BIRKHÄUSER

DE-COD-ING THE CITY

Urbanism in the Age of Big Data

Dietmar Offenhuber, Carlo Ratti (Eds)

Dietmar Offenhuber, Carlo Ratti

Introduction

How to explain the paradox that urbanism, as a profession, has disappeared at the moment when urbanization is everywhere – after decades of constant acceleration – is on its way to establishing a definitive global "triumph" of the urban condition.
Rem Koolhaas (1995)

•

In his 1995 essay "Whatever Happened to Urbanism," Rem Koolhaas diagnosed urbanism as a failed discipline. In the light of rampant global urbanization, the profession has failed to give shape or even to influence the physical, social, and economic realities of cities. While during the nineties, the main challenge was dealing with quantity, now, almost twenty years later, the situation presents itself as more ambiguous. Following more than fifty years characterized by suburbanization and the erosion of urban centers, cities in the developed nations have taken two different paths: They either have returned as global economic players of unprecedented power, or have become shrinking cities – hollowed out by deindustrialization and demographic changes (Ryan 2012). Meanwhile, in the emerging economies of the developing world, urbanization is continuing at an undiminished rate. Globally, it is expected that 67 percent of the world population will live in cities by the year 2050 (United Nations 2012, 2).

For the discipline of urbanism, the struggle continues. Both of these trajectories present challenges in terms of infrastructure provision, housing, and socioeconomic development. But planners, policy experts, and economists are no longer the only specialists responding to these challenges. New actors enter the stage and bring new approaches to the field. Perhaps the most significant developments have happened in the domain of data-intense methodologies. The term *big data* refers to the availability of massive amounts of machine-readable information. This information is generated by the sociotechnical systems in which humans are increasingly entangled, by choice or by necessity: cell phone networks, credit card systems, or social media networks. Since the "digital exhaust" generated by these systems is so closely connected to our daily lives, it becomes a valuable resource for observing the processes and interactions of society at almost no cost. However, because of the massive quantity of these data sets, which were not created and structured with research purposes in mind, new methods are required for analyzing them.

Following the increasing availability of such digital data sources during the past ten years, the social sciences have taken a quantitative turn, often labeled *computational social science* (Lazer et al. 2009). In combination with large data sets, new computational methods allow researchers to address topics such as environmental perception, sentiment, or social connection, which were previously limited to qualitative modes of inquiry. Computational social science has brought together sociologists with physicists, mathematicians, and computer scientists, who have recently discovered cities as a topic that provides numerous intriguing research problems to work on. In particular, the emerging field of network science – the study of complex networks – has made a significant contribution to urban research literature (Börner, Sanyal, and Vespignani 2007). Network science stands for a shift from an exclusively spatial perspective on urban data to a topological perspective, focusing on relationships and interactions between people, places, and institutions at any scale. In this sense, network science actualizes ideas that scholars such as Manuel Castells have introduced into the urban studies literature (Castells 1996). By abstracting cities as spatial social networks of interaction, network science helped to uncover structural commonalities shared by almost all urban systems, allowing researchers to describe and predict how cities evolve and will grow over time (Batty 2013). In engineering, the field of urban informatics stands for the instrumentation of cities with sensor networks (Foth 2009). This includes the ubiquitous integration of technologies such as Global Positioning System (GPS) into everyday devices, which have enabled a real-time representation of urban conditions. Smart cities, both an academic and an engineering discipline, is advanced by systems theorists and companies such as IBM, Siemens, or Cisco. The concept of smart cities promises to improve the management of cities by making its infrastructures more adaptive – able to collect information about its own state and to regulate itself based on the state of the whole system. Finally, perhaps most fundamentally, the role of the citizen in the governance of cities has changed in important ways. The rise of social media led to new forms of participation and social activism. Beyond traditional forms of participation in planning projects, citizens voluntarily fulfill increasingly sophisticated roles in monitoring, management, and governance of the city and its infrastructure – a phenomenon that Eric Paulos called the rise of the "expert amateur" (Kuznetsov and Paulos 2010).

In the light of these developments, we believe that it is necessary to reexamine the state of urban planning, and explore, while avoiding the trap of "Big Data Hubris" (Lazer 2014), how these approaches can lead to a new understanding of the city.

While the discussed approaches are relatively recent, they are not without precedents. The history of urban planning has many examples of paradigm shifts initiated by new technology.

By the 1960s, cybernetics, the science of dynamic feedback systems, had started to leave a mark in the domain of urbanism, with both good and bad results. On the positive side, cybernetic models brought a fresh perspective for investigating urban systems. With its focus on dynamic states, feedback, and systemic processes, cybernetics brought new attention to the temporal and ephemeral, and was in many ways conflicting with high-modernist planning theory with its strict compartmentalization of functions. But cybernetic models have also led to some catastrophic failures, as in the case of the 1970s redesign of New York's fire system by the RAND corporation, which left poor neighborhoods underserviced, resulting in rampant fires and social unrest (Flood 2010). The mismatch between complexity of the problem and the inadequacy of the means is self-evident in the perhaps most ambitious cybernetic experiment: project Cybersyn, designed to control Chile's national economy under the Allende presidency (Pickering 2010, 258).

Apart from the obvious failure to capture the sociopolitical dimension of urban systems, cybernetics can also be challenged from another perspective: it simply does not lend itself to good design theory. It is apt for simulating adaptive, complex dynamic systems, but it provides little guidance for future alternatives. As Andrew Pickering suggests, cybernetics is performative rather than representational: it operates in a black box that adapts to current states, but it does not provide an abstract image of the world in its current or desired state (2010, 19). Nevertheless, the idea that data opens up new territories for urban planning and design has a long tradition that can be traced back to the meticulous mapping efforts by Giambattista Nolli in eighteenth-century Rome or Ildefons Cerdà in Barcelona. In *The Sciences of the Artificial,* Herbert Simon calls for a rigorous design science, a "body of intellectually tough, analytic, partly formalizable, partly empirical, teachable doctrine about the design process" (Simon 1996, 113). Such a design science would serve two functions: first to evaluate the performance of a given design, and second, to guide the identification of alternative scenarios. The role of design is the reconciliation between the "inner" world of physical objects and an "outer" world of its goals and functions. "The natural sciences are concerned with how things are. [...] Design, on the other hand, is concerned with how things ought to be." (Simon 1996, 114)

About This Book

This is a book about models for capturing these phenomena for understanding and improving cities. While big data can improve our understanding of urban systems, little attempt has been made so far to think about the consequences for design. This book represents a cross-sectional view of the research agenda practiced at the SENSEable City Lab at the Massachusetts Institute of Technology. Situated in the Department of Urban Studies and Planning, the Lab is an interdisciplinary institution exploring how real-time technologies can help us to better understand our cities, as well as conceiving possibilities how these technologies can improve our cities. The authors of the individual contributions are current and past researchers at the MIT SENSEable City Lab, as well as frequent collaborators. They come from a range of different backgrounds, including architecture and urban planning, sociology, political science, mathematics, computer science, physics, and visual design.

The contributions in this book address the generation of urban data, their representation and analysis, and finally their relevance for urban planning and design. To that effect, this volume is structured into three parts. The first part focuses on case studies discussing the origin of urban data, their collection and generation, including their inherent gaps and biases. The second part focuses on questions of representation, either as visual or mathematical models. The third part finally focuses on the implications for urban design.

In his opening contribution, Fabien Girardin explains the notion of digital footprints, data left by humans using digital services. Girardin distinguishes between passive footprints, generated without the user's awareness, and active footprints, deliberately created and shared by the users. He illustrates how user-generated data from photo-sharing websites can be used to investigate the travel behavior of tourists. Since an increasing number of photos voluntarily uploaded to such sites contain explicit geographic and temporal information (Geotags), photo-sharing sites allow insights into how different groups of people travel and which interests and values guide them. Michael Szell and Benedikt Groß focus on the possibilities of public data sets collected for accountability purposes. They work with a data set of 170 million taxi trips in New York City over one year, and acquired from the city government through a Freedom of Information Act request. Identifying vast redundancies in the system, the authors explore the possibility of an alternative, sharable urban taxi system. In the third contribution, Tony Vanky investigates the sometimes-ambiguous relationships of trust of people toward urban data. Currently, there are few attempts to measure the relevance of urban real-time data: how they affect the interaction with urban infrastructure. Using Singapore as an example, Vanky describes such measures: whether there is use and appreciation of real-time urban data, and how the information affects the spatial decisions of citizens at an individual level.

David Lee discusses the data collection methodology of participatory sensing (Burke et al. 2006), a way to actively involve volunteers in targeted data collection using location-based technologies. Using the example of the Trash Track project, Lee explores how the experience of participating in such a project in return changes the behavior and perceptions of the participants themselves, for example, whether the helping to investigate the fate of trash might change their attitudes toward waste management and recycling. Finally, Francisca M. Rojas maps the cultural geography of New York City through aggregated cell phone data, contrasting the data set with the official census information in terms of validity and quality. In her analysis of the telecom data, she maps out not only the global activities of New York's economic centers; but also the realities of immigrants and migrant workers who remain in constant connection with their home countries.

The second section of the volume, dedicated to visualization and modeling, is opened by Kristian Kloeckl, who documents an initiative that explores the development of an urban real-time data platform for Singapore, facilitating the collection, combination, and distribution of multiple data streams from urban networks. These "urban demos" provide concrete examples of how meaningful visual representations of data open up possibilities for stakeholders from different domains, facilitating a crosscutting discourse about urban issues. The visualization experts Pedro Cruz and Penousal Machado reflect in their article on the use of metaphors and figurative approaches in urban data visualization. They focus on a dilemma that makes geospatial visualization a difficult problem – the dilemma between the spatial nature of urban systems and abstract nature of data to be represented. The authors explore the relationship between the properties of the underlying data and the representational strategy; distinguish between visualization as a "photograph" or a "caricature" of information. The network scientists Philipp Hövel, Filippo Simini, Chaoming Song, and Albert-László Barabási are concerned with the observation, formalization and prediction of human mobility behavior based on telecom data sets. These data sets from cell phone network providers include implicit information about the spatial movement of cell phone subscribers, making it possible to answer questions such as: how predictable are we in our daily routines? In their mathematical analysis, the authors discover a surprising regularity in the way people travel, and provide a mathematical model for describing human mobility behavior. Kael Greco explores representational strategies for urban data based on a mobility research case study in the city of Riyadh, Saudi Arabia. The text approaches the complexity and nuance of urban data from two related, but antithetical perspectives: first, using spatial data to develop new modalities of seeing the city, and second, using the structure and composition of the city to provide new ways of seeing and understanding social data.

The third part opens with a contribution by the urban planner Andres Sevtsuk, who contrasts the plan as the traditional representational tool of urban design with the representation of the city as a network. Sevtsuk, a researcher in the emerging discipline of configurational studies, shows how structural measures of urban path networks, such as "betweenness" or "reach" offer powerful approaches for explaining attractiveness and locational quality within an urban system. Markus Schläpfer's work also situated in the domain of scaling studies addresses the issue of polycentricism of urban structure in relation to the travel behavior of tourists and city dwellers. Using telecom data sets, Schläpfer investigated the destinations and temporal rhythms of hundreds of thousand people in Singapore, Lisbon and Boston. The physicist and pioneer of the emerging research area of scaling studies, Luís M. A. Bettencourt explains the many ways in which the overall scale of an urban system determines a range of urban qualities and measures, such as the average number of personal contacts, the per-capita economic output and innovation, but also the prevalence of crime. His mathematical theory describes how cities change as they grow, and how these changes affect the lives of their citizens. Stanislav Sobolevsky's contribution focuses on the ramifications of human communication for how regions and their boundaries are defined. Using examples of countries such as Great Britain, France, and Belgium, the author shows how Manuel Castell's concept of the space of flows (Castells 1996) can be charted in geographic space using cell phone data.

How can these methods lead to a new design practice for cities? Koolhaas concludes: "If there is to be a 'new urbanism' it will not be based on the twin fantasies of order and omnipotence; it will be the staging of uncertainty; it will no longer be concerned with the arrangement of more or less permanent objects but with the irrigation of territories with potential [...] it will no longer be obsessed with the city but with the manipulation of infrastructure for endless intensifications and diversifications, shortcuts and redistributions – the reinvention of psychological space." (Koolhaas 1995, 31)

Computational models driven by data are a powerful way to incorporate uncertainty, irrigate potentials, and capture subjective and invisible qualities generally associated with psychological space. Data allows us to model the highly dynamic nature of cities, their social life, and their infrastructure networks at an unprecedented level of detail.

Privacy and Surveillance

You imagine, as does everybody else for that matter, that our organization has for many years been preparing the greatest document center ever conceived, an archive that will bring together and catalogue everything that is known about every person, animal and thing.
Italo Calvino (1995)

•

Privacy and digital surveillance remain central concerns that are tightly connected to the technical nature of digital systems. Often, privacy issues arise as unintended consequences of these technical properties. In 2010 it was revealed that the two major mobile operating systems provided by Apple and Google store and collect the location information of its users. The immediate purpose was not surveillance or targeted marketing: both operating systems relied heavily on information about known Wi-Fi spots and cell towers for establishing the user's location. Since the location of private Wi-Fi spots is hard to come by and changes frequently, both systems relied on users' phones to collect this information through an invisible service, by automatically mapping every hotspot users encountered. Privacy concerns arise not from this original purpose, but from its consequence – the existence of a vast, dynamic database containing detailed information about each user's behavior.

In the early days of digital media, privacy was mainly discussed in terms of data generation – who is allowed to collect data, and who should have access? In the recent decade, this discussion has shifted toward a discussion of control. This means that the user should own and be able to control any data that concern her life, including being able to trade that data in exchange for money or services. In the simplest form, this control can be implemented via opt-in and opt-out mechanisms. Research on digital traces also leads to insights about how privacy can be protected in the age of big data – what works, what doesn't work, and how mechanisms for privacy protection can be improved.

The question of personal privacy is intimately linked to questions of government transparency. Protection of privacy requires mechanisms for controlling that those rules are actually followed by companies and governments, which can be addressed by a rigorous opening of government data sources. Full transparency for the government, full privacy for the citizen is a frequent demand. In reality, however, those two realms cannot be neatly separated, as private citizens and governments interact in many different ways.

However, none of the approaches sketched above are effective against the "deep state," the domain of government secrecy. The two faces of big data, the civic and the threatening, are palpable in current US legislation. The country has both one of the oldest and most highly developed implementations of a Freedom of Information law, which provides mechanisms for the mandatory public access to government documents. At the same time, the government circumvents this law by entertaining a vast network of agencies operating under strict secrecy outside of public accountability mechanisms. Bringing these issues under democratic control through formal and informal measures requires a public discourse that is informed by a high level of data literacy, a differentiated knowledge of technology and digital data. We understand this book as a contribution to developing data literacy by giving insights into the nature and methods of data-driven technologies. Finally, there is also a historical and cultural dimension of personal privacy. Plain text is perhaps the most persistent and explicit digital expression of human speech and thought, requiring only minimal digital storage space. Without a natural expiry date, the question of data life cycles arises – should personal data, including embarrassing messages from one's youth, be deleted at some point, or made inaccessible? On a larger scale, how much culturally relevant information would be lost in such a case about our mostly digitally documented world?

As Calvino's story about the attempt to preserve the world's memory illustrates, the "total archive" would be an insurmountable restraint for all human agency, which in his story ultimately leads to death.

Nevertheless, on closer inspection, the totalizing idea of big data turns out to be a myth. The data distribution is highly unequal; data is least available precisely where we need it most, for example in the hinterlands beyond the digital divide. The contributions in this volume illustrate the immense value contained in the traces generated by our digital lives for better understanding our cities, our cultures, and our society. We consider the variety of public data sets available from governments, research institutions, infrastructures, and voluntary data provided by citizens a basis for civic discourse and ultimately an integral part of public space.

Batty, Michael. 2013. *The New Science of Cities.* Cambridge, MA: The MIT Press.

Börner, Katy, Soma Sanyal, and Alessandro Vespignani. 2007. "Network Science," *Annual Review of Information Science and Technology* 41, no. 1: 537–607.

Burke, Jeff, Deborah Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani Srivastava. 2006. "Participatory Sensing," in *World Sensor Web Workshop,* 1–5.

Calvino, Italo. 1995. "World Memory," in *Numbers in the Dark and Other Stories,* translated by Tim Parks. New York: Vintage, 135–136.

Castells, Manuel. 1996. *The Rise of the Network Society.* Chichester: Wiley-Blackwell.

Flood, Joe. 2010. *The Fires: How a Computer Formula, Big Ideas, and the Best of Intentions Burned Down New York City – and Determined the Future of Cities.* New York: Penguin.

Foth, Marcus, ed. 2009. *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City.* Hershey, PA: Information Science Reference.

Koolhaas, Rem. 1995. "Whatever Happened to Urbanism?" *Design Quarterly* no. 164: 28–31.

Kuznetsov, Stacey, and Eric Paulos. 2010. "Rise of the Expert Amateur: DIY Projects, Communities, and Cultures," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries,* edited by Ann Blandford et. al. New York: Association for Computing Machinery, 295–304.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, et al. 2009. "Computational Social Science," *Science* 323, no. 5915: 721–723.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis," *Science* 343, no. 6176: 1203–1205.

Pickering, Andrew. 2010. *The Cybernetic Brain: Sketches of Another Future.* Chicago: University of Chicago Press.

Ryan, Brent D. 2012. *Design After Decline: How America Rebuilds Shrinking Cities.* Philadelphia: University of Pennsylvania Press.

Simon, Herbert A. 1996. *The Sciences of the Artificial,* 3rd ed. Cambridge, MA: The MIT Press.

United Nations. 2012."World Urbanization Prospects: The 2011 Revision." ESA/P/WP/224. New York: United Nations Publications.