

Connection of curation and metrics in telling the stories of unique data sets: two specific case studies

IASSIST & CARTO 2018, Concordia University, in Montréal, Québec



Lily Troia, MLIS
Engagement Manager | Altmetric
@lilytroia

Dan Valen
Product Specialist | Figshare
@dnvln



Today's Agenda

- Data sharing and social justice: a tale of two geo-spatial data sets
- Study #1: getting outside the echo chamber: a case study examining the unique impact of sharing supplementary data
- Study #2: the intersection of open data, altmetrics, reproducibility
- Connecting data curation and metrics to societal impact
- Conclusion

Megaregions Data for GIS

This Figshare dataset contains the files created by and used in a related PLOS ONE paper, entitled 'An economic geography of the United States: from commutes to megaregions', by Garrett Dash Nelson and Alasdair Rae, published 30 November 2016.

- PLOS Data Policy
- Sheffield + Figshare
- Alasdair Case Study

The screenshot shows the Figshare dataset page for 'United States Commutes and Megaregions data for GIS'. The page header includes the Figshare logo, user 'Dan Valen', and search bar. The dataset is categorized under 'TEXT' and 'ARCHIVE'. It lists several files: 'readme - updated... .txt (4.17 KB)', 'commutes.net (83.83 MB)', 'us_tlw_v3_US_on... .zip (150.95 MB)', 'lower_48.png (6.47 MB)', 'map_1.png (10.16 MB)', 'map_2.png (8.5 MB)', 'map_5.png (8.5 MB)', and 'map_6.png (11.27 MB)'. The page includes a 'Cite' button, 'Download all (888.78 MB)', 'Share', 'Embed', and '+ Collect' options. The dataset description states it contains files created by and used in a PLOS ONE paper by Garrett Dash Nelson and Alasdair Rae, published 30 November 2016. The update date is 27 January 2017. The page also features a 'KEYWORD(S)' section with tags like 'megaregion', 'commutes', 'commuting', 'geography', 'algorithm', 'network', 'partitioning', 'united states', 'combo', 'transit', 'region', and 'journey to work'. A 'LICENCE' section indicates 'CC BY 4.0'. A red box highlights the University of Sheffield logo and the text 'The University of Sheffield'.

United States Commutes and Megaregions data for GIS

Version 5 31.01.2017, 05:01 by Alasdair Rae, Garrett G.D. Nelson

This Figshare dataset contains the files created by and used in a related PLOS ONE paper, entitled 'An economic geography of the United States: from commutes to megaregions', by Garrett Dash Nelson and Alasdair Rae, published 30 November 2016.

Update: 27 January 2017 - see item 7. below

In addition to the files listed below, we have also provided a series of maps here, as high resolution PNGs. The fifth file below can be styled in QGIS using the QML style file provided in number 6.

Information on files

1. A us_tlw_v3_US_only_epsg5070v2 zipped shapefile (for use in GIS software such as ArcGIS or QGIS) containing individual census tract to census tract commuter links for the entire United States, based on the 2006-2010 American Community Survey dataset, as cited in the paper.
2. The commutes.net file is a Pajek format file used in the derivation of our megaregions within Combo, the partitioning algorithm we used in the paper. Pajek (Program for Large Network Analysis) is a free software package. Combo is an open source modularity optimisation program developed by researchers at MIT, and the source code can be accessed at: http://senseable.mit.edu/community_detection/

KEYWORD(S)

megaregion, commutes, commuting, geography, algorithm, network, partitioning, united states, combo, transit, region, journey to work

LICENCE

CC BY 4.0

EXPORT

The impact of sharing supplementary datasets and speaking outside the echo chamber

Alasdair Rae, The University of Sheffield



- Any additional information you can share is valuable.
- Sharing your data encourages conversations outside the echo chamber.
- Sharing data now can save time: no need to field requests



Compliance, altruism, and reuse

Information associated with 'United States Commutes and Megaregions data for GIS' project on Figshare

THIS README FILE WAS LAST UPDATED ON 20 DECEMBER 2016, BY ALASDAIR RAE

This Figshare dataset contains the files used in a related PLOS ONE paper, entitled 'An economic geography of the United States: from commutes to megaregions', by Garrett Dash Nelson and Alasdair Rae, published 30 November 2016. Link: <http://dx.doi.org/10.1371/journal.pone.0166083>

If you are looking to replicate the maps with the glowing lines and the dark background, please see 5 and 6, below.

Compliance (from PLOS,
Sheffield, and the UK):

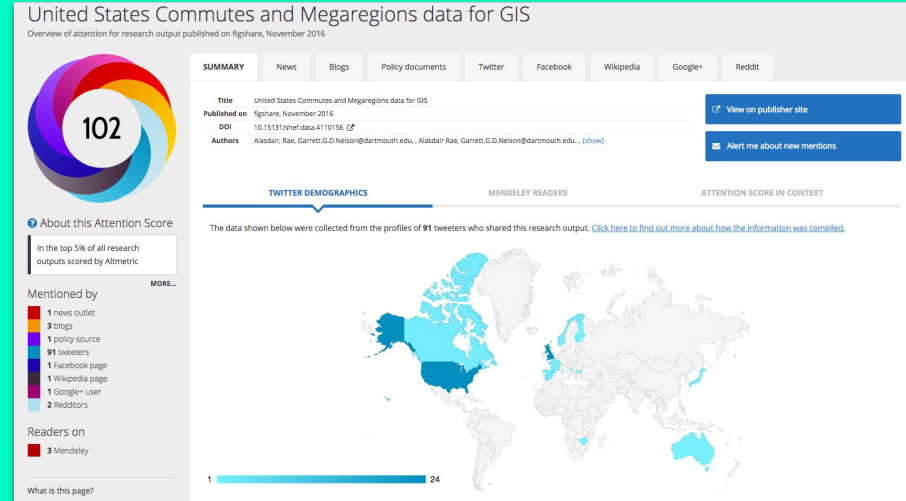
Data Availability: The data are available at the following Figshare link:
<https://doi.org/10.15131/shef.data.4110156>. Link to all Figs in high resolution:
<https://www.dropbox.com/sh/tmqry68m80tg8dz/AAAZBaEjxYWcZniBeZnlbWMLa?dl=0>.

Altruism (from Alasdair):

“ In the social science world,
we don't often think of replicability
and the process of replication.
But it's good to make sure your
data is seen and used.”

Using Altmetrics to Uncover and Analyze GIS Data audiences and communication pathways

- Global and local reach
- Unique audiences = unique impact
- Both “public” and scholarly attention
- Rich info around engagement despite no citation data





Geo Inquiets

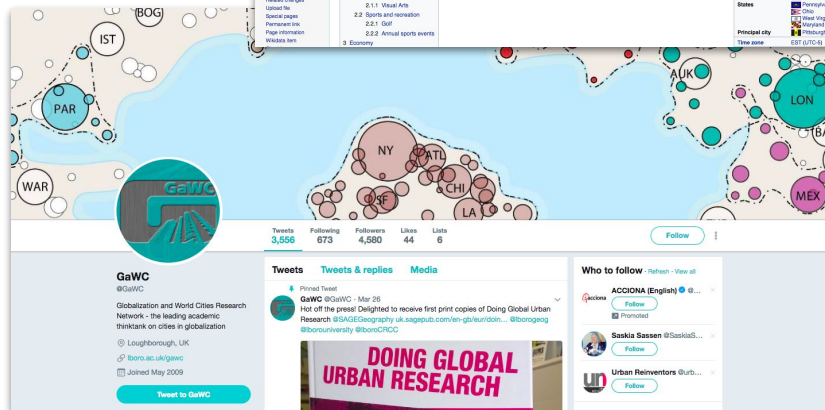
@geoinquiets

A Catalunya som unes quantes ànimes inquietes amb ària per la informació geogràfica i les noves tecnologies.

Barcelona

geoinquiets.cat

Joined February 2011



AIMS Community
@AIMS_Community

Latest news and posts on management and exchange of agricultural research information and knowledge from the AIMS Community of Practice

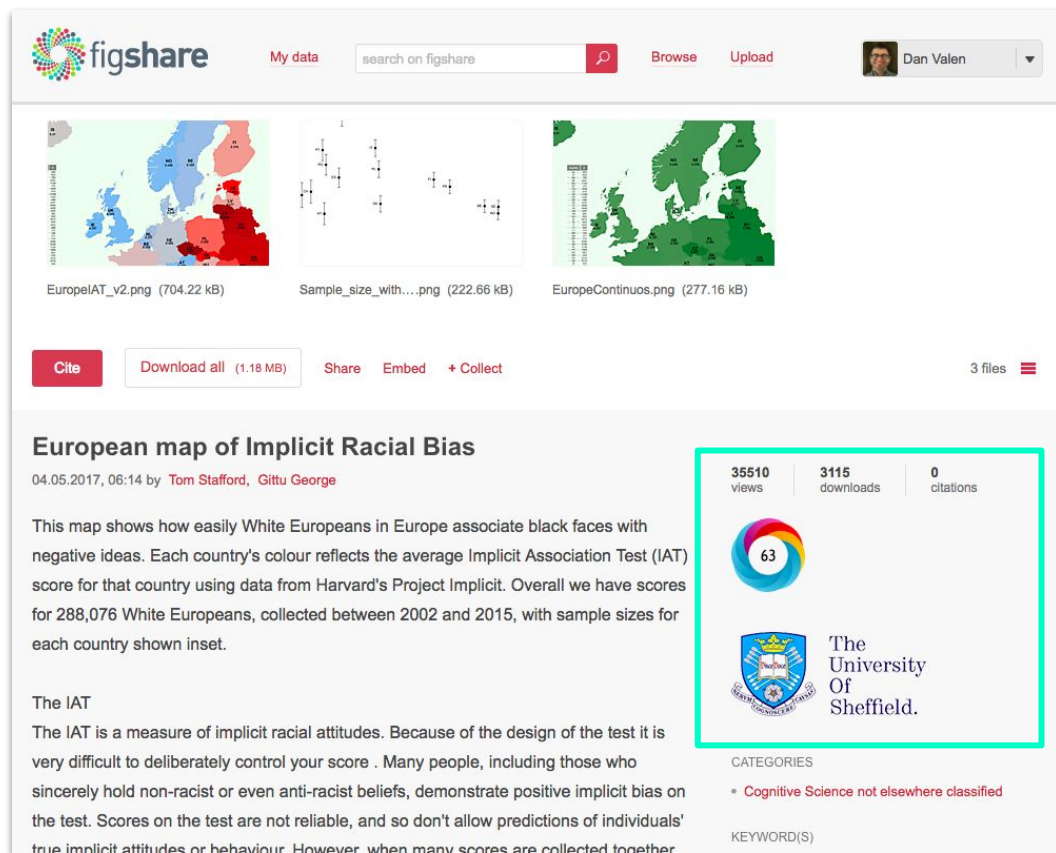
aims.fao.org

Joined August 2010

- Seen by an upward bound of nearly 200k Twitter followers
- Targeted attention from community groups, election reformers/activists
- Evidence of information extending beyond academia

European map of Implicit Racial Bias

This map shows how easily White Europeans in Europe associate black faces with negative ideas. Each country's colour reflects the average Implicit Association Test (IAT) score for that country using data from Harvard's Project Implicit.



<https://doi.org/10.15131/shef.data.4750588.v1>

Project Implicit, reproducibility, and data availability

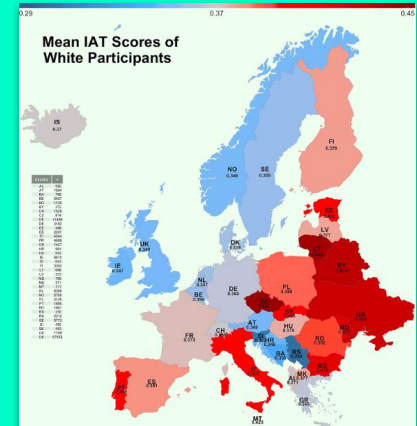
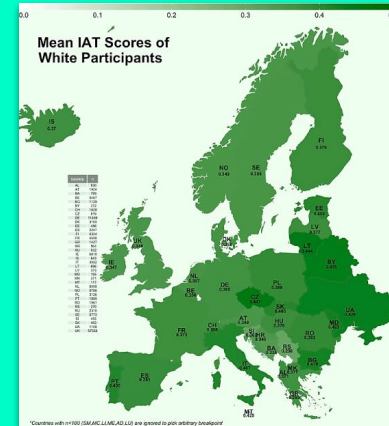
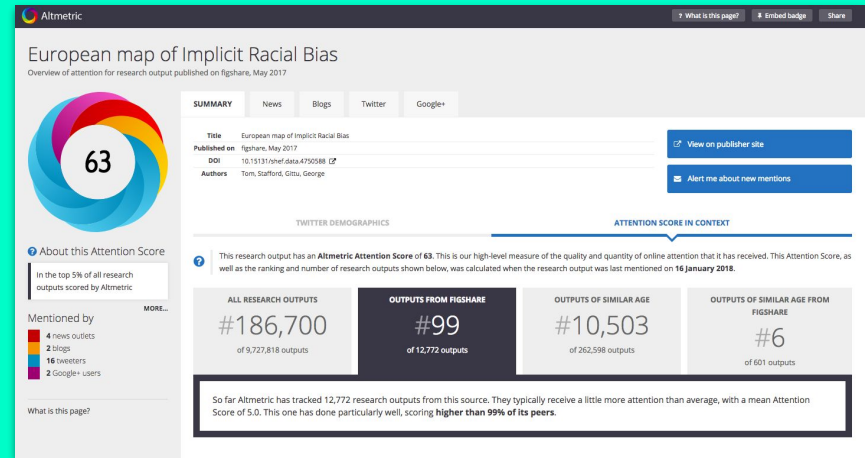
The screenshot shows the OSFHOME interface for the 'Race IAT 2002-2017' project. The header includes the OSFHOME logo, search, support, donate, sign up, and sign in buttons. The project title 'Race IAT 2002-2017' is prominently displayed, along with contributor names (Frank Kalyuan Xu, Nicole Lofaro, Brian A. Nosek, Anthony G. Greenwald) and affiliated institutions (University of Virginia, Center For Open Science). The page is divided into sections: Wiki, Files, and Components. The Wiki section contains a notice about data errors and a link to download updated files. The Files section shows a list of files with columns for Name and Modified. The Components section lists various datasets and codebooks, including 'Datasets & Codebooks', 'Datasets & Codebooks (Touch Screen)', 'Experiment Materials', and 'Skeleton Syntax for .sav Data'.

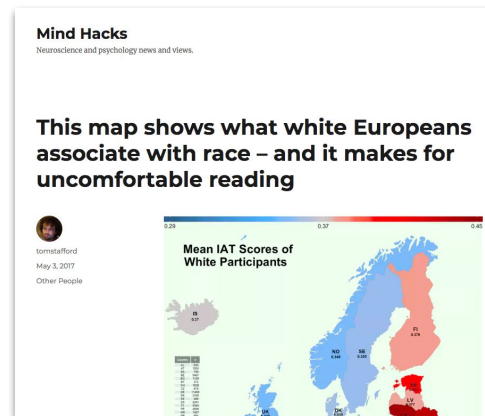
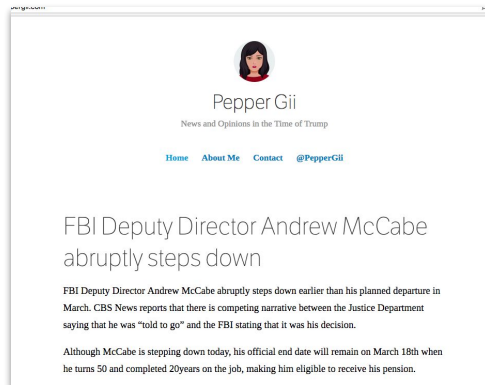
The screenshot shows the GitHub interface for the 'WorldBias' repository by 'george0'. The header includes the GitHub logo, search, pull requests, issues, marketplace, and explore buttons. The repository name 'WorldBias' is displayed, along with watch, star, and fork buttons. The repository statistics show 68 commits, 3 branches, 0 releases, and 2 contributors. The 'Code' tab is selected, showing a list of files and their commit history. The files listed include 'Data', 'Europe', 'WorldMapShapes', 'EuropeContinuos.png', 'IAT_by_ethnicity.md', 'IAT_by_ethnicity.png', 'IAT_by_ethnicity_whiteonly.png', 'MeanIATF2.png', 'README.md', and 'Relat-logs'. The commit history for each file is shown, with the latest commit being '528f6e2' on Jul 5, 2017.

The data is controversial... but it's also built on earlier studies posted at the OSF/Harvard (raw data), code on Github, data visualizations on Figshare, article at? Oh, there is no article :)

Connecting data to those on the ground

- How can academia reach the broader public? Activists?
- How can altmetrics inform communication and outreach efforts?
- How do we define/address “misuse” or distortions?





- 88% of Twitter shares from “members of the public.”
- Interesting global and non-English attention reformers/activists
- Impact of mainstream news attention and institutional blogging

Conclusions (and questions): Recognizing value of datasets in scholarly communication as indicated through Altmetric attention



Lily Troia, MLIS
Engagement Manager | Altmetric
@lilytroia

Dan Valen
Product Specialist | Figshare
@dnvln

