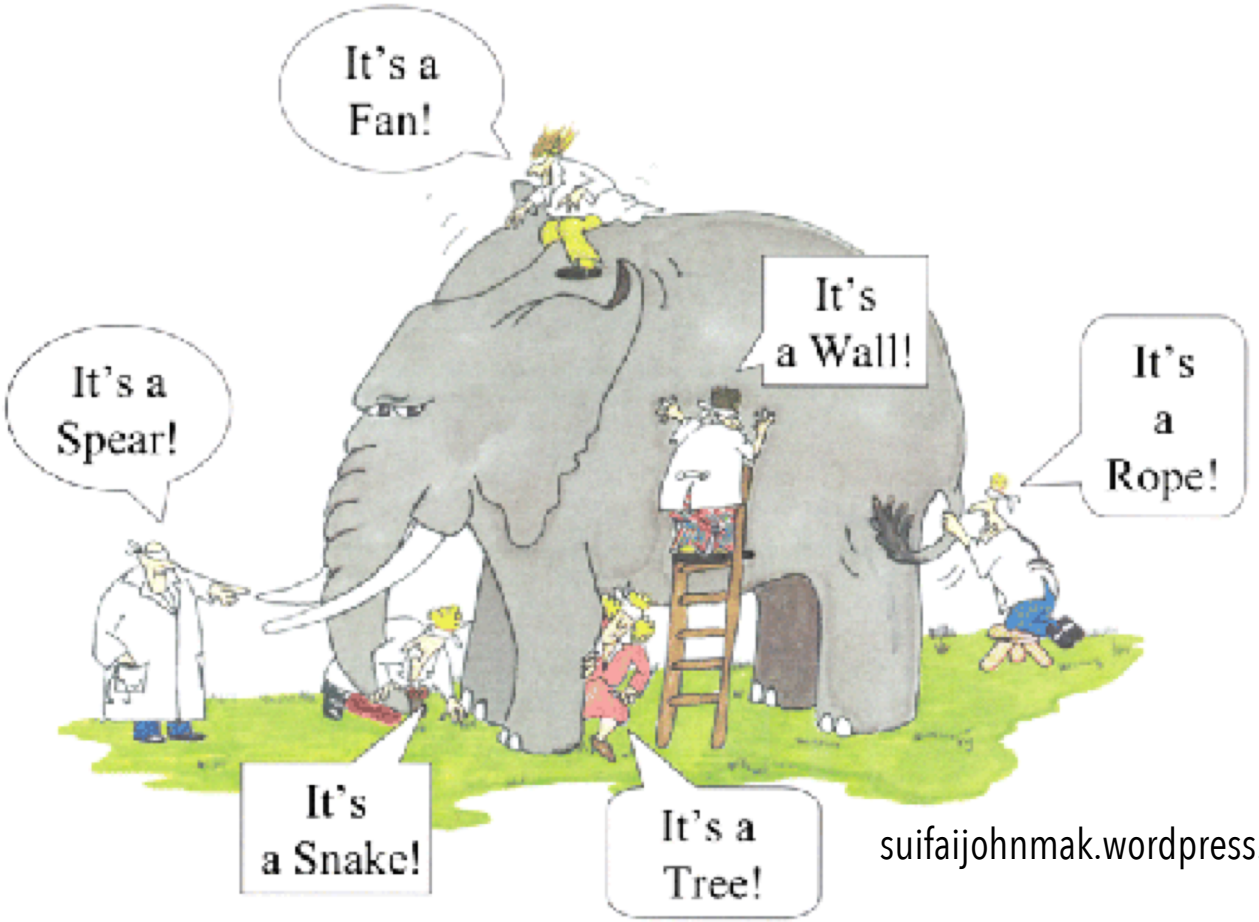
- "Ignorance of crowds" (Carr, 2007)
- Conditions for wisdom
- Menial work, no real innovation/creativity
- Undermining paid work
- Variable quality



2007

Limitations of the paradigm

- Large volumes of information do **not** imply usefulness or fitness for purpose
- We need **representative samples**, not large samples (e.g., random sample of 1,000 > 1M non-random)



suifaijohnmak.wordpress.com

Each CGI source is a **particular viewpoint** and will return a different image of the social and natural world.

Diversity/biases of CGI

- **Thematic**: e.g., tourism, outdoors, typical/atypical behaviour, sharing bias
- **Demographic**: Western Educated Industrialised Rich Democratic (WEIRD) (not always!)
- **Social**: 90%-9%-1%, hyperactive minorities of contributors
- **Geographic**: urban/rural, developed/developing, central/peripheral, human/natural

Diversity/biases of CGI

- **Thematic**: e.g., tourism, outdoors, typical/atypical behaviour, sharing bias
- **Demographic**: Western Educated Industrialised Rich Democratic (WEIRD) (not always!)
- **Social**: 90%-9%-1%, hyperactive minorities of contributors
- **Geographic**: urban/rural, developed/developing, central/peripheral, human/natural

CGI strictures

- Without centralised planning and protocols, data **quality** remains uneven (coverage!)
- **Wikipedia** replaced **Britannica**, but OpenStreetMap is not replacing Google Maps
- CGI **cannot** replace established data collection protocols and sources, but can provide useful **ancillary data**

CGI strictures

- Without centralised planning and protocols, data **quality** remains uneven (coverage!)
- **Wikipedia** replaced **Britannica**, but OpenStreetMap is not replacing Google Maps
- CGI **cannot** replace established data collection protocols and sources, but can provide useful **ancillary data**

Reinventing wheels

- Some CGI **replicates** work that has been done better by professionals
- More useful to focus on "missing" data:

Humanitarian OpenStreetMap Team

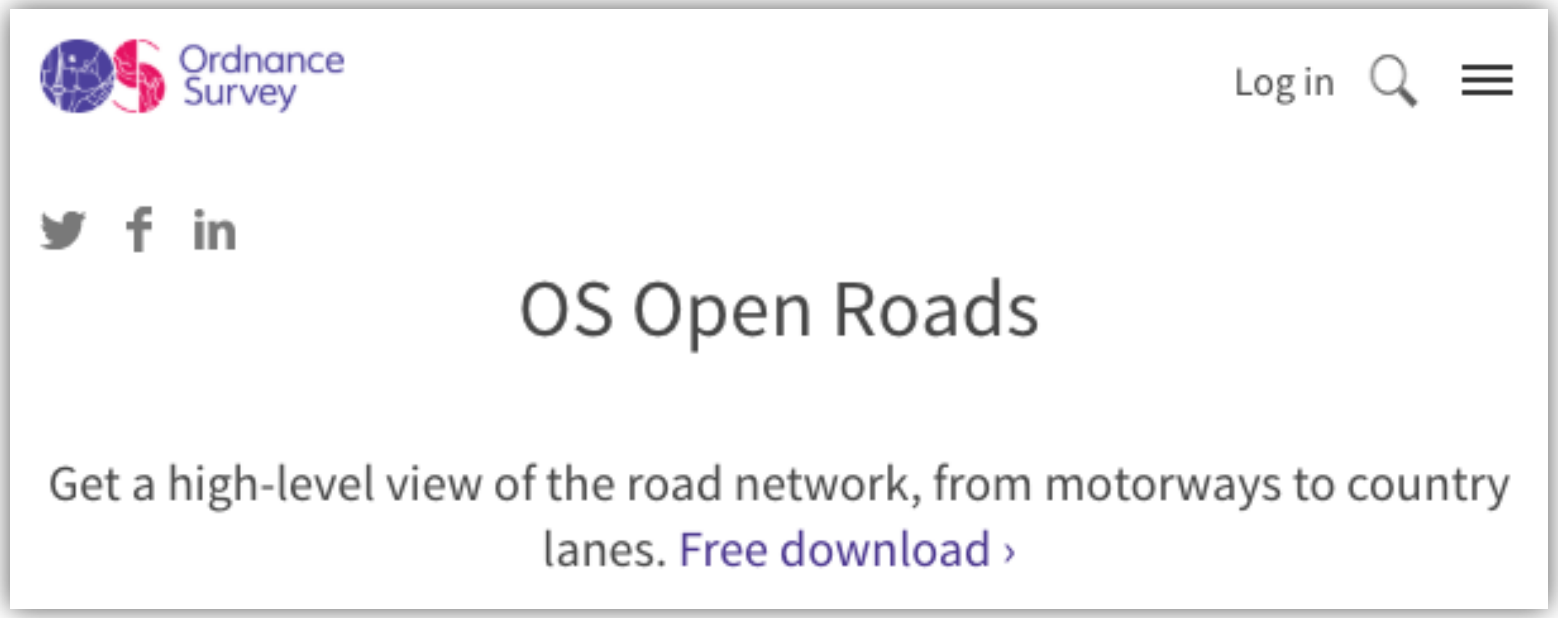
MISSING MAPS

bookscrounger.com

Open data vs CGI

Authoritative datasets are becoming cheaper/free

DATA.GOV.UK Beta
Opening up Government

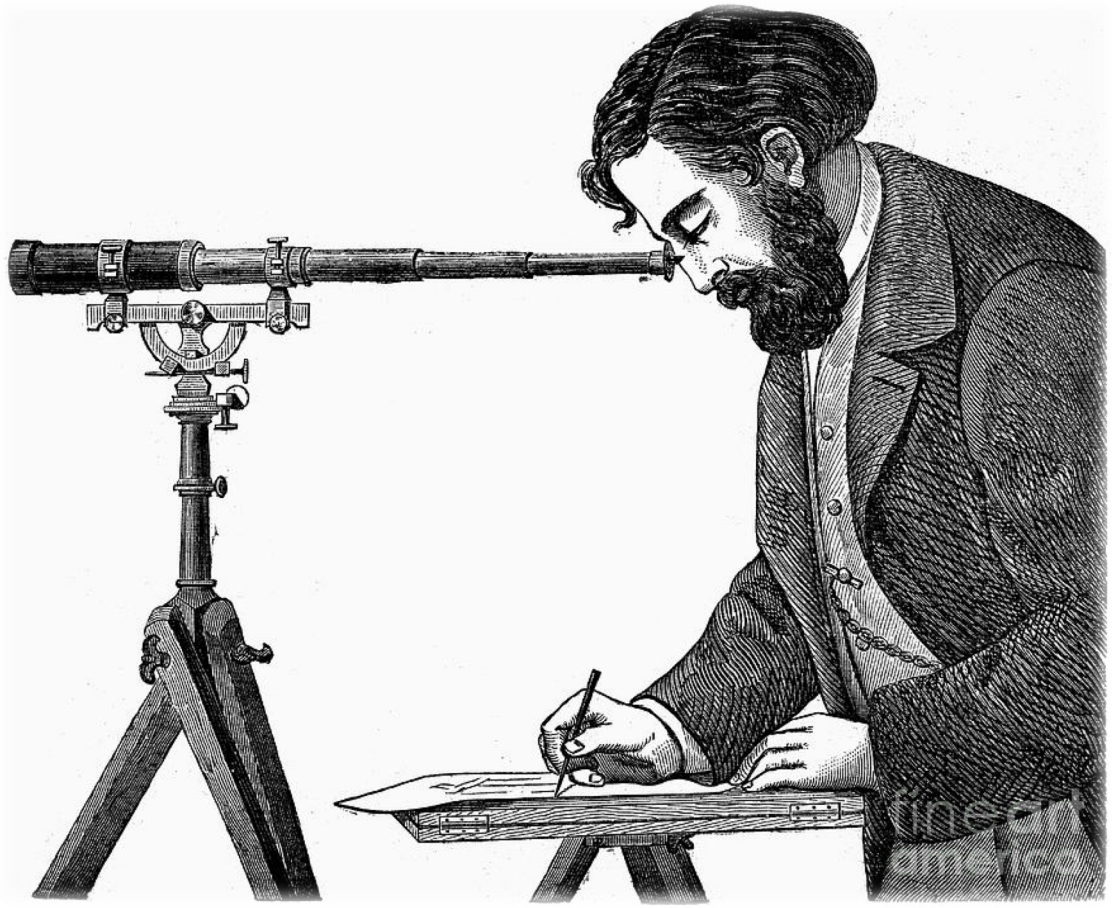


Broadening our horizons

The usual suspects

Most studies on **OSM**, **Wikipedia**, **Twitter**, **Flickr**.

There's more out there!

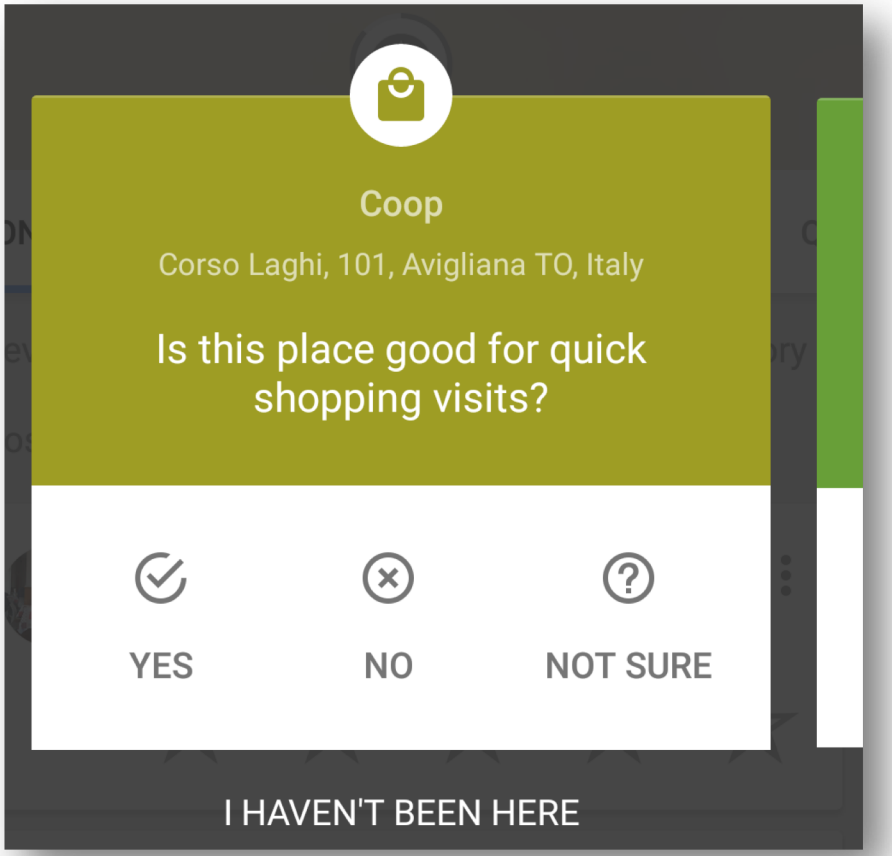


https://fineartamerica.com


Commonplace
geograph
photograph every grid square!
GeoNames
MISSING MAPS
wikimapia
foursquare
craigslist
yelp
airbnb
DBpedia
Insta
reddit
tripadvisor
WIKIDATA
YouTube
Spotify

Local Guides

- Google Maps
- From 5M to 50M contributors in 2017
- 700K new places monthly
- Gamification



- Hundreds of millions of users, billions of reviews
- Measurable effects on spatial and economic behaviour
- Sentiment about points of interest, cities, and neighbourhoods

PREMISE

- Micro-economic data (e.g. price of onions in India, new shops in Ghana)
- For profit, contributors are paid
- Applications: International Development, Government, Global Security, and Business

Case studies

crowdsourcing

examples of crowdfunding and crowdsourcing **include**
appen crowdsourcing
define crowdsourcing
benefits of crowdsourcing
amazon crowdsourcing
types of crowdsourcing
advantages of crowdsourcing
lego crowdsourcing
logo crowdsourcing
jeff howe crowdsourcing

Google Search I'm Feeling Lucky

Online visibility of CGI projects

- **Search engines** are the key entry point to discover new information
- Feedback loop between Wikipedia and Google Search to attract **new contributors**
- Making CGI **findable** and **consumable** for search engines and social media
- Study on **CGI** on **Google Search** (2018)

Online visibility of CGI projects

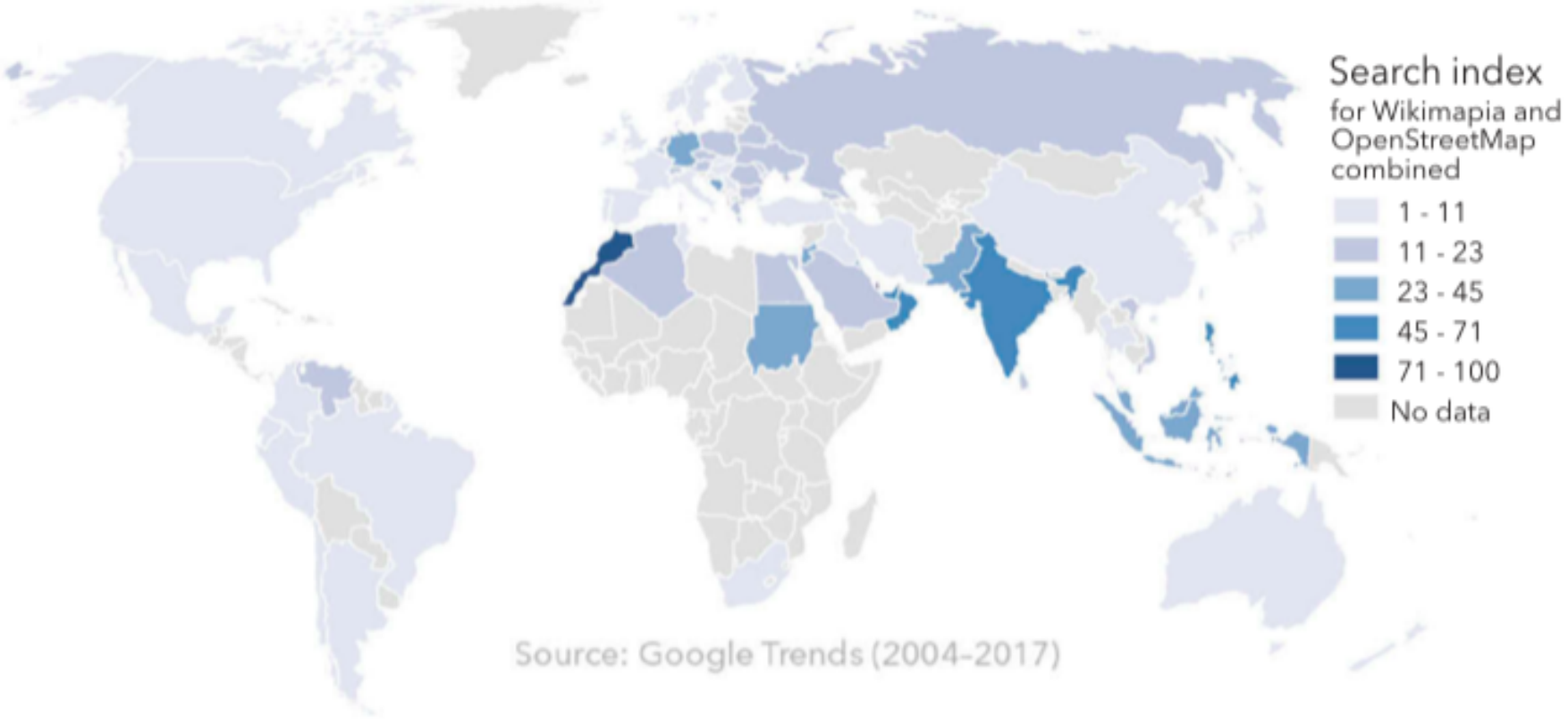
- **Search engines** are the key entry point to discover new information
- Feedback loop between Wikipedia and Google Search to attract **new contributors**
- Making CGI **findable** and **consumable** for search engines and social media
- Study on **CGI** on **Google Search** (2018)

Interest in CGI projects

Figure 4: Google searches for Wikimapia and OpenStreetMap on a monthly basis (source: Google Trends worldwide from 2004 to 2017, accessed on 15 April 2017).

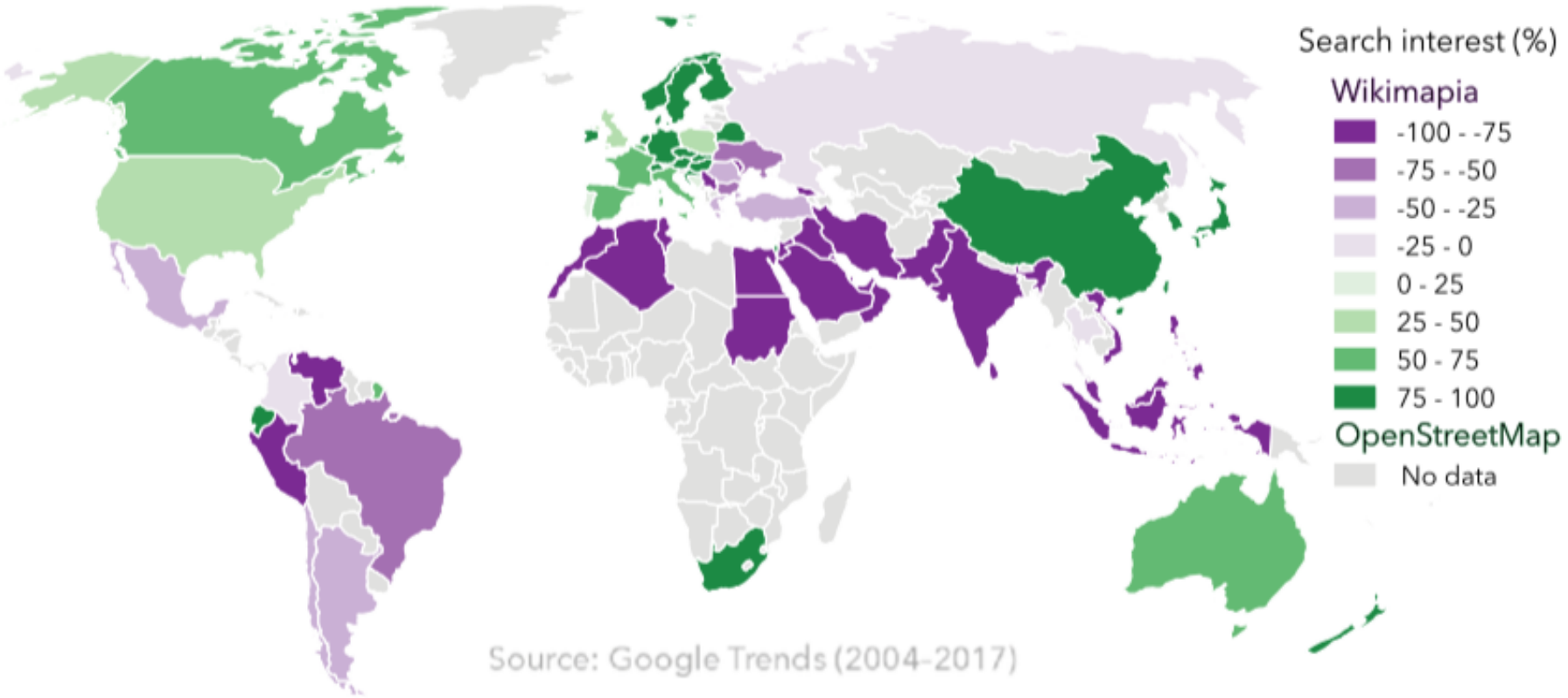
(Ballatore & Jokar Arsanjani, 2018)

Search Interest in Wikimapia/OSM combined

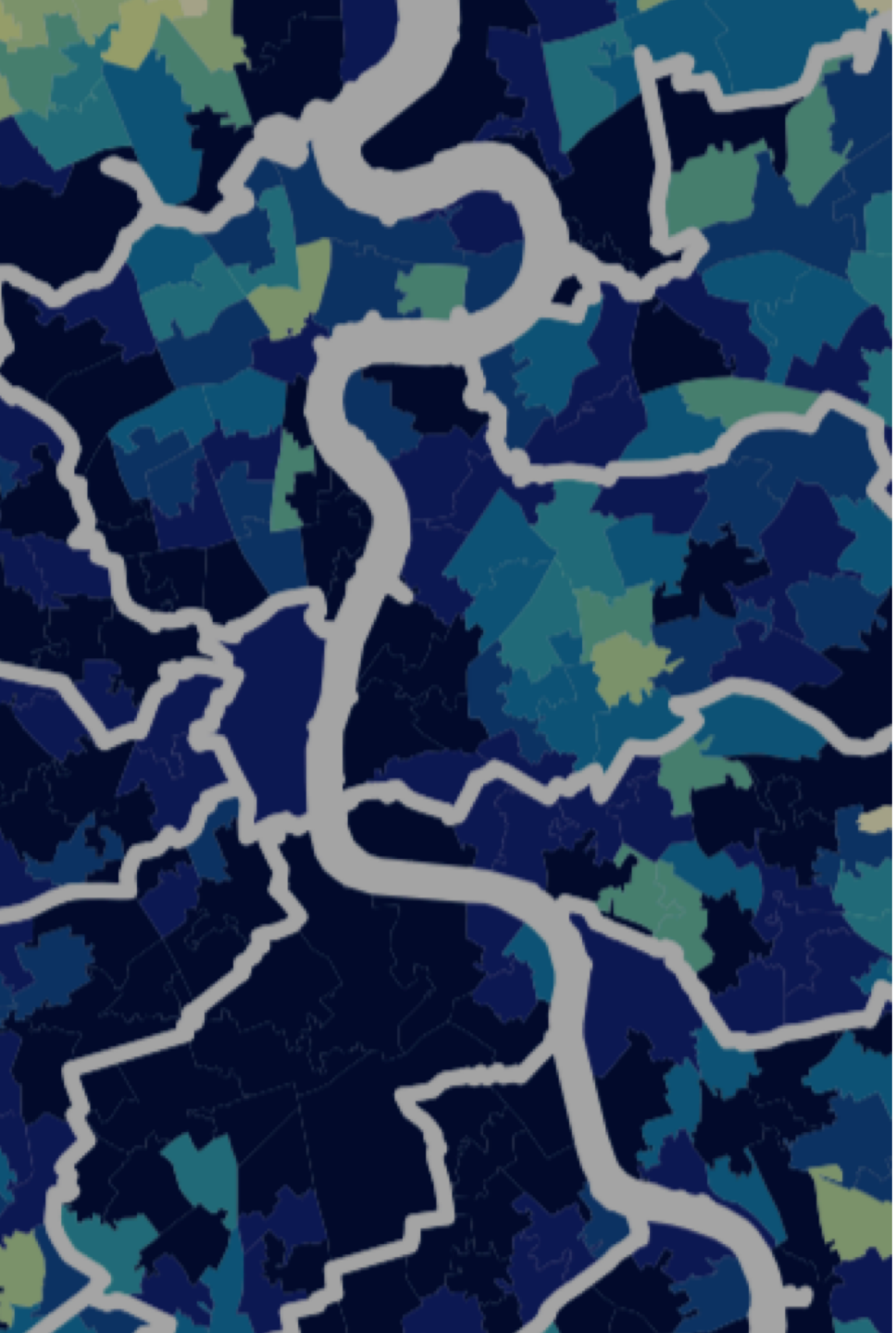


(Ballatore & Jokar Arsanjani, 2018)

Search Interest in Wikimapia vs OSM



(Ballatore & Jokar Arsanjani, 2018)



Thanks!

a.ballatore@bbk.ac.uk

aballatore.space

@a_ballatore

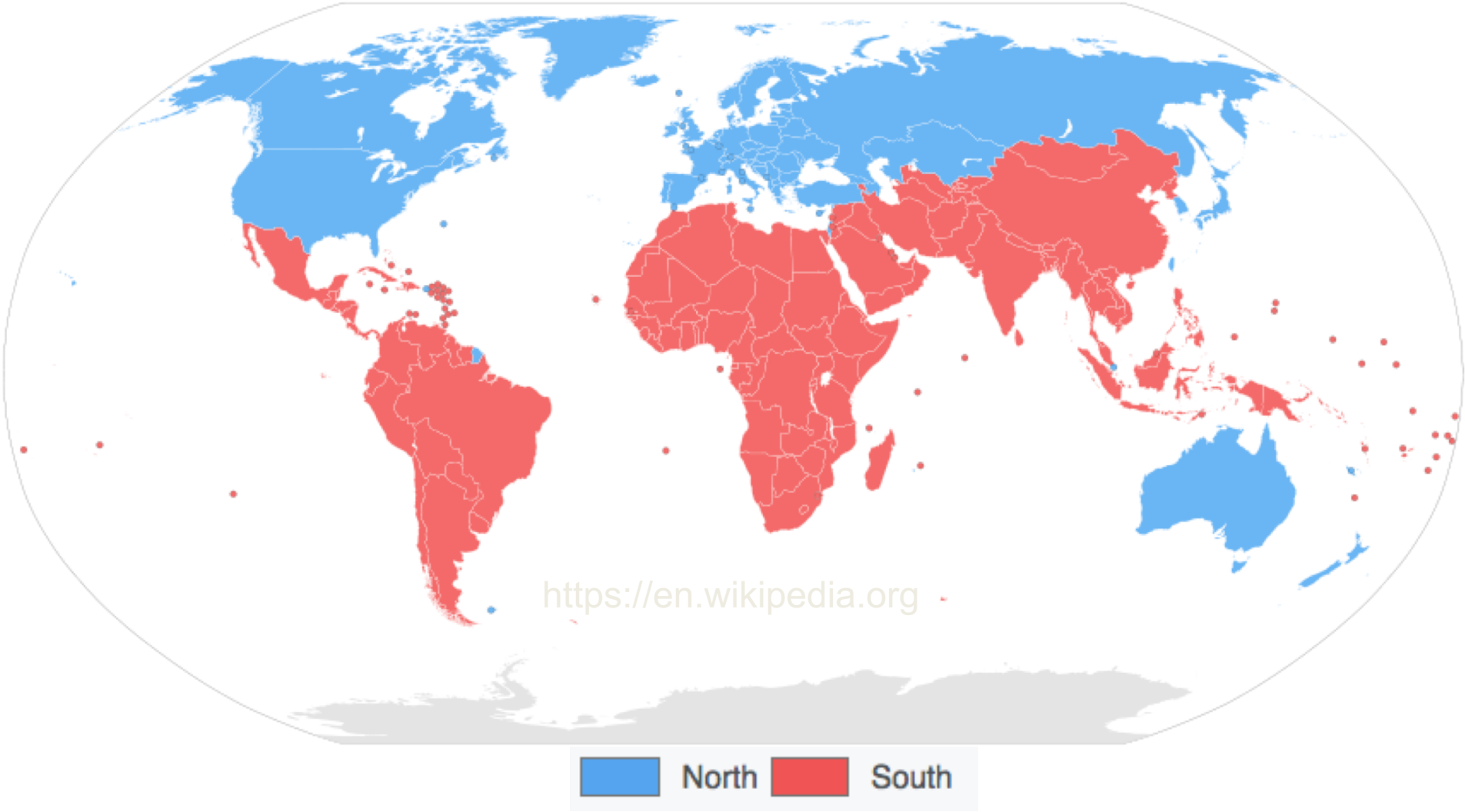
Birkbeck
UNIVERSITY OF LONDON

Diversity in CGI

- **Theme**: every project has its own scope
- **Demography**: specific populations, non-representative
- **Geography**: project activities are concentrated according to diverse geographies

(Ballatore & De Sabbata, 2018)

North/South divide



Genius loci (the spirit of place)

Bacchus and Serpent Agathodaimon, Pompeii (15 AD ca)

craigslist

- 50M active users, 20B views per month
- Broad topics (housing, jobs, sales)
- City-specific lists

community

activities
artists
car pools
childcare
classes
events
general
groups
local news
lost+found
missed connections
musicians
pets
politics
rants & raves
volunteers

services

beauty
cars/automotive
cell/mobile
computer
creative
cycle
event
farm+garden
financial
household
housing/real estate
labour/move
legal
lessons
marine
pet
skilled trade
sm biz ads
travel/hol

housing

flats/housing
holiday rentals
housing/real estate for sale
housing swap
housing wanted
office / commercial
parking / storage
rooms/shared
rooms wanted
sub-lets / temporary

for sale

antiques
appliances
arts+crafts
atv/utv/sno
auto parts
aviation
baby+kid
barter
beauty+hlth
farm+garden
free
furniture
garage sale
general
heavy equip
household
jewellery
materials