Morgan Freeman is dead and other big data stories

Forthcoming viewpoint in cultural geographies

Matthew W. Wilson, PhD
Graduate School of Design and Center for Geographic Analysis, Harvard University
Department of Geography, University of Kentucky

Contact:
Graduate School of Design
48 Quincy Street
Cambridge, MA 02138
(mobile) 1.206.817.4762
matthew_wilson@harvard.edu

Abstract:
In the context of the feverish pace in which the social sciences are grappling with the implications for a turn toward 'big data', I suggest a different starting point: that big data are not necessarily social science data. In this somewhat speculative provocation, I argue that we should lean more on the notion that social media are phenomena and less on the notion that social media are evidence of phenomena. In doing so, I sketch four areas of potential criticality for an emerging big data studies.

> It is not enough to simply ask, as Anderson has suggested 'what can science learn from Google?', but to ask how the harvesters of Big Data might change the meaning of learning, and what new possibilities and new limitations may come with these systems of knowing.
> danah boyd and Kate Crawford, in *Critical Questions for Big Data*[1]

On 5 September 2012 a new Facebook page was created titled "R.I.P. Morgan Freeman"[2] causing a flurry of social media activity that echoed a similar statement about Freeman in 2010 that quickly became globally trending on Twitter. At the time of this writing, Morgan Freeman is *still* quite alive. Indeed, while it's generally known that the content produced through Twitter and Facebook is not fact-checked nor peer-reviewed, per se, the virality of this kind of content sustains fantastic stories. Perhaps it's the seemingly inconsequential status of tweets and posts that make their existence nearly effortless; for many, tweeting, posting, retweeting, and sharing is akin to breathing. And yet, the aggregation of these moments of content production have become much more weighty, moving academic disciplines and funding agencies, reconfiguring industry and government, and increasingly becoming part of everyday life, for some.

The potential of these activities lies in their capacities to focus and extend attention. Consider the map as an internet meme, as an object that circulates through mimicry. Buzzfeed, Upworthy, and other media aggregators[3] perennially rediscover 'the map', with posts that proliferate through social media as 'The 40 Maps That Will Change The Way You See The World'. Putting aside the complaint by cartographers and graphic designers that many of these representations come up short on cartographic standards, these kinds of media sharing events gather more publicity, with several thousands of shares, likes, and retweets, than perhaps the most well-written publication in our flagship journals. Impact delivered at the click of a 'thumbs-up' icon[4].

Big data marks a particular production of information, the "fourth paradigm"[5], and as boyd and Crawford suggest above, big data also rewires both how we know and the significance of that learning process. Furthermore, big data marks the establishment of who stands to benefit from this process. While there are different genres of big data (volunteered and not, ambient and not, firehose and trickle, etc.), I specifically intend to push back on the proliferation of studies that propose to utilize social media as 'big data' evidence[6]. I argue that we should lean more on the notion that social media are phenomena and less on the notion that social media are evidence of phenomena. In other words, the aggregation of social media as big data is not necessarily social science data. However, I do not mean to argue that social scientists should not pay attention; here, Manovich's pushback on the digital humanities is instructive: "These objections do not imply that we should not use new data sources about human culture and human social life or not analyze them with computational tools… But we need to carefully understand what is possible in practice as opposed to in principle."[7] Social media as big data signals a multiplicity of strategies -- some strategies more transparent than others. This is, strangely perhaps, an opinion formed *not* due to the erroneous and entertaining substance of much of social media (e.g. toilet tweets or the latest death hoax), but an argument that seeks to

situate the shifting conditions of knowledge amid the rush toward big data as The Next Big Thing. Totes OMG.

Sponsored Post

Google, Facebook, and Twitter are advertising and marketing regimes. They derive profit(ability) from their platforms. As each have become a basic infrastructure for the internet -- replacing a perhaps more heterogenous and distributed html-based architecture (denounced as 'web 1.0') -- they provide a basic communicative function, to create (often through an engineered 'serendipity') opportunities for consumption[8]. To draw upon the content delivered through these platforms is to add legitimacy to their corporate values. To produce maps of geotagged Twitter content works to showcase the hegemony of Twitter (and Twitter-like systems in other parts of the world) as a primary technique for the distribution of thoughts and observations. Treating this social media content as social science data reinforces the operative field generated by Twitter, Facebook, and Google -- by establishing them, discursively, as the preferred infrastructure of global communication.

There are few instances of walking the line between directing attention to the stream of content interrupted by 'sponsored' and 'suggested' posts and critically investigating the underlying attentional controls furthered by this content stream. For instance, the blog operated by the research group known as Floating Sheep refines the methods for visualizing the geoweb portion of this content, while also satirizes any serious social science treatment[9]. The point is to elevate the partiality and the commercialization of this infrastructure by recognizing that social media scholars are not mere observers or utilizers of social media content but are promoters of this infrastructure and its underlying advertising schemes.

We Are The 1%

Location-tracking is generally something that prosumers must choose to enable for their social media posts. Muki Haklay is at pains to remind geographers of the implications for this limitation, in the utilization of geosocial big data[10]. While Twitter is perhaps the most frequently used example of geographic representation of social media, only around one percent of tweets are geotagged[11]. However, for geographers, this has still meant a number of projects (proposed and funded) that seek to enroll geosocial content in various studies across the physical and human dimensions of geographic scholarship, from climate science to revealing various essentialisms. This social media content does not emerge from representative motivations. Instead, these maps enact what I term the 'represensational': sensational representation -- to lean on the production of world mappings of geosocial content, to admire the expansion of this corporate digital infrastructure, just as we tilt forward in our seats to examine maps of night lights, made more interesting by the negative spaces of the image.

In addition to the shearing created by these represensational maps, Monica Stephens further calls attention to the gendering of geosocial media[12], recognizing that this content is embodied, despite the posturing of this content as everywhere and evoked by nearly everyone. In the leap toward maps of global conditions using the infrastructure of suggestion and sponsorship, geographers must twist and contort the representative from the represensational, to constitute knowledge beyond the evocative. How do we shift the valuing of these contortions? Can we recover a representational approach that invites more questioning, more investment in the subject?

I Agree

Few users read the terms and conditions of the platforms they utilize to distribute their social content. Indeed, a kind of muscle memory is enacted and we click 'I Agree' as a barrage of text

insists on our permission. Google's most recent (11 October) update to their terms of service included the following statement: "We call these recommendations *shared endorsements*..."[13]. Our agreement to use these services, sets into motion an apparatus that monetizes sharing. Advertisers are empowered through access to these 'shares' -- enrolling the social networks of Google's users in attentional control techniques that draw more traffic to their sites and products.

Concerns about this have been lumped under discussions of privacy, and critical GIS scholars like Sarah Elwood and Agnieszka Leszczynski have documented the shifting discourses that organize these concerns[14]. What is signaled by 'I Agree' is not only a calculated and strategic publicity of individual content in social media, but is more -- holding hostage global communication through the reterritorializations of digital infrastructure for the purposes of commodification. Consider the ways in which even our universities move to adopt Google Gmail services in place of our existing university email systems, or the taking up of Facebook and Twitter as primary ways for nonprofit organizations and municipal governments to connect with constituents and advocates. 'I Agree' signals a new political economy.

Analytics™

Something one does has become a thing one buys. Data analysis has been a commodity for some time, for instance with the mid-century rise in municipal planning firms armed with data and new computing power (for instance, see McKinsey & Company's more recent shift into this sector: "We turn data into actionable insights and improved performance."[15]). Now, following an increase in the informatics sector of many marketing firms, the social sciences are witnessing the potential for training new 'data analysts' to peddle a vision of smart cities, ubiquitous health-monitoring systems, real-time global brand nurturing, and geointelligence and security. While this may be 'good news' for academic units competing to attract potential student/customers into classrooms increasingly

funded by tuition over state support (e.g. the popularity of Penn State certificates in homeland security and geospatial intelligence[16]), this development is troubling for units that do not 'play well' with the rise of big data. In a move that reminisces of Haraway's[17] marking of biotechnological commodities ("OncoMouse™"), I conceptualize Analytics™ as a way to signal the rise of a commoditization of big data.

Analytics™ enacts a narrowing of what is meant by analysis, just as it places limitations on the types of data that are permissible. Where do we press the button for 'critical analysis'? How do we input interviews and public discourse into Analytics™? It is impractical to view the rise of real-time big data analytics as distinct from the sloganeering that permeated the rise of GIS in the early 1990s, and the rebranding of GIS as GIScience in the late 1990s. However, the nagging skepticism around logical positivism of that decade within our discipline should be perhaps redressed now as a surging commodification of method. Analytics™ limits to what we pay attention, while also conditions the mechanics of those attentional systems. In other words, in the wake of fantastic mechanisms for tuning the urban-regional system, paid for by taxpayers, how do we advocate for the transparency of these analytic regimes, in order that we might reconfigure their lines of questioning?

* * *

Certainly, there are other forms that big data take. Here, I'm referring only to social media as big data, and I do recognize that scholarship has and may emerge that analyzes social media as evidence of some phenomena, much in the way that Google may use search activity as a proxy to understand the spread of flu. However, this viewpoint is a challenge to consider the ways in which this kind of scholarly attention actually reinforces the shifting political economy of data[18], rather than attempt to

call attention to and even rework that economy. Instead, I suggest that criticality in big data studies requires a duality similar to the technopositional stance of critical GIS[19] -- to both make critical use of big data and critically situate its provenance. I have sketched four hooks into this criticality, beginning with the presumption that social media are not necessarily evidence of phenomena, but are phenomena. This skepticism provides a clearer starting point for big data, a more sobering position amid infectious hype, and a reminder that knowledge is formed, not found or scraped.

That Morgan Freeman is *not* dead serves as a reminder of these contingencies of knowledge-in-formation. The survival of critical geographic representations of social media depends upon such a vigilant skepticism, of a 'nice map, but' response, that actually serves to extend discussion of our complicity in the advance of digital culture, rather than solidify our neutral vantage point on the proliferation of studies that utilize social media as big data.

---

[1] boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society, 15(5), 662-679. doi: 10.1080/1369118X.2012.678878

[2] Facebook. (2012). R.I.P. Morgan Freeman. Available at: http://www.facebook.com/pages/RIP-Morgan-Freeman/274408782669578

[3] Online media aggregators curate and organize content that resides online at other websites. These aggregators effectively allow a speeding-up of content distribution and provide the infrastructure for virality.

[4] The 'thumbs-up' icon has become ubiquitous as a way to respond affirmatively to a piece of shared media content, made more interesting by the lack of 'thumbs-down' icons.

[5] DeLyser, D., & Sui, D. (2013). Crossing the qualitative-quantitative chasm III: Enduring methods, open geography, participatory research, and the fourth paradigm. *Progress in Human Geography*. doi: 10.1177/0309132513479291

[6] For example, see Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. Paper presented at the Web Intelligence and Intelligent Agent Technology (WI-IAT), IEEE/WIC/ACM International Conference.

[7] Manovich, L. (2012). Trending: The promises and the challenges of big social data. In M. K. Gold (Ed.), *Debates in the Digital Humanities* (pp. 460-475): Minnesota; see also DeLyser, D., & Sui, D. (2013). Crossing the qualitative-quantitative chasm III: Enduring methods, open geography,

participatory research, and the fourth paradigm. *Progress in Human Geography*. doi: 10.1177/0309132513479291

[8] Wilson, M. W. (2012). Location-based services, conspicuous mobility, and the location-aware future. Geoforum, 43(6), 1266-1275. doi: http://dx.doi.org/10.1016/j.geoforum.2012.03.014

[9] See also, Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. A. (2013). Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. Cartography and Geographic Information Science, 40(2), 130-139. doi: http://dx.doi.org/10.1080/15230406.2013.777137

[10] See also, Haklay, M. (2013). Neogeography and the delusion of democratisation. Environment and Planning A, 45(1), 55-69.

[11] Murdoch, V. (2011). Your mileage may vary: on the limits of social media. *SIGSPATIAL*, 3(2), 62-66. doi: 10.1145/2047296.2047309

[12] Stephens, M. (2013). Gender and the GeoWeb: divisions in the production of user-generated cartographic information. GeoJournal.

[13] Google. (2013). Policies & Principles: Terms of Service update. Available at: https://www.google.com/intl/en/policies/terms/changes/

[14] Elwood, S. A., & Leszczynski, A. (2011). Privacy, reconsidered: New representations, data practices, and the geoweb. Geoforum, 42, 6-15.

[15] McKinsey & Company. (2013). McKinsey Advanced Data & Analytics. Available at: http://www.mckinsey.com/client_service/advanced_data_analytics

[16] http://www.worldcampus.psu.edu/degrees-and-certificates/homeland-security-geospatial-intelligence/overview

[17] Haraway, D. J. (1997). *Modest_Witness@Second_Millennium. FemaleMan©_Meets_OncoMouse™: feminism and technoscience*. New York: Routledge.

[18] Leszczynski, A. (2012). Situating the geoweb in political economy. Progress in Human Geography, 36(1), 72-89; Wilson, M. W., & Graham, M. (2013). Situating Neogeography. *Environment and Planning A, 45*(1), 3-9.

[19] Wilson, M. W. (2009). Towards a genealogy of qualitative GIS. In M. Cope & S. A. Elwood (Eds.), *Qualitative GIS: A Mixed Methods Approach* (pp. 156-170). London: Sage.