I/S: A JOURNAL OF LAW AND POLICY FOR THE INFORMATION SOCIETY

Space-time Data Science for a Speedy World

HARVEY J. MILLER*

Abstract: Space-adjusting technologies such as transportation and information/communication technologies are accelerating our world in complex ways. A speedy world has benefits but also challenges attempts to make it more sustainable and resilient. Our capabilities for observing human dynamics have improved dramatically, but less well developed are capabilities for extracting relevant space-time knowledge and making decisions while that knowledge is still fresh. This paper reviews the challenges and issues involved in developing space-time data science to deliver actionable knowledge quickly in a speedy world.

I. INTRODUCTION

The world has become speedy. Transportation and information/communication technologies have collapsed the time and cost required for mobility and interaction. This has generated levels of travel and communication that would have seemed astonishing a generation or two ago. While there are benefits to individuals and societies, more people conducting more activities in more places at more times is creating faster dynamics with higher complexity, greater resource consumption and the potential for failures among subsystems that move at different speeds (such as human and physical systems).

Mobility and connectivity are also generating massive amounts of data that can help manage the faster dynamics they generate. However, it remains unclear how to deliver appropriate space-time knowledge from these torrents of space-time (time-stamped and geo-referenced) data. Data are flowing at faster speeds from the environment, but our

* Department of Geography, The Ohio State University, 1036 Derby Hall, 154 North Oval Mall, Columbus, OH 43210-1361.

capabilities for extracting and delivering knowledge remain slow. Our knowledge is stale before we can act, limiting our ability to manage the complexity of a faster world.

This paper discusses issues surrounding space-time data science for a speedy world. The next section discusses the role of space-adjusting technologies in creating a smaller and speedy world; as we will see, these effects are non-uniform and intricate. The subsequent section discusses the role of space-time data on human activities in the social sciences and planning. Following this is a review of the challenges of extracting knowledge from the new data flowing from people, cities and regions. The following section focuses on a specific question: What type of knowledge should be extracted and delivered to act quickly in a speedy world? Some brief comments conclude the paper.

II. SPACE-ADJUSTING TECHNOLOGIES

Space-adjusting technologies are techniques, systems and methods of organization that alter the nature of experienced space with respect to the time, cost and effort to overcome in movement or communication. Consequently, space-adjusting technologies redistribute human activities in geographic space and time (Abler 1971, 1975).

Space-adjusting technologies include transportation systems such as horse-drawn vehicles, ships, canals, ocean-going vessels, railroads, bicycles, automobiles and commercial aviation. These technologies and systems were revolutions in the way people interacted with geographic space, and consequently where people located key activities such as home, work and markets. Horse-drawn vehicles allowed people to carry greater weight over longer distances. Ships expanded mobility along rivers and coasts. Canals expanded the navigability of river systems, railroads opened the interior of continents away from hydrologic features, bicycles gave mobility to the masses, automobiles expanded personal mobility and commercial aviation allows global scale mobility within hours rather than months. These revolutions led to shifts in the locations and shape of settlement and commerce, leading to our current global urban landscape.

Space adjusting technologies also include information/communication technologies (ICTs) such as the telegraph, telephone, computers, the Internet, wireless communication and mobile telephony. The telegraph was magic: for the first time, a message could be transmitted faster than a person could move. This helped facilitate the growth of downtown business districts and the rise of global business. The telephone gave this magic to the masses in homes, computers and the Internet expanded the spectrum of information accessibility, wireless communication untethered people and computers from cables, and mobile telephony facilitates contacting people directly rather than calling a place (and hoping the person is there).

The relationships between transportation-based mobility and ICT-based communication are complex. It is not necessarily one of simple *substitution*: more ICT-based communication means less transportation-based mobility. ICTs can also *complement* transportation; for example, the ability to email colleagues can lead to more meetings. Similarly, capabilities for accessing online reviews of restaurants and travel destinations can lead to more travel to those locations. Transportation and ICTs can also *modify* activities by changing the location and timing of activities without a net increase or decrease in frequency (Mokhtarian 1990). In addition to these direct effects there are indirect effects through growth of new economic sectors such as e-commerce and applications such as teleconferencing that generate new mobility demands for both people and objects (Bannister and Stead 2004; Miller 2007).

III. The World is Shrinking, Shriveling and Fragmenting

The rise of mobility and communication technologies has created *space-time convergence*: the world as experienced has become smaller (Janelle 1969). For example, the time distance between Portland, Maine and San Diego, California, has shrunk from 2 years on foot in the 16th century CE, 8 months on horseback in the 17th century, 4 months by stagecoach in the 19th century, 4 days by rail in the early 20th century, to 5 hours by airplane in the late 20th/early 21st centuries. If we use the 16th century as a base with a walking speed of 4.8 km per hour, San Diego has 'moved' to a location only 24 km from Portland, with the apparent distance between the cities shrinking over 4023 km in 400 years (Lowe and Moryadas 1975).

ICTs have also generated space-time convergence in a similar manner to transportation technologies. Abler (1971) notes that a trans-continental phone call in the United States required fourteen minutes to establish in the 1920s, one minute in the 1950s and thirty seconds in the 1970s, implying a convergence rate of 16.2 seconds per year (Lowe and Moryadas 1975). In the 21st century, an email can be delivered across the globe in well under one second, and Michael Lewis' latest book *Flash Boys* (Lewis 2014) describes high frequency traders who grow rich by shaving *milliseconds* from exchanges in asset markets (and I can start reading his book in under a minute on my Kindle!).

Space-time convergence is not uniform. As Waldo Tobler argues, the world is also *shriveling*: absolute transport costs are decreasing but relative differences in transportation costs – especially time – are increasing (Miller 2007). In the pre-modern era everyone traveled at the slow rates of speed generated by muscle (human or animal) or wind, although at varying levels of comfort. In the contemporary world, wealth buys speed. If you live in Europe and North America, you can travel to major cities in the world within hours. However, travel to a major city can still take days in parts of Africa, the Amazon Basin and the interior of Asia (World Bank 2009). Even in affluent parts of the world, some enjoy travel via private jets while others suffer flight delays and cancellations at dreary airports. Within cities, those who can afford premium housing have easier access to jobs and entertainment while others are relegated to traffic and long commutes. In many U.S. cities, the resources and capabilities to own and operate a private automobile provide mobility well beyond those who rely on public transport.

ICTs are not just shrinking and shriveling: they are *fragmenting* the world (Couclelis 2000). ICTs allow activities to be disconnected from specific times and places. For many people, there is no longer a specific location to perform work: for example, I am typing this sentence while waiting for a flight in Arlanda airport in Sweden, 6800 km from my so-called workplace. Of course, this is not universally true: the airport employee who served my latte this morning had to travel to a specific place to work, as did the gate agent who will let me on my flight and the crew that will fly the plane and provide inflight services. But even these jobs may be remotely sourced in the future: McDonald's has experimented with centralized call centers that process drive-through orders from restaurants thousands of miles away (Richtel 2006). Unmanned aerial vehicles (drones) can deliver air passengers as easily as they deliver missiles or packages from Amazon (Riberio 2013). Beyond work, other activities such as shopping, entertainment and socializing have disconnected from specific locations and times: one may still choose to travel to a store, theater or nightclub, and many still do, but this is not required.

IV. THE WORLD IS SPEEDING UP

A shrinking, shriveling and fragmenting world has profound impacts. Mobility has exploded at all geographic scales, from local to global, to levels that would have astonished previous generations. Today, the average citizen of The Netherlands travels roughly 40km per day; this was the average *per year* at the turn of the 18th century (Bertolini and Dijst 2003). There is evidence that we may be reaching peak travel – at least in the Global North – but the peak is pretty high: in one year, the United States experiences nearly 4 trillion passenger miles and 1.3 trillion motor carrier ton miles from 250 million vehicles, and 550 billion air passenger miles between major airports (TRB Executive Committee 2013). The demand for mobility – particularly individual vehicles – continues to grow in sub-Saharan Africa, China and southeast Asia: we may be facing a world with two billion cars by 2030 (Sperling and Gordon 2009).

ICT growth is also gob-smacking. Here we go: the first email was sent on ARPANET in 1971. By 2001, global email users sent 31 billion messages per day, a rate of 22 million per minute (Bryant 2011). By 2012, global email users sent 204 million messages per minute; a tenfold increase. Also during that minute - Google received over 2 million search queries, Twitter users sent over 100,000 tweets, Facebook users shared 685,000 content items and YouTube users uploaded 48 hours of new video (Tepper 2012).

Hyper-mobility and hyper-communication are changing the nature of activity organization. People conduct more activities in more places and times than would have been imaginable a generation or two ago. Activity intensification may be increasing *social metabolism*: rates of resource consumption and waste output from cities, regions and societies. Making faster systems more efficient will not necessarily make them more sustainable. Higher efficiency can increase rather than decrease resource consumption by lowering costs and consequently increasing demand: an effect known as *Jevon's paradox* from its 19th century discoverer, the British economist William Stanley Jevons (Alcott 2005). Jevons paradox is certainly apparent from the explosion of mobility and communication over the past two generations (Miller 2013). Although human systems have sped up, physical systems move at their same pace. This increases the possibility of *shearing forces*: cascading and amplifying failures generated when disruptions occur in systems with components operating at different speeds (Zolli and Healy 2012).

V. SPACE-TIME DATA AND THE SOCIAL SCIENCES

Mobility, communication and data are best friends forever. Mobility and communication are major consumers of data, e.g., applications such as travel maps, phone directories and online review sites. Mobility and communication also generate streams of data as byproducts – their *data exhaust* (Williams 2013). These data are exhaust but not pollution: they have vital roles to play in managing the speedy and complex world they are enabling.

There are now a wide range of *geospatial technologies* that can capture data on mobility, communication and coupled social and environmental systems such as cities, societies and coupled biological and physical systems. *Location-aware technologies* (LATs) are technologies that can frequently report their location in geographic space. LATs are generally associated with mobile entities: they include the global positioning system (GPS), radiolocation and dead-reckoning techniques coupled with computers and tablets on the desks and laps of humans, carried in their pockets and transported by vehicles. *Geosensor networks* are wirelessly communicating, sensor-enabled, small computing devices distributed in geography to enable in-situ monitoring of dynamic properties such as change and movement (Duckham 2013). *Remote sensing* devices are passive and active sensors carried on aircraft and satellites for environmental monitoring over urban and regional scales suing both passive (reflected light) and active (laser) methods. Helping to manage all these data are *geographic information systems* (GIS), spatial database management systems and moving objects databases: all have seen remarkable growth in their capabilities to handle dynamic geographic phenomena.

The privacy and ethical issues in a sensor-saturated world are profound and urgent. They are an active area of investigation in science and policy research, generating concepts such as *locational privacy* (Beresford and Stajano 2003). While people are justified to have qualms about living in a Panopticon, these data can revolutionize social sciences such as geography, economics, urban planning, sociology, environmental science and public health.[1] The theories and models of previous generations were developed in a computation-scarce and data-poor world: these are static and aggregate sketches of things that are individualistic and dynamic – and much more interesting - in the real world.

The social sciences have rich concepts for nearly a century to describe human activities in space and time. Most notable is *time geography*: a conceptual theory and notation system for representing the basic necessary conditions for human mobility, communication and activities in space with respect to time, and the multilevel systems that emerge from interlinked individual dynamics (Hägerstrand 1970). In recent years, advances in the geospatial technologies described above have enabled analytical and computational methods in time geography that are suitable for real world applications using high resolution space-time data (Ellegård and Svedin 2012; Miller, in press). In a wide range of fields such as transportation, urban planning, environmental science and public health, it is now possible to think about human systems as constituted by people rather than abstract stocks and flow. Outcomes such as traffic jams, social trends and epidemics emerge from the intricate interactions; they are not manufactured from scratch by the modeler (Flake 1998).

[1] See for example, Batty (2012), Kitchin (2014), Mayer-Schonberger and Cukier (2013), Miller (2010), Murdoch and Detsky (2013).

VI. GENERATING KNOWLEDGE FROM SPACE-TIME DATA

While lots of data and computations are changing the social sciences, the increasing flow of spatio-temporal data from LATs and sensors is overwhelming. It is not trivial to generate knowledge from the torrents of data flowing from objects and processes in the real world. The challenges associated with so-called "big data" - data that exceeds our capabilities to analyze – have at least three major facets. *Volume* is the amount of data that can be processed: we have much more data than can be processed using traditional analytical methods and computing technologies. *Variety* is the expansion of data sources beyond traditional formats and sources. *Velocity* is the speed at which data flows from the real world (Dumbill 2012).

Data volume outstripping capacity to analyze is not completely new: an early data challenge occurred in the 1890s when the U.S. government realized it would take more than a decade to process census data by hand, rendering the effort useless. This led to the creation of automated tabulating technologies and a company now known as IBM. In the 1970s, the ability to download remotely sensed data from the satellite-based LANDSAT system exceeded analysis capabilities at that time (Miller and Goodchild, 2014). Massive spatio-temporal data volume is nevertheless a crucial challenge, leading to research frontiers such as *cyberGIS*: high-performance, distributed and collaborative processing for spatio-temporal data (Wang et al. 2013).

Processing spatio-temporal data is only part of the story: we must also be able to describe it in a language that humans can understand; in other words, generate knowledge. However, most of the techniques and processes we call science were developed for a data-scarce world. Massive databases may contain hidden knowledge that is difficult to uncover using the painstaking, deliberate methods of traditional science. How do we extract hidden space-time knowledge from these data?

One option is to use it during the initial discovery phase of the scientific process when the researcher is formulating new theories and hypotheses. This is the motivation behind techniques collectively known as *knowledge discovery from databases* or *data mining* (Han and Kamber 2012) and specialized siblings such as *geographic knowledge discovery* (Miller and Han 2009), *spatiotemporal knowledge discovery* (Roddick and Lees 2009) and *mobility mining* (Giannotti and Pedreschi 2008). Massive databases and knowledge discovery techniques are similar to the telescope and microscope: they allow scientists to see things they could not see before, leading to new hypotheses that to be tested using traditional confirmatory methods such as experiments, inferential statistics and analytical modeling (Gahegan 2009; Miller 2010).

A second strategy for analyzing massive data is *data-driven modeling*. In traditional modeling, the researcher develops a conceptual theory, formalizes the conceptual theory using logic and mathematics, generates predictions using rule-based manipulations, and tests these predictions against real world data. But this is too painstaking to explore the wide range of possible models in a large database. Also, the scientist can only postulate models she can imagine; hidden knowledge is difficult to imagine. Data-driven modeling inverts the modeling process: scalable techniques such as machine learning and genetic algorithms automatically (with varying degree of human supervision) grow models from data using information search and statistical tests. A pioneering example of data-driven spatial modeling is Stan Openshaw's automated system for exploring a universe of possible spatial interaction (or "gravity") models for describing flows among locations based on their site-specific properties and distances (see Fotheringham and O'Kelly 1989). This early attempt presciently highlighted a weakness: the models that emerged were incomprehensible and cannot be summarized other than pointing to them. But, if explanations are incompressible – they cannot be described as narratives or summaries - are they explanations (Miller and Goodchild 2014)?

Google FluTrends is another example of the foibles of data-driven modeling. It predicted flu very well - at first. But it was overfitted to data from a process (search queries) that changed due to an artifact (algorithm tweaking by engineers). It was also caught in an echo chamber as Google FluTrends received media attention and generated more queries. Avoiding getting tripped-up by what Lazer et al. (2014) call *data hubris* requires understanding the processes that generate the data and not ignoring existing knowledge and traditional data.

The wider variety of data from non-traditional sources using non-traditional formats is also challenging. Science and scientific management involve carefully sampled and precisely measured data that can be stored as numbers in tables and relations. New sources of data are self-selected with unsure representativeness – e.g., not everyone can afford a smartphone and data plan. Also, data is often shared consciously, meaning that it also may not be representative of behavior. Much of these data are unstructured and difficult to quantify, such as text, audio and video (Miller and Goodchild 2014).

Geographic space and time are powerful ways to integrate a wide range of authoritative (official) and structured data with naturalistic and unstructured data. Spatial and temporal referencing allows access to the time-honored concepts and techniques of *cartography* (map design and production), a technology that has served humanity well for many millennia. It also leverages contemporary techniques such as *geovisualization* (exploratory analysis of large georeferenced datasets using the map as a central metaphor). We can georeference unstructured data such as text, sound and video using the location where the data were collected, or based on their content.

Going beyond structured and authoritative data to the naturalistic data generated through daily life is a game changer in the social sciences. Ubiquitous, ongoing space-time data flows allow the researcher to capture spatio-temporal dynamics directly (rather than inferring them from snapshots every month, year or decade) and at multiple geographic and time scales. The data are collected on an ongoing basis, meaning that both mundane and unplanned events can be captured. We also do need to seek improbable but consequential or "black swan" events (Taleb 2007): we can measure all events and determine later which are consequential. We can also discover how seemingly inconsequential or obvious events can combine to form major and surprising consequences (Watts 2011). These data greatly expand the social sciences' ability to witness natural experiments in the real world instead of relying on artificial experiments or statistical controls (Miller and Goodchild 2014). The increasing speed at which data flows from the real world challenges our ability to understand and make decisions based on that data. Our world is speeding up but our analytical and decision processes remain slow. This was not a problem in a slower world where models provided only broad-brush generalities. But a speedier world requires richer descriptions of complex and intricate behaviors. The potential for shearing forces between the faster pace of human dynamics in the 21st century and our slower ability to understand and act on those dynamics are increasing.

VII. DELIVERING SPATIO-TEMPORAL KNOWLEDGE QUICKLY

There are technical and conceptual dimensions to delivering spatio-temporal knowledge quickly enough to make a difference in a speedy world. The technical dimension involves detecting patterns quickly enough. *Real-time stream processing* involves fast processing of high-volume data streams before spending time to store the data in a database. This requires capabilities to query directly from data streams, gracefully handle data stream imperfections such as delayed, missing or out-of-order data, and the high availability and rapid fusion of streaming and previously stored data (Stonebraker, Çetintemel and Zdonik 2005). A challenge is *space-time data fusion* (integrating data in real-time based on its spatial reference, time-stamp, semantics, or some combination). Syntactic differences concerning rules about how the world or its measures are constructed is easy to resolve; more difficult are semantic differences regarding what the data means in the real world. A vexing property of human concepts is ambiguity and fluidity. For example, what is a "community" or a "neighborhood"? What does it mean to be "poor" or "middle-class"? Even administrative data can have semantic heterogeneity across regions and countries, and even within an organization or agency (Goodchild and Li 2012; Miller and Goodchild 2014).

The conceptual challenge to delivery spatio-temporal knowledge in a timely manner relates to the types of knowledge we can generate quickly and the decisions we can make before it is stale. Traditionally, causality is king in science: we want to know *why* something happens. But massive amounts of streaming data favor correlations over causality since the former can be derived quickly and easily while the later requires deliberate theorizing and testing. Some have argued that causality is irrelevant in the new data-rich environment: science can advance without coherent models or unified explanations since reality is much richer than we can imagine (Anderson 2008). However, as discussed above, thus far data-driven modeling has not been promising with respect to generating useful knowledge.

The way forward may not be the simple choice between painstaking knowledge versus quick but vapid correlations. Duncan Watts argues that, although for first time in history we can observe in high fidelity the real-time behavior of large groups and societies, the type of knowledge we can derive from these data may be more modest than the grand explanation attempted in earlier eras (Watts 2011). In the absence of data and computations, explanation relied on the scientific intuition to derive simple and tractable models, typically by assuming rational behavior by individuals and collectives. However, intuition – especially about humans - is often wrong: behavior is diverse, fluid and often driven by emotions rather than calculations. Collective social behavior also cannot be derived from individual actions: human systems are complex, and social phenomena emerge from the interactions of individuals in surprising ways (Flake 1998). Applying generic and rigid models of cause and effect to complex human systems is doomed to failure and disillusionment: these systems are not predictable or controllable even in principle.

Instead of a quixotic quest for grand explanations, social scientists should use new data to derive richer but more circumscribed explanations of particular social situations and solutions for specific social problems. These empirically-grounded and context-specific explanations can be obtained more quickly from data than grand explanations. Although more modest, they can be effective. They can be used to derive hypotheses for more deliberate investigation and can serve as stepping stones to more general theories and laws derived at a more deliberate pace (Watts 2011). Low-hanging fruit are still fruit, and space-time data flowing from environments are generating orchards where there used to be barren scrub.

Richer but circumscribed explanations of human phenomena are useful for the types of policy and planning interventions we can conduct quickly with fast flowing data in a speedy world. Rather than trying to predict the unpredictable and controlling the uncontrollable, Watts (2011) suggests *measuring and reacting*. Instead of large interventions that require specific responses, measure directly how people react to a wide range of possibilities and reacting accordingly. This involves naturalistic observations since there are many possibilities in the real world and we can now see more. It also includes observing quasi-experiments generated by events that happen naturally (such as a neighborhood becoming trendy or blighted) or deliberately (such as a concert in the center city or the construction of a new light rail line). It may also include innocent experiments repeated over large populations (such as varying the design of an online transit map)[2] Measuring and reacting is also a more humble approach to managing human dynamics than prediction and control. It involves crowd-sourcing and boot-strapping: harnessing local knowledge, searching for existing but partial solutions and propagating best practices, allowing humans to self-organize solutions nurtured by knowledge.

As Mike Batty points out with respect to cities, the scientific and engineering philosophy that dominated policy and planning in the 19th and 20th centuries emphasized dramatic and radical changes over the long-run with little concern for the small movements and mundane behaviors that maintain cities (Batty 2012). Space-time data flowing from people, cities and societies allow planners to rediscover these small but consequential spaces and activities. It also may suggest nuanced interventions that nudge human dynamics towards more equitable, resilient and sustainable outcomes.

VIII. CONCLUSION

We are accelerating the world through our space-adjusting technologies, and we are scrambling to understand and manage these dynamics. The rich space-time data flowing from these activities at faster speeds provides capabilities for shaping human dynamics towards more sustainable and resilient outcomes. We must extract space-time knowledge from these data while it is fresh and relevant for the types of planning and policy interventions that will be successful in a speedy world. There are challenges to shortening the pipeline from data to decisions, some of which involve recognizing our limited ability to understand, predict and control human dynamics. But accepting mystery may bring openness to new discoveries. It also requires trusting

[2] However, the boundary between innocent and not-so-innocent experiments is not always clear; an example is the recent controversy over Facebook's manipulation of users' feeds to gauge reaction to positive versus negative posts (Vertesi 2014).

humans to help manage their own dynamics through open data and transparent processes that invite collaboration (Gurin 2013).

REFERENCES

Abler, R. (1971). Distance, Intercommunications and Geography. Proceedings of the Association of American Geographers 3, 1– 4.

Abler, R. (1975). "Effects of Space-Adjusting Technologies on the Human Geography of the Future," in: Abler, R., Janelle, D. and Philbrick, A. (eds) Human Geography in a Shrinking World. North Scituate, MA: Duxbury Press, 35–56.

Alcott, B., (2005) "Jevons' Paradox," *Ecological Economics* 54, 9–21.

Anderson, C. (2008) "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired*, June 23.

Bannister, D., and Stead, D. (2004) "Impact of Information and Communications Technology on Transport," *Transport Reviews,* 24, 611–632.

Batty, M. (2012) "Smart Cities, Big Data," *Environment and Planning B*, 39, 191–193.

Beresford, A. R. and Stajano, F. (2003) "Location Privacy in Pervasive Computing," *Pervasive Computing*, 2, 46–55.

Bertolini, L., and Dijst, M. (2003) "Mobility Environments and Network Cities," *Journal of Urban Design* 8, 27–43.

Bryant, M. (2011) "The First Email was sent 40 Years ago this Month," *The Next Web*, October 8 2011; http://thenextweb.com/insider/2011/10/08/the-first-email-was-sent-40-years-ago-this-month.

Couclelis, H. (2000). "From Sustainable Transportation to Sustainable Accessibility: Can We Avoid a New Tragedy of the Commons?" in Janelle, D. G. and Hodge, D. C. (eds) *Information, Place and Cyberspace: Issues in Accessibility*. Berlin, Germany: Springer, 341–356.

Duckham, M. (2013) *Decentralized Spatial Computing: Foundations of Geosensor Networks*, Berlin: Springer.

Dumbill, E. (2012) "What is Big Data? An Introduction to the Big Data Landscape," http://strata.oreilly.com/2012/01/what-is-big-data.html.

Ellegård, K. and Svedin, U. (2012) "Torsten Hägerstrand's Time-Geography as the Cradle of the Activity Approach in Transport Geography." *Journal of Transport Geography* 23: 17–25.

Flake, G. W. (1998) *The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems, and Adaptation*, Cambridge, MA: MIT Press.

Fotherigham, A. S. and O'Kelly, M. E. (1989) *Spatial Interaction Models: Formulations and Applications*, Kluwer.

Gahegan, M. (2009) "Visual Exploration and Explanation in Geography: Analysis with Light," in H. J. Miller and J. Han (eds.) *Geographic Data Mining and Knowledge Discovery*, second edition, Taylor and Francis, 291-324.

Giannotti, F. and Pedreschi, D. (2008) *Mobility, Data Mining and Privacy*, Springer.

Goodchild, M. F. and Li, L. (2012) "Assuring the Quality of Volunteered Geographic Information," *Spatial Statistics* 1, 110–120.

Gurin, J. (2013) *Open Data Now: The Secret to Hot Startups, Smart Investing, Savvy Marketing, and Fast Innovation,* McGraw-Hill.

Hägerstrand, T. (1970). "What about People in Regional Science?" *Papers of the Regional Science Association* 24: 1–12.

Han, J.. and Kamber, J. (2012) *Data Mining: Concepts and Techniques*, 3rd edition, Elsevier Inc.

Janelle, D. G. (1969) "Spatial Organization: A Model and Concept," *Annals of the Association of American Geographers* 59, 348–364.

Kitchin, R. (2014) "Big Data and Human Geography: Opportunities, Challenges and Risks," *Dialogues in Human Geography*, 3, 262-267.

Lazer, D., Kennedy, R., King, G. and Vespignani, A. (2014) "The Parable of Google Flu: Traps in Big Data analysis," *Science*, 343, 1203-1205.

Lewis, M. (2014) *Flash Boys: A Wall Street Revolt*, W. W. Norton & Company.

Lowe, J. C., and Moryadas, S. (1975). *The Geography of Movement*. Prospect Heights, IL: Waveland Press.

Mayer-Schonberger, V., Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*.

Miller, H. J. (2007) "Place-based versus People-based Geographic Information Science," *Geography Compass*, 1, 503-535.

Miller, H. J. (2010) "The Data Avalanche is Here. Shouldn't we be Digging?" *Journal of Regional Science*, 50, 181-201.

Miller, H. J. (2013) "Beyond Sharing: Cultivating Cooperative Transportation Systems through Geographic Information Science," *Journal of Transport Geography*, 31, 296-308.

Miller, H. J. (in press) "Time Geography and Space-time Prism," *Wiley-AAG International Encyclopedia of Geography*.

Miller, H. J. and Goodchild, M. F. (2014) "Data-driven Geography," *GeoJournal*. DOI 10.1007/s10708-014-9602-6

Miller, H. J. and Han, J. (2009) *Geographic Data Mining and Knowledge Discovery*, second edition, CRC Press.

Mokhtarian, P. L. (1990) "A Typology of the Relationships between Telecommunication and Transportation," *Transportation Research A*, 24A, 231–242.

Murdoch, T. B. and Detsky, A. S. (2013) "The Inevitable Application of Big Data to Health Care," *Journal of the American Medical Association*, 309, 1351-1352.

Openshaw, S. (1988) "Building an Automated Modeling System to Explore a Universe of Spatial Interaction Models," *Geographical Analysis*, 20, 31-46.

Riberio, J. (2013) "Amazon Plans to use Aerial Drones to Deliver Packages via a new Prime Air Service," *TechHive*, December 1; http://www.techhive.com /article/2068040 /amazon-plans-to-use-drones-to-deliver-packages.html.

Richtel, M. (2006) "The Long-Distance Journey of a Fast-food Order," *New York Times*, April 11.

Roddick, J. F. and Lees, B. G. (2009) "Spatio-Temporal Data Mining Paradigms and Methodologies," in H. J. Miller and J. Han (eds.) *Geographic Knowledge Discovery*, second edition, CRC Press, 27-44.

Sperling, D. and Gordon, D. (2009) *Two Billion Cars: Driving Towards Sustainability*. New York: Oxford University Press.

Stonebraker, M., Çetintemel, U. and Zdonik, S. (2005) "The 8 Requirements of Real-Time Stream Processing," *SIGMOD Record,* 34: 4, 42-47.

Taleb, N. N. (2007) *The Black Swan: The Impact of the Highly Improbable*, Random House

Tepper, A. (2012) "How Much Data is Created Every Minute?" *Mashable*, June 22, http://mashable.com/2012/06/22/data-created-every-minute.

TRB Executive Committee (2013) *Critical Issues in Transportation 2013*, Washington DC: Transportation Research Board, National Research Council, The National Academies.

Vertesi, J. (2014) “The Real Reason You should be Worried about that Facebook Experiment,” *Time*, http://ti.me/1iXMVUx.

Wang, S., Anselin, L., Bhaduri, B., Crosby, C., Goodchild, M.F., Liu, Y. and Nyerges, T. L. (2013) “CyberGIS Software: a Synthetic Review and Integration Roadmap,” *International Journal of Geographical Information Science*, 27, 2122–2145.

Watts, D. J. (2011) *Everything is Obvious – Once You Know the Answer*, Crown Business.

Williams, A. (2013) “The Power of Data Exhaust,” http://techcrunch.com/ 2013/05/26/ the-power-of-data-exhaust.

World Bank (2009) *World Development Report 2009: Reshaping Economic Geography*, Washington DC: The International Bank for Reconstruction and Development / The World Bank. http://go.worldbank.org/FAV9CBBG80.

Zolli, A. and Healy, A. M. (2012) *Resilience: Why Things Bounce Back*, New York: Free Press.