

Applications of Big Data

The seeds for this *Geographical Analysis* special issue on “Applications of Big Data” date back to a series of sessions we organised at the annual conferences of the American Association of Geographers in Chicago (2015) and San Francisco (2016). We originally noted the rise of what was starting to be known as Big Data, or, in other words, the explosion of new forms of data generated thanks to cheap computing and geospatial technology. We also realised that the Human Geography community was less active in engaging with these new developments than other areas of (Social) Science, despite obvious opportunities and potential. So, we decided to “test the waters” and see if other researchers felt the same way. Despite the usual room (mis)allocation and unfortunate time slots (the Chicago session was scheduled at 8am and the San Francisco one took place on the last slot of the conference), the rooms were packed with not only presenters but more interested researchers to whom this offered a fresh breath of air.

Fast-forward to the beginning of 2018, when these lines are being written, the landscape in Geography has radically changed. We seem to have passed “*peak Big Data*”, as there is a feeling that many of the reasons why it made sense to call this confluence of technological advances (e.g. data sets were very large) were maybe not the most useful ones to understand the underpinnings of these new forms of data. The underlying conditions that originally prompted the Big Data craze however are very much still with us. What has changed is the realization that size might not be the best way to characterise this new *golden age* of data. Instead, there is more and more talk of “new forms of data”, or variants thereof. This shift recognises that, irrespective of the size, many of these more recent data sources are different from traditional ones in fundamental ways, and thus cannot simply be treated as *more of the same*. This mindset is behind the shift towards the idea of developing a Data Science ([Donoho, 2017](#)) that creates methods to address these specific characteristics and fully exploit the opportunities these data provide. Similar initiatives are starting to take shape in Geography as well.

These two years of editorial process have also allowed us to take some perspective and reflect on the key areas of interest for researchers interested in new forms of data and their applications in Human Geography. We consider these to be mainly three: development of new methods; exploration of the nature and properties of these new sources of information; and brand new things we could not do before that are possible now. Thankfully, we believe the three of them are represented in this special issue. [Pavlis et al. \(2017\)](#) depart from a breed of datasets that, although not new, it has been little accessible until recently: internal corporate data or, what is now being termed “consumer data”. Using an extensive and detailed database of individual retail locations, they devised a modified version of the popular machine learning clustering technique DBSCAN to delineate retail centre areas. In a similar spirit, but very different context, [Arribas-Bel & Tranos \(2017\)](#) present a new visualisation approach that allows to explore and understand urban dynamics when captured from fine-grained datasets such as mobile phone activity. This exercise enables them to capture the fast and slow changes of the city of Amsterdam. [Folch, Spielman & Manduca \(2017\)](#) instead focus on an area that has not

received nearly enough attention: characteristics, quality, and accuracy of the data, particularly as compared to more traditional sources. Using the case of the social network Yelp in the Phoenix metropolitan area, they benchmark its spatial distribution against that of an administrative source. Their study finds that, as has been suggested in the literature, these sources of data contain certain inherent biases. More surprisingly however, they also use the Yelp dataset to document equally systematic biases in the administrative dataset. Finally, [Poorthuis \(2017\)](#) represents an excellent illustration of some of the genuinely new opportunities new forms of data afford us. Using data from a geo-referenced social network, the paper revisits the old question of regionalisation and boundary delineation, providing solutions that would not be possible without the nature and resolution of his dataset.

Overall, we think that the collection of these papers exemplifies the critical use of a side-product of the digital revolution mentioned above in order to improve our capacity to understand cities and urban systems from a social science perspective. Having overcome the hype around Big Data, we are now better positioned to assess the utility of these data sources and create new and appropriate methods to deal with their complexity. Looking back, this has been a great journey for us as guest editors and we hope the same for the authors. We are thankful to Prof. Sergio Rey who originally instigated the idea of a special issue and Dr Rachel Franklin who took over halfway the process for their support and patience; the authors for all their efforts; the reviewers for their thoughtful and constructive comments; the audience of the AAG special sessions. And to the GA readers: we hope you find this issue useful and enjoy it nearly as much as we did. That would be a clear sign of success for us.