DIGITAL LIBRARY SERVICES

Open Geospatial Data - Compound Interests

UCT Open Data Day 2020

Friday 6th March, 2020

Nicholas Lindenberg

GIS Specialist, ORCID: 0000-0003-0783-0966

The GIS Lab | Digital Library Services, UCT Libraries

University of Cape Town

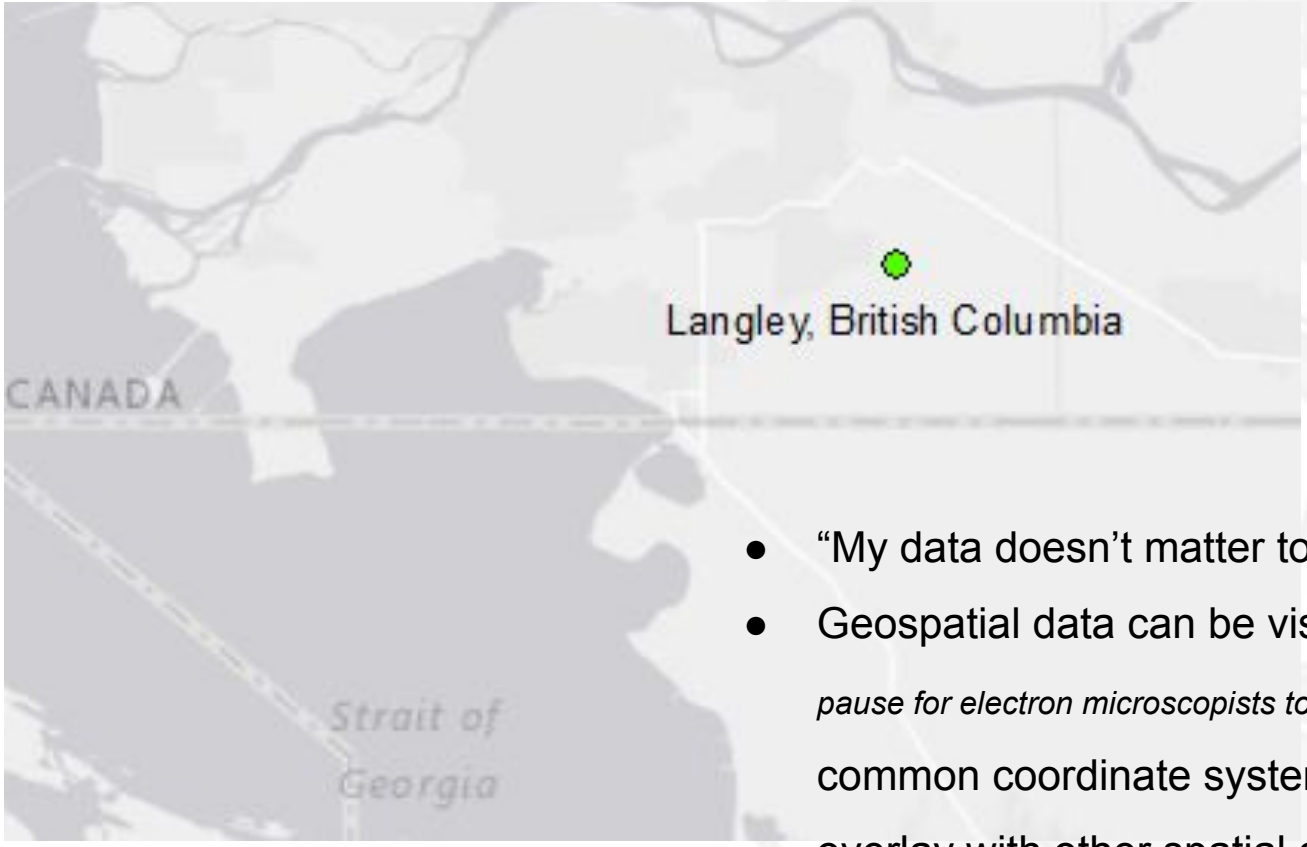
DOI: 10.25375/uct.11946225

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Lindenberg, Nicholas (2020): Open Geospatial Data - Compound Interests.

1. Insignificant Figures
2. Particulates Matter
3. Sift Data
4. Bank Leafs
5. Ethical Dots
6. Convenience Costs
7. Thank You For Your Service

Insignificant Figures



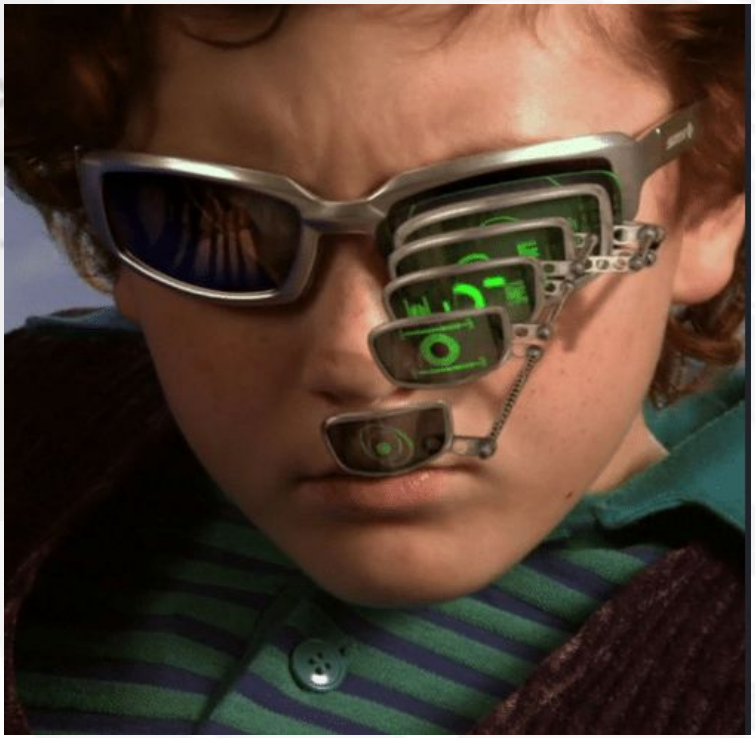
- “My data doesn’t matter to anyone else.”
- Geospatial data can be visibly insignificant - *pause for electron microscopists to calm down* - due to common coordinate systems allowing easy overlay with other spatial data

Insignificant Figures

Dispersal, home-range size, and habitat use of an endangered land snail, the Oregon forestsnail (*Allogona townsendiana*)

"To investigate the dispersal distances and habitat-use patterns of Oregon forest snails, we tagged and tracked 21 adult snails at Langley, British Columbia, for up to 3 years (2005-2008). The maximum daily dispersal distance for a snail was 4.5 m and the maximum displacement that we observed for a snail was 32.2 m during 3 years. Snails occupied home-range areas of 18.4-404.4 m2, often overlapping both forest and meadow habitat."

Edworthy, Amanda & Steensma, Karen & Zandberg, H & Lilley, Patrick. (2012). Dispersal, home-range size, and habitat use of an endangered land snail, the Oregon forestsnail (Allogona townsendiana). Canadian Journal of Zoology. 90. 875-884. 10.1139/Z2012-056

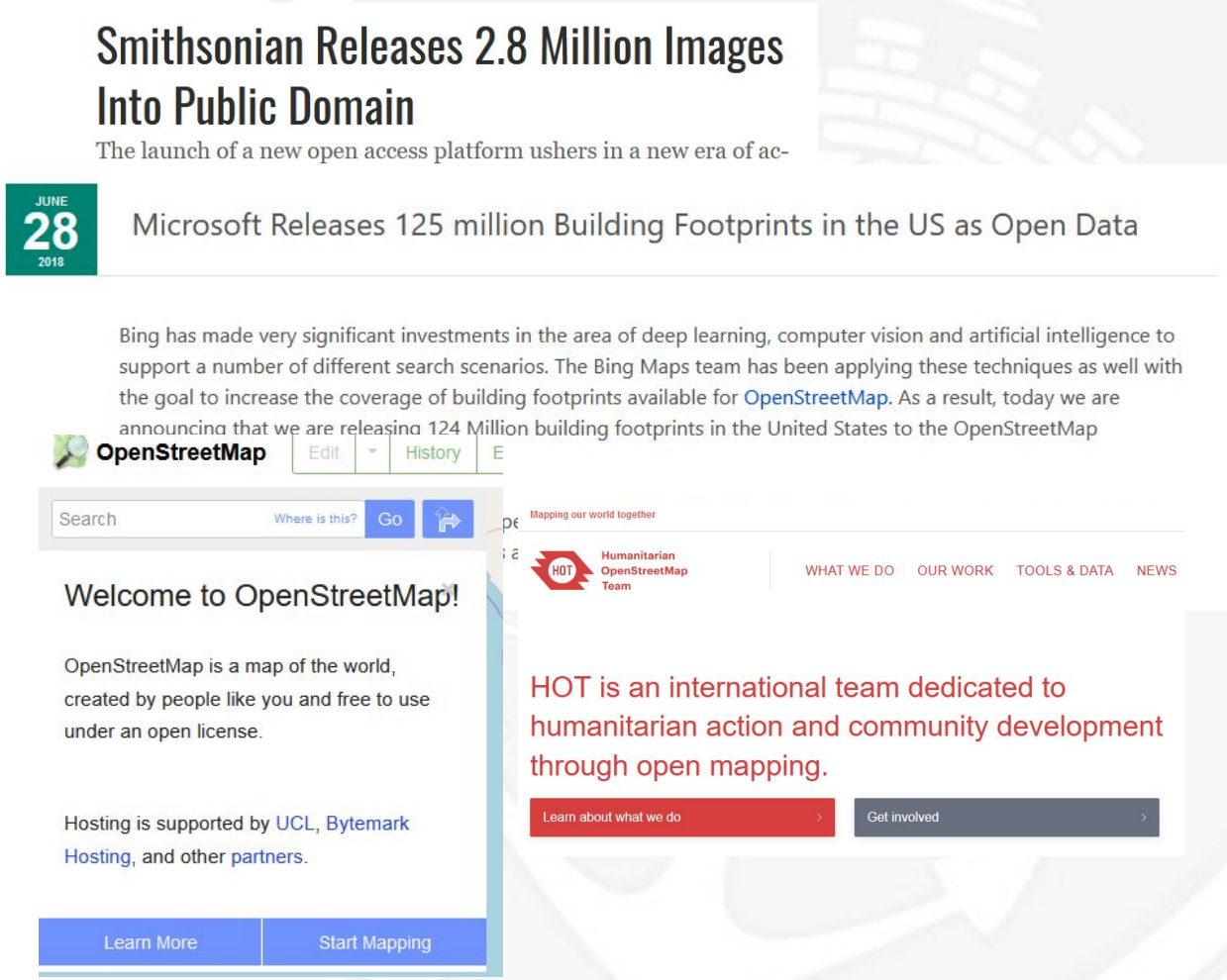


Particulates Matter

- “Only volume and scale count!”
- Your data is the medium, not the container
  - Lego blocks,
  - not Lego kits

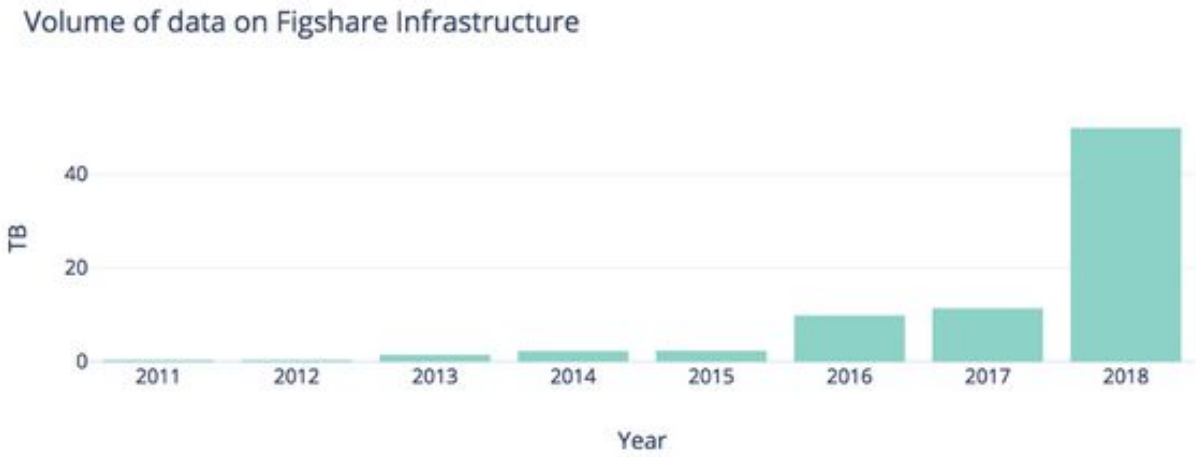
Particulates Matter

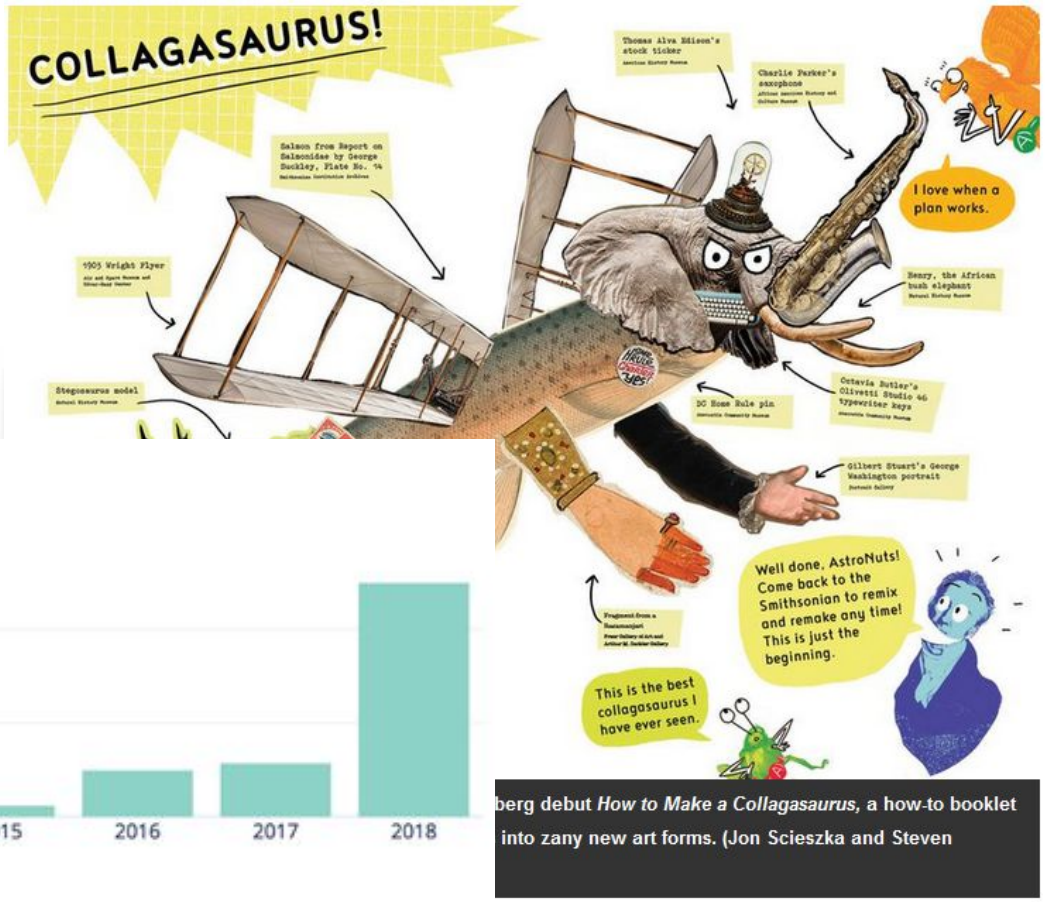
- You don't need to be a corporation or join a project and produce data in bulk for it to be of value.
- *Pause for astrophysicists, climatologists and bio-geneticists to calm down.*



Particulates Matter

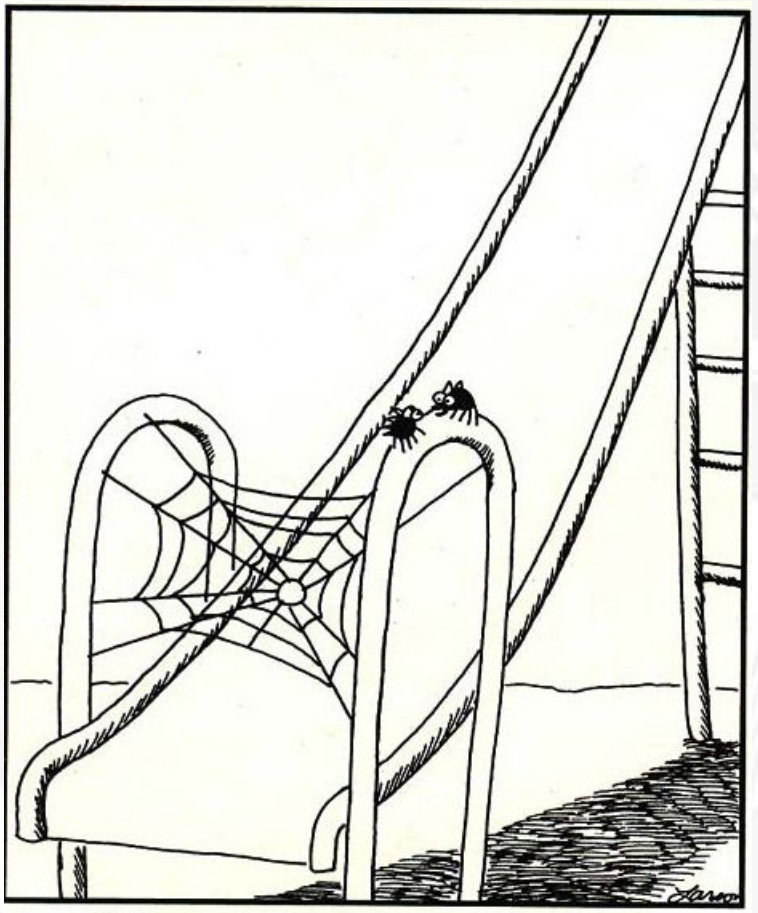
- Everything counts (in large amounts)
- FAIR data accumulates faster than you think





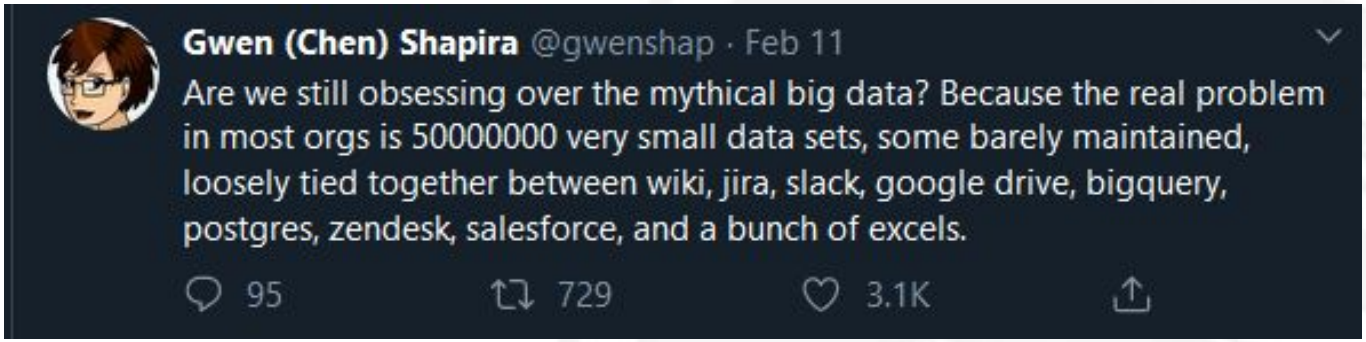
Sift Data

- Spatial data accumulates even faster than non-spatial data due to the locational component - *everything happens somewhere.*
- This allows serendipitous discovery without subject specific search relevance
- Locations allow overlay and amalgamation at many scales


"If we pull this off, we'll eat like kings."

"*GIS-enabled applications are thus especially valuable for almost all GEO data, although they are not discussed separately here*" The Value of Open Data Sharing GEO Report, 2015

Bank Leafs

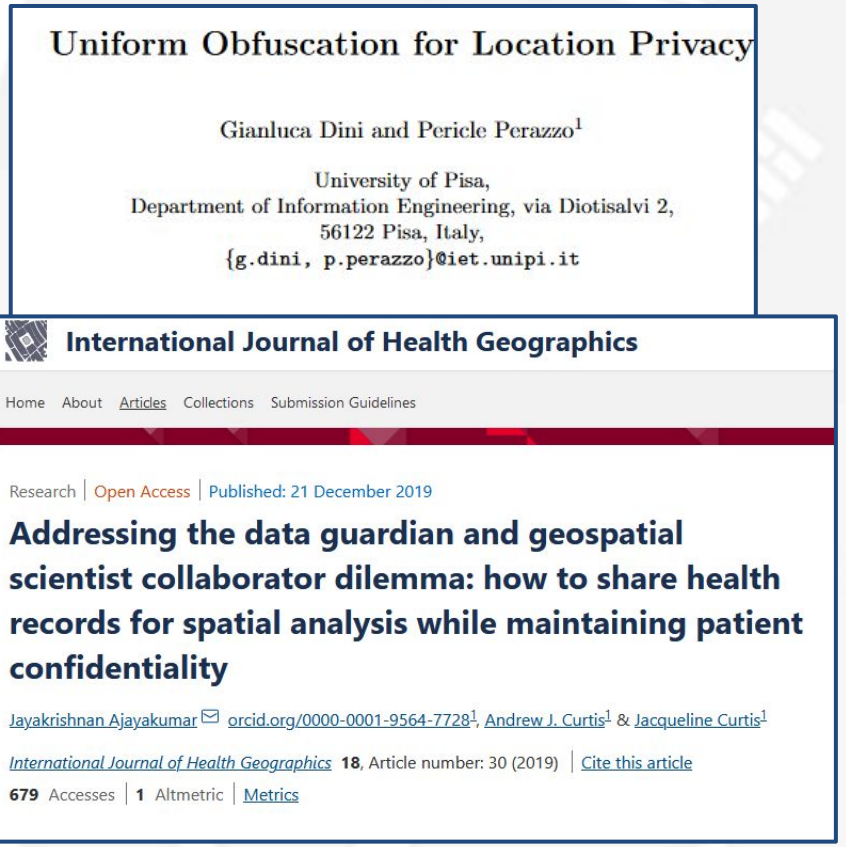


- Fragile spider webs of data exist within and between individuals and organisations
- FAIR policies and processes allow accretion of data beyond subject specific boundaries
- Disconnecting the concept of research outputs as strictly final publications vastly increases accessibility and survivability of data

"However, long-tail models only work when there is sufficient content to occupy the tail. In order to achieve this scale of content in a sustainable manner, the outputs listed above need to become a frictionless by-product of the standard practice, rather than the outcomes of isolated projects."
Weller, M; 2011: The Digital Scholar

Ethical Dots

- Locational data may be too sensitive to share directly
- POPIA! GDPR! Ethics! Oh My!
- Obfuscation:
  - Nearest Place Assignment
  - Rounding Coordinates
  - Clustering techniques
- Use platforms with tiered access levels



Uniform Obfuscation for Location Privacy

Gianluca Dini and Pericle Perazzo[1]

University of Pisa,
Department of Information Engineering, via Diotisalvi 2,
56122 Pisa, Italy,
{g.dini, p.perazzo}@iet.unipi.it

International Journal of Health Geographics

Home About Articles Collections Submission Guidelines

Research | Open Access | Published: 21 December 2019

Addressing the data guardian and geospatial scientist collaborator dilemma: how to share health records for spatial analysis while maintaining patient confidentiality

Jayakrishnan Ajayakumar orcid.org/0000-0001-9564-7728[1], Andrew J. Curtis[1] & Jacqueline Curtis[1]

International Journal of Health Geographics **18**, Article number: 30 (2019) | Cite this article

679 Accesses | **1** Altmetric | Metrics

Ethical Dots

- Data resists anonymisation

"Using our model, we find that 99.98% of Americans would be correctly re-identified in any dataset using 15 demographic attributes."

Rocher, L., Hendrickx, J.M. & de Montjoye, Y. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun* **10,** 3069 (2019). https://doi.org/10.1038/s41467-019-10933-3

Article | Open Access | Published: 23 July 2019

Estimating the success of re-identifications in incomplete datasets using generative models

Luc Rocher orcid.org/0000-0002-9956-1187[1,2,3], Julien M. Hendrickx[1] & Yves-Alexandre de Montjoye[2,3]

Nature Communications **10**, Article number: 3069 (2019) | Cite this article

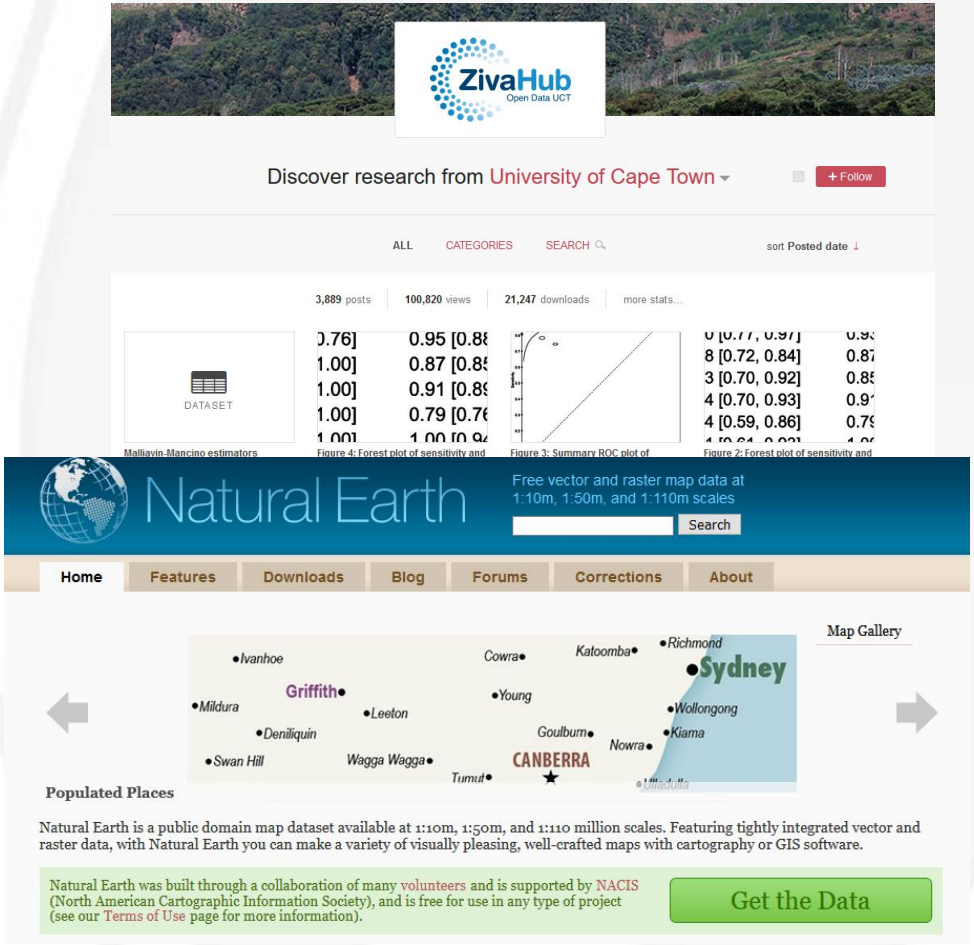
97k Accesses | **15** Citations | **2075** Altmetric | Metrics

Subjects

Computational science **Social sciences**

Convenience Costs

- FAIR != Open Data, T&C's apply and add the cost of administering and policing TOU's
- Convenience Counts - Ease of access, search and sharing are strong attractors.
- Clean Open Data is enormously valuable
- Natural Earth & WorldCLIM underpin many basic maps by UCT researchers.



Thank You For Your Service

- Thank you for sharing your data and gifting that effort to others.
- Open Data is made by (and of) People.
- Including you!

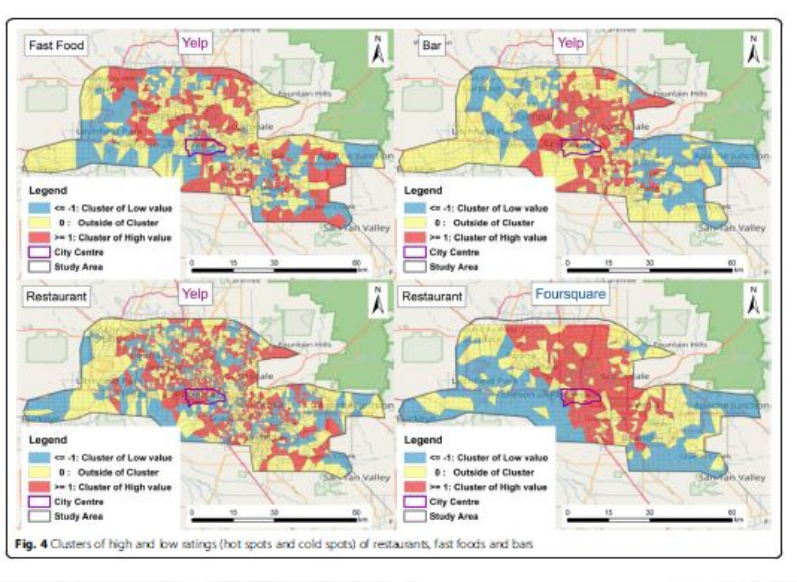


Fig. 4 Clusters of high and low ratings (hot spots and cold spots) of restaurants, fast foods and bars

Research article | Open Access | Published: 20 October 2018

Harvesting the wisdom of the crowd: using online ratings to explore care experiences in regions

Roy J P Hendrikx orcid.org/0000-0003-3088-6407[1,4], Marieke D Spreeuwenberg[2,3], Hanneke W Drewes[4], Jeroen N Struijs[4,5], Dirk Ruwaard[3] & Caroline A Baan[1,4]

BMC Health Services Research **18**, Article number: 801 (2018) | Cite this article

595 Accesses | **2** Citations | Metrics

Spatial analysis of users-generated ratings of yelp venues

Article (PDF Available) · February 2017 *with* 171 Reads

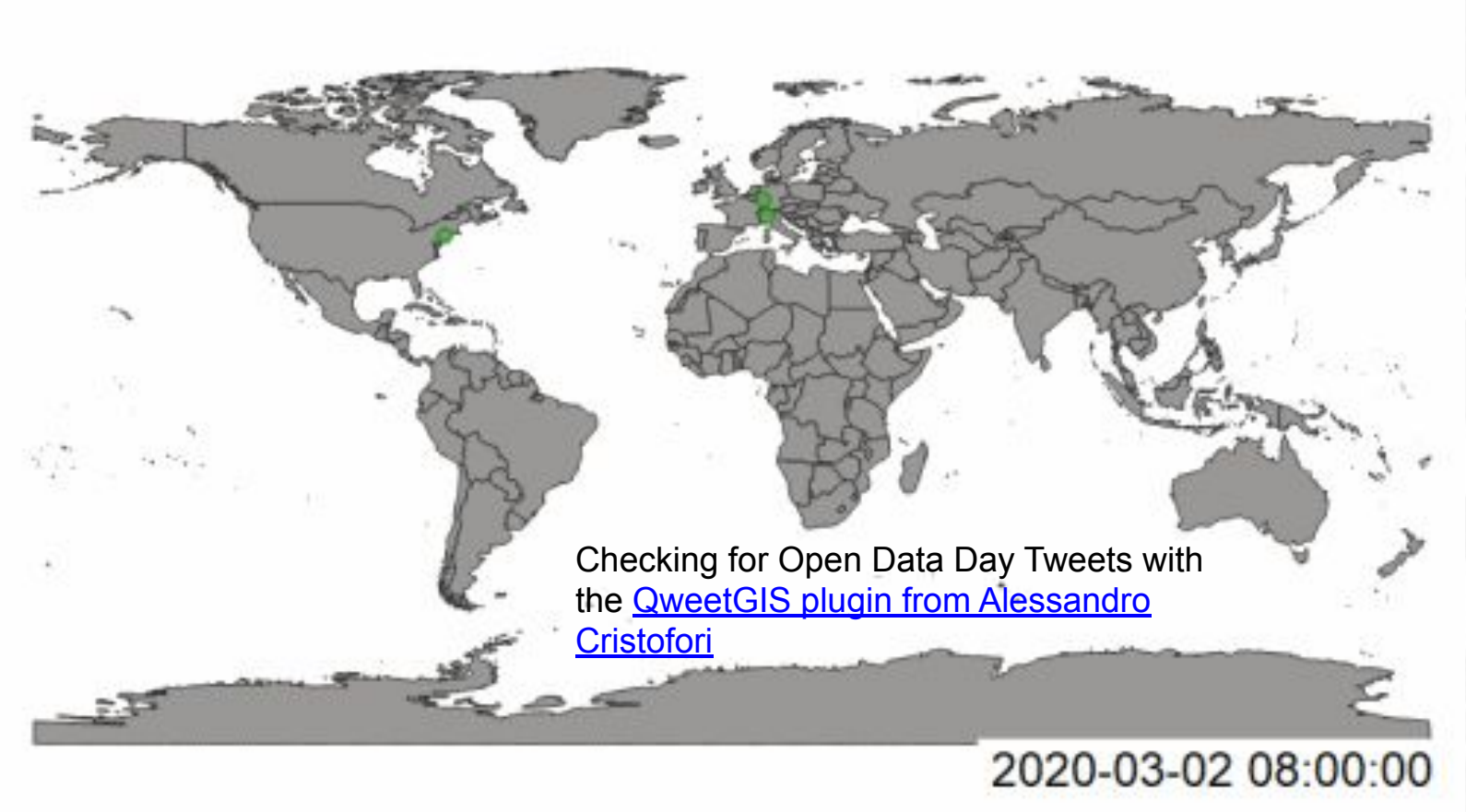
DOI: 10.1186/s40965-017-0020-9

Cite this publication

Yeran Sun
18.71 · University of Glasgow

Jorge David Gonzalez Paule
4.29 · University of Glasgow

Thank You For Your Service



Thank You For Your Service

- For Now ...

Fig 1. Important Computer Vision tasks applied to GIS

Figure 3. Upload AI generated roads through OSM iD tool, modified to highlight generated roads in green

"*AI technologies have the potential to foster an inclusive science community. But a good AI is dependent on the variety and quality of data. Open data can play a key part for AI algorithms and machines to function and produce good outcomes*[27]."

Burgelman Jean-Claude, Pascu Corina, Szkuta Katarzyna, Von Schomberg Rene, Karalopoulos Athanasios, Repanas Konstantinos, Schouppe Michel, 2019: Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century; DOI: 10.3389/fdata.2019.00043

DIGITAL LIBRARY SERVICES

Thank You