



## Understanding urban China with open data



Xingjian Liu<sup>a</sup>, Yan Song<sup>b</sup>, Kang Wu<sup>c</sup>, Jianghao Wang<sup>d</sup>, Dong Li<sup>e</sup>, Ying Long<sup>f,\*</sup>

<sup>a</sup>The University of Hong Kong, China

<sup>b</sup>University of North Carolina – Chapel Hill, USA

<sup>c</sup>Capital University of Economics and Business, China

<sup>d</sup>Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, China

<sup>e</sup>China Academy of Urban Planning and Design, China

<sup>f</sup>Beijing Institute of City Planning, China

### ARTICLE INFO

#### Article history:

Available online 29 March 2015

#### Keywords:

China  
Open data  
Urban analytics

### ABSTRACT

A solid understanding of urbanizing China – the world's largest and most rapidly transforming urban society – calls for improved urban data provision and analysis. This paper therefore looks at major technological, social-cultural, and institutional challenges of understanding urban China with open data, and showcases our attempt at understanding Chinese cities with open urban data. Through our showcases, we hope to demonstrate the usefulness of open urban data in (1) mapping urbanization in China with a finer spatiotemporal scales; (2) reflecting social and environmental dimensions of urbanization; and (3) visualizing urban China at multiple scales.

© 2015 Elsevier Ltd. All rights reserved.

### Introduction

With more than 700 million urban residents and another 30 million people being added to its cities every year, China becomes the world's largest and most rapidly transforming urban society (Bai, Shi, & Liu, 2014). Urbanization in China has therefore attracted wide attentions from researchers, policy makers, investors, and the general public. The number of articles featuring “urban China” has risen by 600% in the last decade, based on bibliographic information gathered from the ISI Web of Science. Public awareness about a wide range of urban issues, ranging from public safety, air pollution, to food security, has reached a historical height (Enserink & Koppenjan, 2007). Meanwhile, the recently released “National New-Type Urbanization Plan (2014–2020)” has marked a new consensus among the country's ruling elites to foster sustainable and people-centered urbanization patterns (Bai et al., 2014; Wallace, 2014).

A solid understanding of urbanizing China hinges on improved measurements of Chinese cities, although gaining more data is seldom a solution by itself, as rigorous analyses and insightful communications are necessary to deliver derived information to a wide variety of audiences. Recent studies in urban analytics identify open urban data as a promising complement to conventional authoritative data sources in understanding urban form and

functions (Crooks et al., in press; Goetz & Zipf, 2012; Jia & Jiang, 2010). Furthermore, for most developing countries that lack conventional data infrastructures, marshalling existing open data from various sources seems to be a viable option. The purpose of this report is therefore twofold. First, it reviews the major technological, socio-cultural, and institutional challenges of assembling openly accessible urban data for Chinese cities. Second, the paper showcases one specific attempt – the Beijing City Lab – at understanding Chinese cities with open urban data.

### What are open urban data?

Before discussing the challenges in assembling and using open urban data in China, we briefly provide a working definition of open urban data within the context of our discussion. We use urban data as an overarching umbrella term for all datasets that would characterize some aspects of urban form and functions (Crooks et al., in press). Our definition of open urban data subsequently is therefore broad and refers to urban data that are openly accessible to the public. Preferably such data are available in convenient and adjustable formats, but as we will elaborate, this is not always the case. Open urban data mainly come from three overlapping though different sources (Gurin, 2014): official data portals, big data initiatives, and the broader open data community.

Urban data have conventionally been produced and maintained by government agencies. These traditional urban datasets range from cadastral maps prepared by land departments and traffic

\* Corresponding author at: No. 60, South Lishi Rd, Beijing 100045, China. Tel.: +86 10 8807 3660.

E-mail address: [longying1980@gmail.com](mailto:longying1980@gmail.com) (Y. Long).

records maintained by transportation authorities, to business transactions monitored by taxation agencies. Recent open government initiatives have propelled governments to grant public access to a greater amount of authoritative data sources. This is best evidenced by mushrooming data portals for US cities in the wake of The Digital Accountability and Transparency Act of 2014 and other related local legislations. For example, NYC OpenData, the official online data portal of New York City (NYC), provides open access to more than 1300 governmental datasets, some of which are quite big in terms of volume.

Big data enable us to gauge urban dynamics at finer spatial and temporal scales (Kitchin, 2014; Yue et al., 2014). The generation of big data is characterized by large volume, great variety, and high velocity. Typical big data in the urban environment include mobile phone activities, vehicle trajectories, public transit smart card data, business catalogues, as well as other information generated by various smart cities programs (Batty, 2012). These types of data often enable researchers to capture urban dynamics at very fine spatiotemporal scales. Many of these datasets are often made available by private data holders via public–private partnerships and data challenges. For example, the ‘Data For Development’ challenge (<http://www.d4d.orange.com/en/home>) organized by the Orange Group released mobile phone usage datasets for African countries Ivory Coast and Senegal.

Volunteer Geographic Information (VGI) and Crowdsourcing (Goodchild, 2007; Crooks et al., in press) are major contributors to urban data within the general open data community. As with other grassroots information campaigns (Howe, 2006), urban data in this category are contributed by the general public and represent a ‘bottom-up’ data generating process. Notable examples include collaborative VGI mapping platforms (e.g., OpenStreetMap) and location-enabled social media applications (e.g., Foursquare, Twitter, and Flickr).

Open urban data can be further differentiated along three dimensions: (1) Online/offline: direct downloading from the Internet is the preferred access mode of open data, and indeed many examples we have given so far are online sources. Nevertheless, we note that not all open urban data can be accessed online. Offline open urban data include *inter alia* hard-copy government reports, historical paper maps, as well as yet-to-be-digitized library collections. (2) Explicit/implicit: Crooks et al. (in press) emphasize the distinction between explicit and implicit urban data. Many open urban data are of the former type, being readily available for use in the desired form. For example, datasets from the NYC OpenData portal are stored in standardized tabular and geospatial data formats. By contrast, implicit urban data capture information about urban dynamics, but cannot be applied directly in analyses. Such implicit data need to be derived and pre-processed. For example, the business cataloguing website Yelp contains a wealth of business establishments (e.g., location, ratings, and descriptions), however such information needs to be crawled from individual webpages and geocoded before any subsequent analyses. (3) Usable/re-usable/redistributable: Open datasets are subject to different sharing requirements and codes. The more open the data, the less strings attached: At one end of the spectrum, open data allow for free use, re-use, and re-distribution by anyone. However, on other occasions, the use of datasets may be restricted for education purposes. For example, datasets released by the aforementioned ‘Data for Development’ data challenge are for education and research only.

### Open urban data in China

The increasing supply of scientific data, spearheaded by numerous open data initiatives, has transformed research practices in

many fields (see for example, a *Science* magazine special issue on data and new science initiatives; *Science* Vol. 331, Issue 6018). These new waves of data initiatives are also opening up new research opportunities in urban studies and planning (Batty, 2013; Crooks et al., in press). Meanwhile, researchers still lament the dearth of urban data in developing countries. While European and North American cities are being measured with greater precision and granularity, cities in the developing world are oftentimes less well-charted. In other words, there seems to be a widening gap in terms of openly available data between cities in developed and developing worlds (Graham, Hogan, Straumann, & Medhat, 2014; Lang, 2011). Overcoming such gaps in data availability seems to be of high priority for urban studies and planning in developing countries in general and China in particular.

As in many developing countries, government agencies are the primary supplier of urban data in China and exercise tight data regulations for reasons such as national security. Nevertheless, the Chinese government is pushing for more open access to data via open government initiatives and loosening control of data use (see for example, the launch of an official national data repository: <http://data.stats.gov.cn/>). Recasting Reichman, Jones, and Schildhauer (2011)’s insights and framework about challenges of open data in ecology, we have identified the following technological, social–cultural, and institutional challenges of collecting open urban data in China.

A first technological challenge concerns the dispersion of urban data, similar to the cases in ecology (Reichman et al., 2011). In general, efforts to produce urban data are rather fragmented, as most publicly available data about Chinese cities are scattered in online and offline sources. As discussed previously, data sources range from user-generated contents on the Internet, governmental yearbooks and websites, datasets released by private companies,<sup>1</sup> to datasets produced by research projects. Additionally, there exist substantial overlaps between data collected via different venues, creating discrepancies for data use and wasting resources. For example, due to the rather complicated *tiao/kuai* administrative framework in China (Mertha, 2005), (urban) land use data are being collected independently by a number of government agencies, such as land use and resources, urban planning, as well as agricultural and forestry departments.

With the dispersion of data sources comes the second technological challenge: the heterogeneity of data. Data from different sources are usually produced and maintained in different formats, focusing on different urban issues, capturing different years, and covering different geographic areas (Reichman et al., 2011). For example, information about urban construction permits is usually available from individual municipal governments’ websites (see for example, <http://www.bjjs.gov.cn/tabid/660/Default.aspx>). Such information is critical for understanding urban growth patterns, as construction permits data capture the ‘where’ of urban development at a very fine spatial scale. However, permits data of different cities do not come along in the same format: some exist on fully-fledged websites that enable online information retrieval, others take the form of offline archives and require digitization; some are explicit and encoded in tabular forms, others are implicit and need to be derived from plain texts; some are spatially explicit and contain longitudinal and latitude coordinates and are thus readily available for mapping, others need further geocoding.

Furthermore, the pace of urbanization in China posits a unique challenge: it is unprecedented and faster than the frequency of many regular governmental monitoring and census programs.

<sup>1</sup> In addition to businesses operated by large consulting and data companies, proprietary data, ranging from digitized yearbook to sales records of individual companies, can be purchased via many small vendors on taobao.com – the Chinese equivalent of eBay.

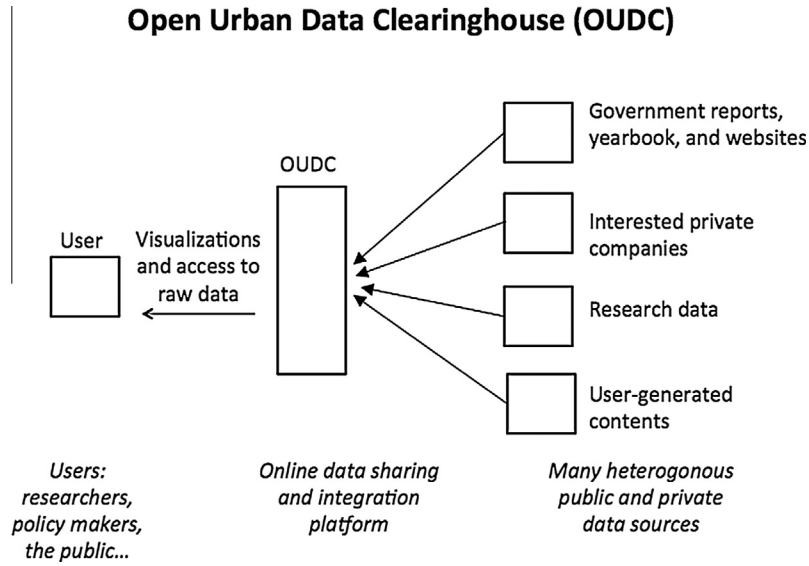


Fig. 1. Open urban data clearinghouse.

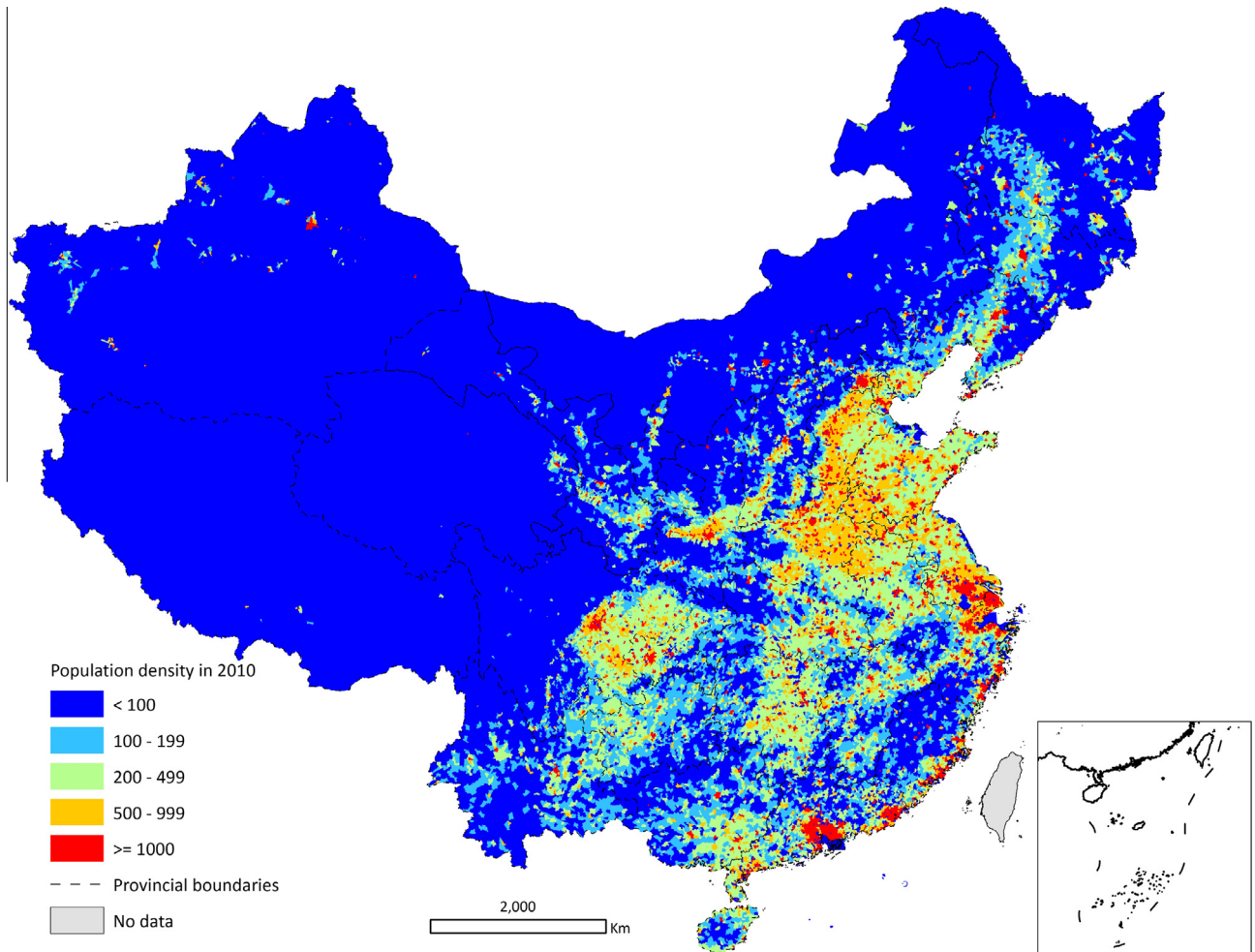


Fig. 2. Population density at sub-district level.

For example, the census in China is conducted every 10 years, while around 200 million people became new urban residents in China during 2000–2010. Urban dynamics in interstitial years thus need to be measured via alternative ways. This calls for more

frequent sharing and integration of data sourced from multiple (public and private) sources. These sources may not have the large thematic and geographic coverage as those governmental programs, however they reveal the most 'updated' situations of

Chinese cities. In a sense, the entire dataset about Chinese cities, from which we derive our perception of urbanizing China, could be updated incrementally via “patches” submitted by individual contributors, instead of regular full-scale overhauls by the governmental programs. Still, issues such as meta-data documentation and data ownership specification need to be resolved before publicly sourced data can be used for scientific and policy research (Overpeck, Meehl, Bony, & Easterling, 2011).

These technological challenges call for inter-linked, inter-operable, and frequently updated urban datasets. While international attempts, such as the Urban Data Observatory (<http://www.urbanobservatory.org/>) and the New York City (NYC) Open Data (<https://nycopendata.socrata.com/>), are under way with support from both public and private sources, systematic collections and publications of data for Chinese cities are still in their infancy (notable exceptions include a number of urban form indicators shared by The Peking University – Lincoln Institute Center).

Although overcoming these technological challenges for sharing urban data seems to be daunting, the institutional and socio-cultural hindrance incurs additional difficulties (Reichman et al., 2011). The State has been playing the wild card in shaping open data movements in China. For example, the few open data initiatives operated by large governmentally-linked institutions enforce certain access control. For example, these initiatives mostly grant restricted access to researchers who are affiliated with public universities and research institutions in China, and have gone through data application processes, in a sense limiting data access for *inter alia* overseas researchers, private institutions, and the general public. Nevertheless, we are witnessing a promising trend towards more open governmental data portals and programs such as the Shanghai Municipal data network (<http://www.datashanghai.gov.cn>).

In addition, providing incentives for data contributors has been a fundamental challenge for data sharing in general (Poldrack & Gorgolewski, 2014). As for the example of research generated data, Reichman et al. (2011) point out that granting access to data before the final publication of results might effectively jeopardize publication/analysis plans of individual researchers (see also Sui, 2014 for a more general discussion on open science). In this regard, like many of their international colleagues, Chinese urban researchers may have little incentive to share datasets and embrace the idea of open data: the centrally-designed academic performance evaluation system puts much emphasis on publication outputs (see for example, Wang, 2011), which are intricately related to the uniqueness, novelty, and ultimately ‘publishability’ of data. This is also reflected by the fact that, among researchers who are willing to share their data, most still prefer to release datasets after the results from these data have been published.

While it is unlikely that such social, cultural, and institutional norms against data sharing will change soon, we believe that one potential remedy at the current stage might include (1) a carefully designed and centrally managed data clearinghouse system; and (2) a devoted group of researchers who operate the clearinghouse, possess data integration capabilities, and have strong willingness to share their own data, thus essentially providing ‘public goods’ to the research community. The data clearinghouse matches the scattered data supply with the growing needs for urban data. More specifically, such a clearinghouse platform would standardize, spatialize, synthesize, and categorize datasets from governments, interested private enterprises, and academic studies, while providing online data access to the public (Fig. 1). Meanwhile we hope that the sharing behavior of an initially small group of researchers would set the change of social-cultural traditions in motion and facilitate the formation of an ‘open urban data’

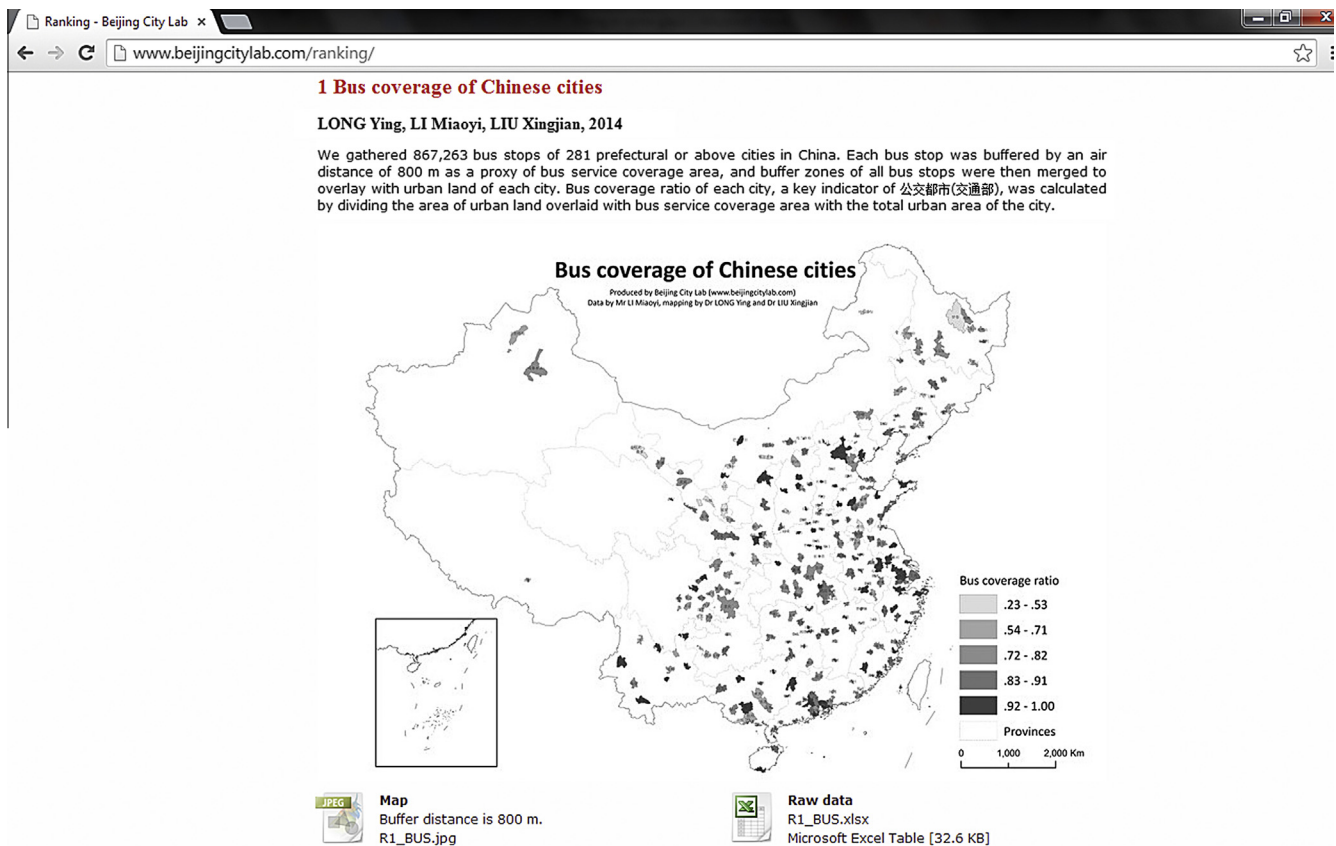


Fig. 3. The BCL ranking of bus coverage.



community. The clearinghouse would have a built-in curating component, which grants controlled access to users (Reichman et al., 2011). For example, as one remedy for the jeopardized publishability of shared data, the curating system would aggregate submitted datasets at coarser geographical scales, thus fulfilling the goal of sharing data to a wider audience as well as safeguarding the more detailed datasets for original contributors' own use.

### An attempt at understanding urbanizing China with open data

Modeled on the urban data clearinghouse system envisioned in Fig. 1, we setup the Beijing City Lab (BCL; <http://www.beijingcitylab.org>), an online research network to produce and collate data about Chinese cities. The network is consisted of a group of core urban researchers and a larger group of concerned planners, architects, geographers, economists, and policy analysts. The core researchers' roles include among others (1) identifying gaps between data needs and supplies based on their domain knowledge; (2) generating and enhancing urban data via their own research; (3) synthesizing and standardizing datasets shared by others; and (4) maintaining the online data repository and serving as curators of the data. The larger group of Lab members is not involved in the daily operation of the project, but these members are active suppliers and users of the data collection. Taken together, major sources for the BCL data repository include (1) standardized open governmental data: these datasets include official

information (reports, yearbooks, and governmental websites) digitized, geocoded, and integrated by the core members; (2) research data produced by BCL members: this includes among others some big data (smart card data) released by proprietary data holders for research purposes; and (3) information extracted from social media (e.g., geotagged Sina Weibo – the Chinese equivalent of Twitter) and crowdsourcing projects (e.g., OpenStreetMap) by the core members.

To date, we have consolidated and shared more than 30 nationwide urban datasets. The identification of these datasets draws upon both domain knowledge and technical capacity, i.e., being able to identify technically obtained measurements that address important urban questions. Though the number of datasets may not seem large, the data being shared cover many central topics in urban studies, such as block boundaries, land use mixing, population exposure to environment pollution, and the supply of public goods. For example, the shared data have been cited/used in multiple media reports, policy documents, and research studies (e.g. media reports on several BCL rankings and projects, see [www.beijingcitylab.com](http://www.beijingcitylab.com) for more details), and the website of the Beijing City Lab has gained more than 20,000 visits during the network's first six months of existence.

We encourage individual contributors to submit their data at 1 km × 1 km grid, in the hope of providing more incentives for data sharing. As for the example of crawled geotagged Sina Weibos, instead of uploading the detailed coordinates of several million

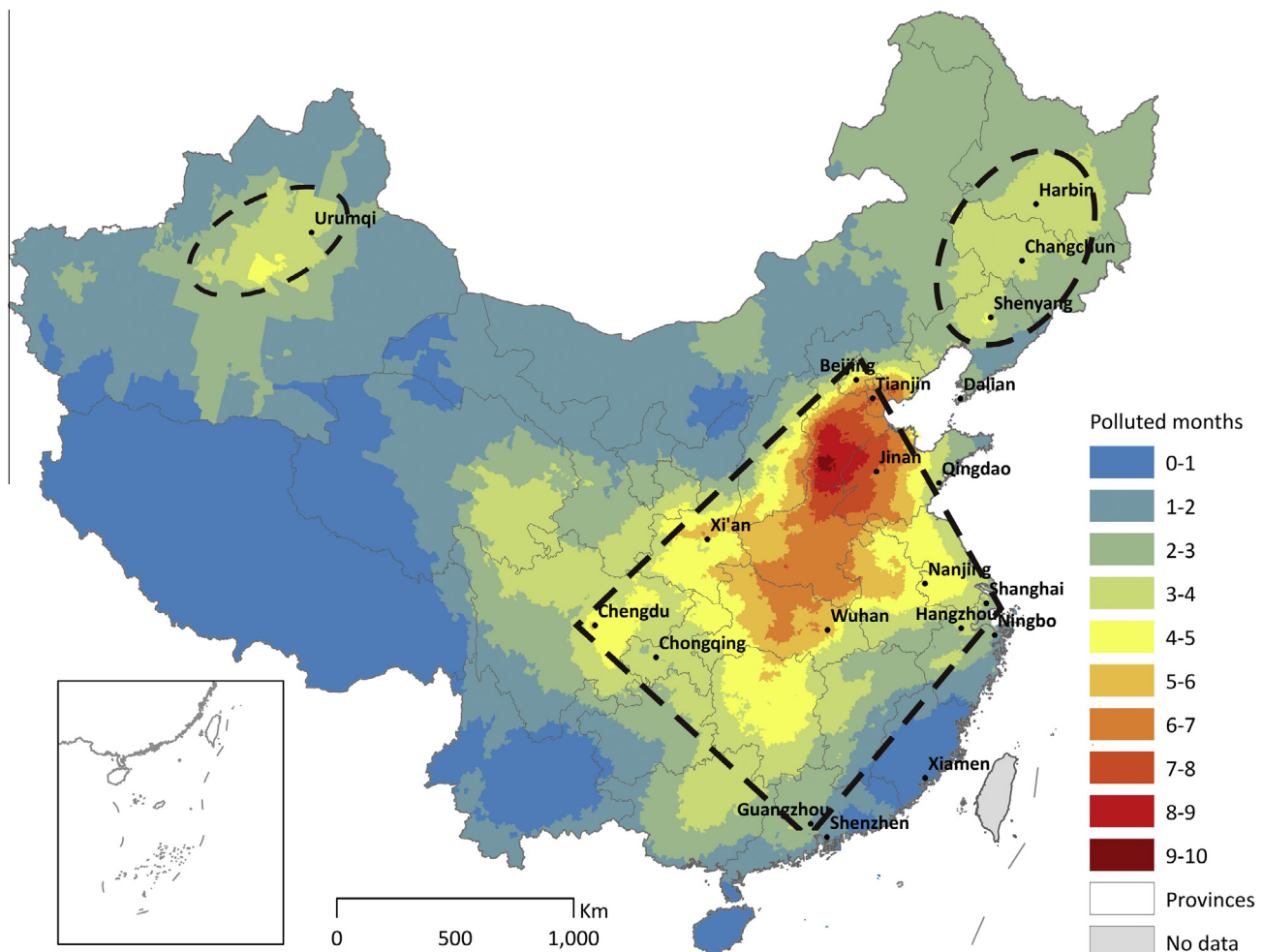


Fig. 4. Urban environments at fine spatial scales: The number of polluted months in a year for each Chinese sub-district violating national PM<sub>2.5</sub> standard. Note that the dashed lines indicate the most PM<sub>2.5</sub> polluted areas.

Weibos, the data contributor can opt for submitting the aggregated number of Weibos within individual grids. On the one hand, individual data holders would still keep their datasets at finer spatial scales, thus not jeopardizing opportunities for publication. On the other hand, data aggregated at  $1\text{ km} \times 1\text{ km}$  grid can still be useful for regional and urban analysis. Relatedly, we have provided software tools and manuals to help potential data contributors aggregate their datasets onto  $1\text{ km} \times 1\text{ km}$  grids.

With a repository of urban data about Chinese cities in hand, we endeavor to remedy some of the previously mentioned difficulties and limitations, and, more specifically, we have strived to: (1) model urbanization in China with a greater geographic coverage and at finer spatial scales; (2) introduce a people-centered perspective in the understanding of cities, which emphasizes more the social and environmental dimensions of urbanization; and (3) disseminate and explore data via various visualizations. Below we enumerate a list of our recent projects and briefly discuss how we further the understanding of urban China with open data.

#### Examining the national urban system at finer spatiotemporal scales

Cities that contain small populations and/or are economically less advanced are usually overlooked and in a sense ‘off the map’ in public media, government reports, and academic studies. Most of our analyses cover all 654 cities at county-level and above, our

project effectively puts all these less significant places ‘on the map’ and enables systematic comparisons of cities on a variety of fronts. For example, while large cities along the Chinese eastern seaboard such as Beijing, Shanghai, and Guangzhou are competing for the status of ‘global cities’, relatively little is known about economically less-advanced cities in the vast Chinese inland. Part and parcel of this problem is the lack of information and essentially ‘off’ our mental maps. Therefore, gathering data about cities would generate more information about these ‘off-the-map’ cities and facilitate their efforts to (re)gain their public visibility.

Our projects also strive to combine large geographic coverage with fine spatial scales. As one example, we have looked into the basic building blocks of the city proper in China: *Jiedao* (sub-districts). More specifically, we geocoded all township-level units based on the Population Census of China, and estimated population densities for the years of 2000 and 2010 using official township boundaries (Fig. 2; Wu, Long, Mao, & Liu, 2015). Along the same lines, we have produced block-boundaries for all Chinese cities, identified urban blocks based on their functionalities, and simulated urban growth for five year periods at the block level (Long, Wu, & Mao, 2014). These datasets substantively improve the basic spatial units for further exploration of Chinese cities, as cities are now being measured on a unit that corresponds to the spatial extent of daily routines, i.e., the urban block, rather than the conventionally used

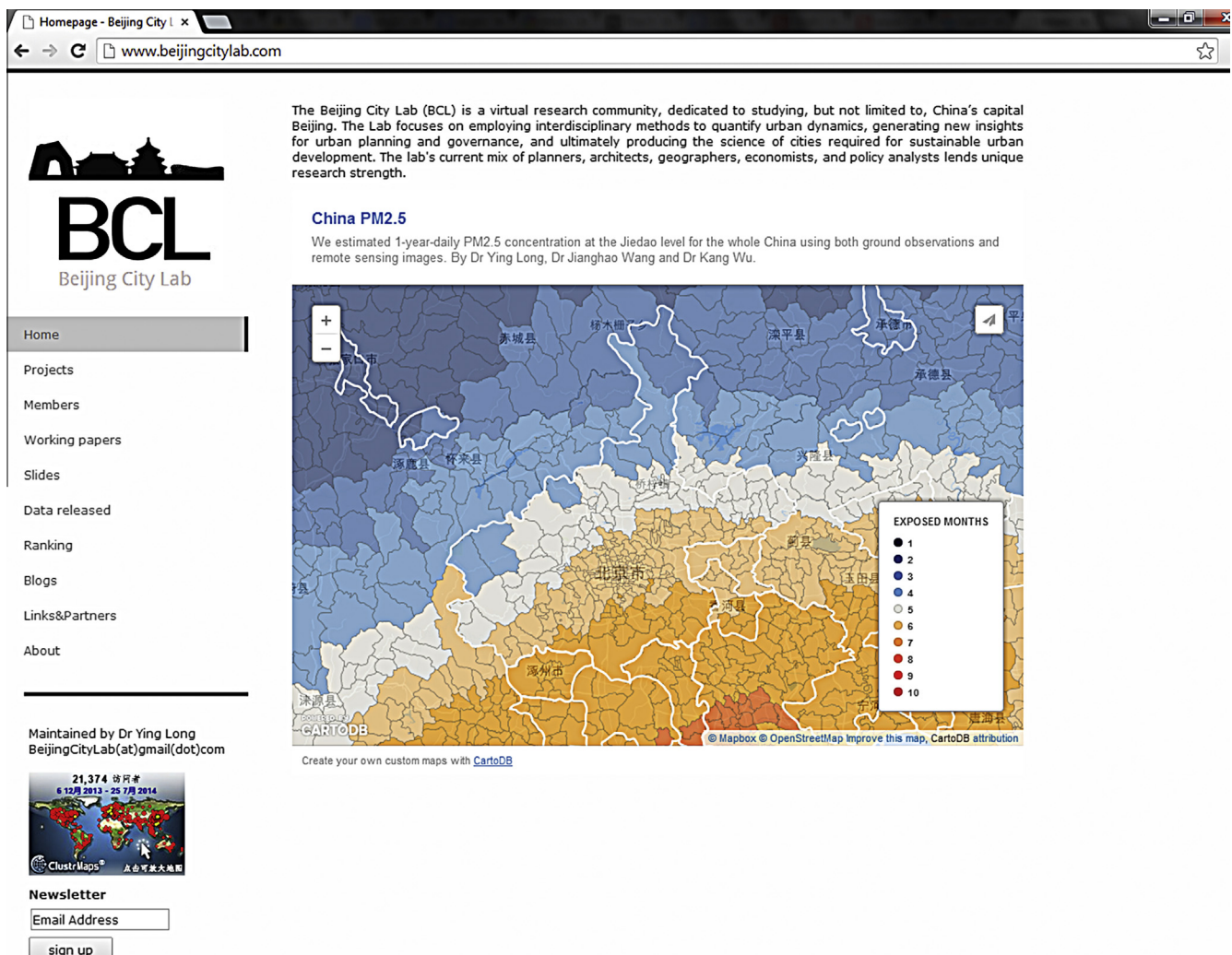


Fig. 5. Maps for multi-scale exploration on the BCL website.

regular grids that bear little resemblance to urban functional areas (Long & Liu, 2013). The academic community has started to take up new research opportunities provided by our datasets: see for example, Waddell (2014) who has generated 3-d visualizations of urban blocks within major Chinese cities with the BCL data. In addition, we also attempt to extend temporal analyses with open available data. For example, planning permits issued in Beijing during the past ten years are displayed in text formats on the city's planning commission website. We have crawled, integrated, and geocoded this information, which provides a valuable source for tracking Beijing's development over the past decade.

*Measuring people-centered urbanization*

Our projects exemplify a people-centered perspective, featuring environmental and social consequences of rapid urbanization. This is evidenced by the collection of BCL rankings of Chinese cities, which rank cities based on a number of non-conventional indices that focus on social well-being, such as bus coverage, per capita religious facilities, and mobile Internet penetration (Fig. 3). These seemingly unorthodox indicators introduce an individual-based everyday-life related perspective on cities, emphasizing citizens' rights to the cities. The cities are interpreted as places to provide facilities and services to the urban residents. While conventionally Chinese cities are deemed as spaces of production, these people-

centered indicators may help to refocus our attention on cities' role as social spaces (Pacione, 2003).

For individual rankings, we provide corresponding maps, data tables, and research highlights, as well as technical supporting materials. While the BCL rankings can be readily accessed by the general public, the associated research insights provide additional details to be used for policy reports. For example, the ranking of Chinese cities by bus coverage have been cited by a number of media outlets and used by at least two policy reports submitted to municipal governments.

Furthermore, driven by the urgent environmental issues in China, one ongoing investigation systematically gauges population exposure to particle matter (PM<sub>2.5</sub>) in China, integrating shared ground monitoring data and MODIS AOD images and thus producing the first mapping of fine-scale and one-year PM<sub>2.5</sub> for China (Fig. 4; Long, Wu, Wang, & Shen, 2014). Using the national government's threshold for air pollution, the map suggests among other things that (1) nearly a quarter of the nation's land territory – 243 million square kilometers to be exact – has PM<sub>2.5</sub> pollution over more than 3 months; and (2) most Chinese townships have 1–2 months pollution per year. Combining pollution patterns with town-level population census, more alarming results emerge as for example that more than 800 million people are exposed to 'pollution' for more than 3 months every year. Other projects concerning the social and environmental dimensions of urban life

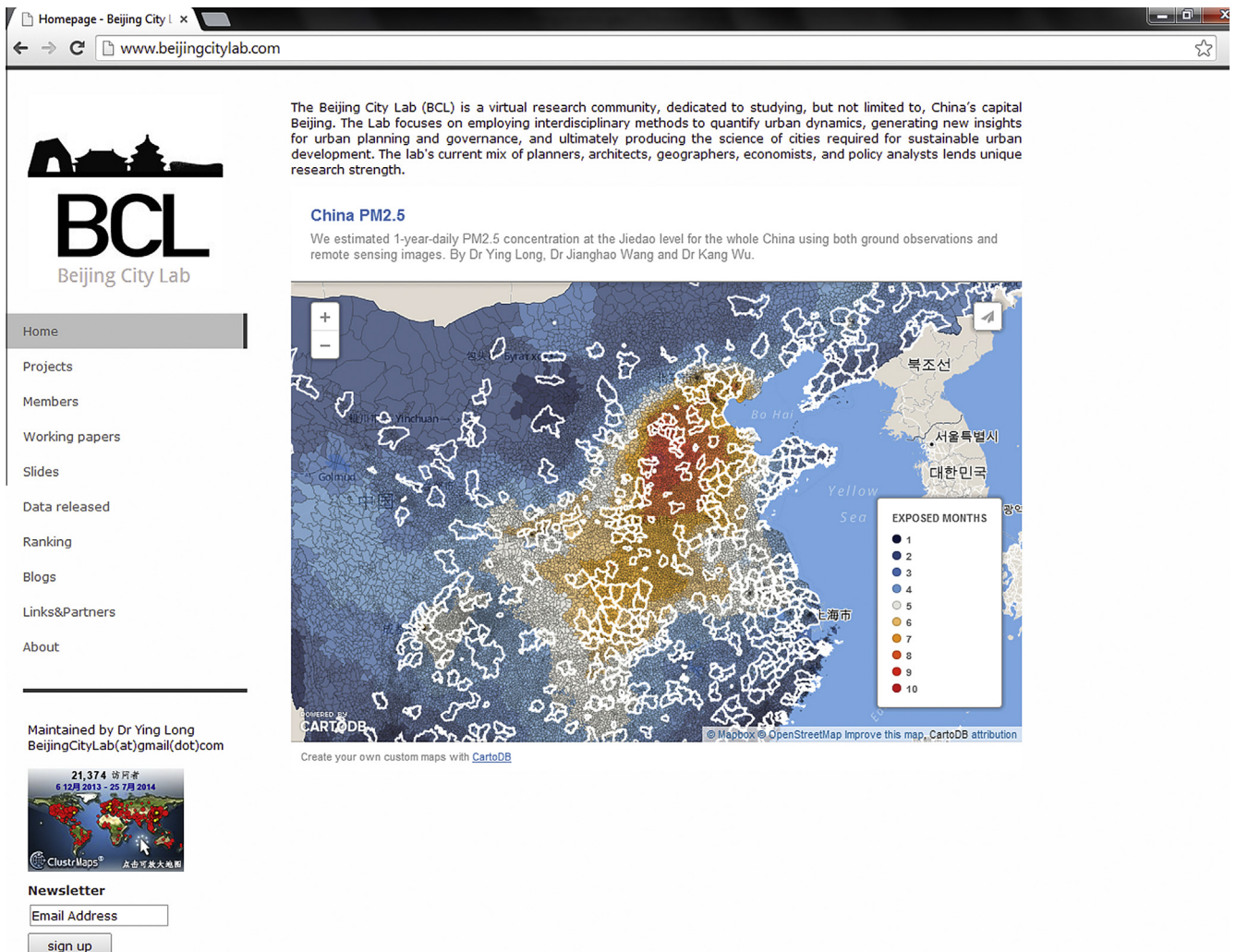


Fig. 5 (continued)



include modeling the spatial patterns of frequent public transit riders with released smart card data of Beijing; comparing housing prices for major Chinese cities based on real estate websites; as well as mapping urban gastronomy through the density of restaurants for major Chinese cities (Liu, 2014).

#### Using visualization as an exploration tool

As already shown in previous maps, we deem data visualizations as exploration tools (Fox & Hendler, 2011). Visualizations stand for one core dimension of spatial science, elaborating the spatial distribution and dynamics of social, economic, and political activities through visual instruments such as maps (Crampton, 2009). Many of our datasets cover most Chinese cities at fine spatial resolutions for multiple periods, thus making it difficult to render all information with a few static visualizations. Furthermore, as urban issues are often multi-faceted, multiple representations of the same urban phenomena could corroborate or contrast with each other, generating more complete understandings. We have therefore uploaded most of our datasets onto CartoDB – an online mapping platform – to allow for multi-scale and oftentimes dynamic explorations. Figs. 5(a) and (b) reveal how PM<sub>2.5</sub> exposure can be explored at various geographic scales on the CartoDB platform, with the former detailing exposure patterns at the town-level and the latter characterizing pollution spread at national level. Fig. 5(a) indicates that the southeast corner of Beijing suffers more PM<sub>2.5</sub> pollution, while as illustrated by Fig. 5(b), the city itself anchors the northern tip of the heavily polluted ‘urban quadrilateral’ in eastern China. In addition, online mapping platforms such as CartoDB allow for more flexible, interactive, and user-friendly access by the general public.

#### Concluding remarks

In this short report, we have identified several major technological, institutional, and socio-cultural challenges in assembling open urban datasets for Chinese cities, and showcased our attempt at understanding Chinese cities with open urban data. Our efforts represent a step toward creating and applying novel open data for understanding cities in China. Our initiative is first and foremost an online data clearinghouse for open urban data. Through the efforts of numerous data contributors, BCL has successfully marshalled and integrated a sizeable amount of open urban datasets over a relatively short time span. Through our showcase examples, we hope to demonstrate the usefulness of open urban data in (1) mapping urbanization in China with a finer spatiotemporal scale; (2) reflecting social and environmental dimensions of urbanization; and (3) visualizing urban China at multiple scales.

As a project in its infancy, the Beijing City Lab inevitably suffers a series of limitations, which in turn call for future research. Firstly, the Beijing City Lab, along with several other data sharing projects, marks the inception of a new wave of open urban data in China. However, the scope and amount of data being shared is still rather limited, and thus the data collection calls for participation from the wider research communities. More importantly, it matters how we link these data and models with the practices (including but not limited to planning practices, advanced urban governance, spatial information system open to the public) to improve urban spaces and the well-being of citizens.

Secondly, more empirical evidences and heuristic validations need to be collected before conclusions about the usefulness of open data in urban models can be drawn. This is intricately linked with the quality control, error monitoring, and cleaning of open data. Thirdly, the current organizational and legal frameworks of data sharing platform such as the Beijing City Lab are ambiguous, thus requiring careful treatment for issues such as data ownership

and user licenses. Finally, as our approach partly draws upon online data sources, such as user-centered, locational information, we also reflect critically how our method is situated in the broader neogeography and volunteered geographic information debates (Crampton, 2009; Zook & Graham, 2007). Future research is needed to critically examine the uneven geographies and power relations of online content as well as how this unevenness may result in transformations in both perceptual and physical aspects of places (Gurstein, 2011; Wilson & Graham, 2013).

#### Acknowledgment

The authors thank the editor and anonymous reviewers for their invaluable insights and suggestions. The authors are thankful for financial supports given by the National Natural Science Foundation of China [grant numbers 41401178 and 51408039] and for an MoE Grant for Humanities and Social Sciences [grant number 14YJC90132]. The usual disclaimer applies, and all errors remain the authors' own.

#### References

- Bai, X., Shi, P., & Liu, Y. (2014). Realizing China's urban dream. *Nature*, 509, 158–160.
- Batty, M. (2012). Smart cities, big data. *Environment and Planning – Part B*, 39(2), 191.
- Batty, M. (2013). *The new science of cities*. Cambridge, UK: The MIT Press.
- Crampton, J. W. (2009). Cartography: Maps 2.0. *Progress in Human Geography*, 33, 91–100.
- Crooks, A., Pfoser, D., Jenkins, A., Croitoru, A., Stefanidis, A., Smith, D., et al. (2014). Crowdsourcing urban form and function. *International Journal of Geographical Information Science* (in press). <http://dx.doi.org/10.1080/13658816.2014.977905>.
- Enserink, B., & Koppenjan, J. (2007). Public participation in China: Sustainable urbanization and governance. *Management of Environmental Quality: An International Journal*, 18, 459–474.
- Fox, P., & Hendler, J. (2011). Changing the equation on scientific data visualization. *Science*, 331, 705–708.
- Goetz, M., & Zipf, A. (2012). Towards defining a framework for the automatic derivation of 3D CityGML models from volunteered geographic information. *International Journal of 3-D Information Modeling (IJ3DIM)*, 1(2), 1–16.
- Goodchild, M. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69, 211–221.
- Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014). Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Annals of the Association of American Geographers*, 104, 746–764.
- Gurin, J. (2014). *Open data now: The secret to hot startups, smart investing, savvy marketing, and fast innovation*. New York, NY: McGraw-Hill.
- Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16(2).
- Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, 14(6), 1–4.
- Jia, T., & Jiang, B. (2010). Measuring urban sprawl based on massive street nodes and the novel concept of natural cities. arXiv preprint arXiv:1010.0541.
- Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1–14.
- Lang, T. (2011). Advancing global health research through digital technology and sharing data. *Science*, 331, 714–717.
- Liu, X. (2014). Featured graphic: Visualizing urban gastronomy in China. *Environment and Planning A*, 46, 1012–1013.
- Long, Y., & Liu, X. (2013). Featured graphic: How mixed is Beijing, China? A visual exploration of mixed land use. *Environment and Planning A*, 45, 2797–2798.
- Long, Y., Wu, K., Mao, Q. (2014). Simulating parcel-level urban expansion for all Chinese cities. arXiv preprint arXiv:1402.3718.
- Long, Y., Wu, K., Wang, J., & Shen, Z. (2014). Big models: From Beijing to the whole China. arXiv preprint arXiv:1406.6417.
- Mertha, A. (2005). China's “soft” centralization: Shifting tiao/kuai authority relations. *The China Quarterly*, 184, 791–810.
- Overpeck, J., Meehl, G., Bony, S., & Easterling, D. (2011). Climate data challenges in the 21st century. *Science*, 331, 700–702.
- Pacione, M. (2003). Urban environmental quality and human wellbeing—A social geographical perspective. *Landscape and Urban Planning*, 65, 19–30.
- Poldrack, R. A., & Gorgolewski, K. J. (2014). Making big data open: Data sharing in neuroimaging. *Nature Neuroscience*, 17(11), 1510–1517.
- Reichman, O., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science*, 331, 703–705.
- Sui, D. (2014). Opportunities and impediments for open GIS. *Transactions in GIS*, 18, 1–24.
- Waddell, P. (2014). GeoCanvas: A new platform for 3d urban informatics and visualization. In *Applied urban modelling conference 2014*, Cambridge, UK.
- Wallace, J. (2014). *Cities and stability: Urbanization, redistribution, and regime survival in China*. Oxford: Oxford University Press.



- Wang, N. (2011). China's chemists should avoid the Vanity Fair. *Nature*, 476, 253.
- Wilson, M. W., & Graham, M. (2013). Neogeography and volunteered geographic information: A conversation with Michael Goodchild and Andrew Turner. *Environment and Planning A*, 45, 10–18.
- Wu, K., Long, Y., Mao, Q., & Liu, X. (2015). Mushrooming Jiedaos, growing cities: An alternative perspective on urbanizing China. *Environment and Planning A*, 47, 1–2.
- Yue, Y., Lan, T., Yeh, A. G. O., & Li, Q. (2014). Zooming into individuals to understand the collective: A review of trajectory-based travel behaviour studies. *Travel Behaviour and Society*, 1, 69–78.
- Zook, M. A., & Graham, M. (2007). The creative reconstruction of the Internet: Google and the privatization of cyberspace and DigiPlace. *Geoforum*, 38, 1322–1343.