Modernizing government's approach to transportation + land use data

Challenges and opportunities

Adie Tomer and Ranjitha Shivaram

The Brookings Institution | Metropolitan Policy Program | JULY 2017

Executive Summary

Data is already a fundamental part of how transportation and land use professionals plan and build places. These professionals map infrastructure assets, zoning plans and related socioeconomic data to better understand the supply side of transportation and land use in their communities. The federal government offers overarching support via public data sets, database standards, and collection mandates. Entire ecosystems of private data providers and software publishers support public sector capacity. But, on the transportation demand side, infrastructure performance measures and traveler surveys provide limited or often outdated measures.

Today, this approach to data is poised to change. The rapid penetration of connected devices into everyday life provides geospatial movement data at a scale never seen before. Smartphones, vehicle telematics, fitness trackers, credit card transactions, and online map searches offer real-time and highly detailed data points about people's transportation demands. Digitized records, scalable processing power, and falling storage prices facilitate data sharing and production of new analytics. Combined, such emerging data assets introduce new opportunities to plan and build communities around shared economic, social, and environmental objectives.

However, this transition will not be easy. Using a mix of primary research and interviews, this paper outlines the challenges to integrating new data and techniques into current governance frameworks. Limited staff capacity and restricted budgets can slow adoption. Procurement policies are stuck in an analog era. Privacy concerns introduce risk and uncertainty. Private data could be invalid or unavailable to public consumers. And even if governments could acquire all of the new data and analytics that interest them, their planning and investment models must be updated to use the new resources.

Addressing these challenges will be a difficult but crucial step in modernizing how the public sector approaches urban transportation and land development.

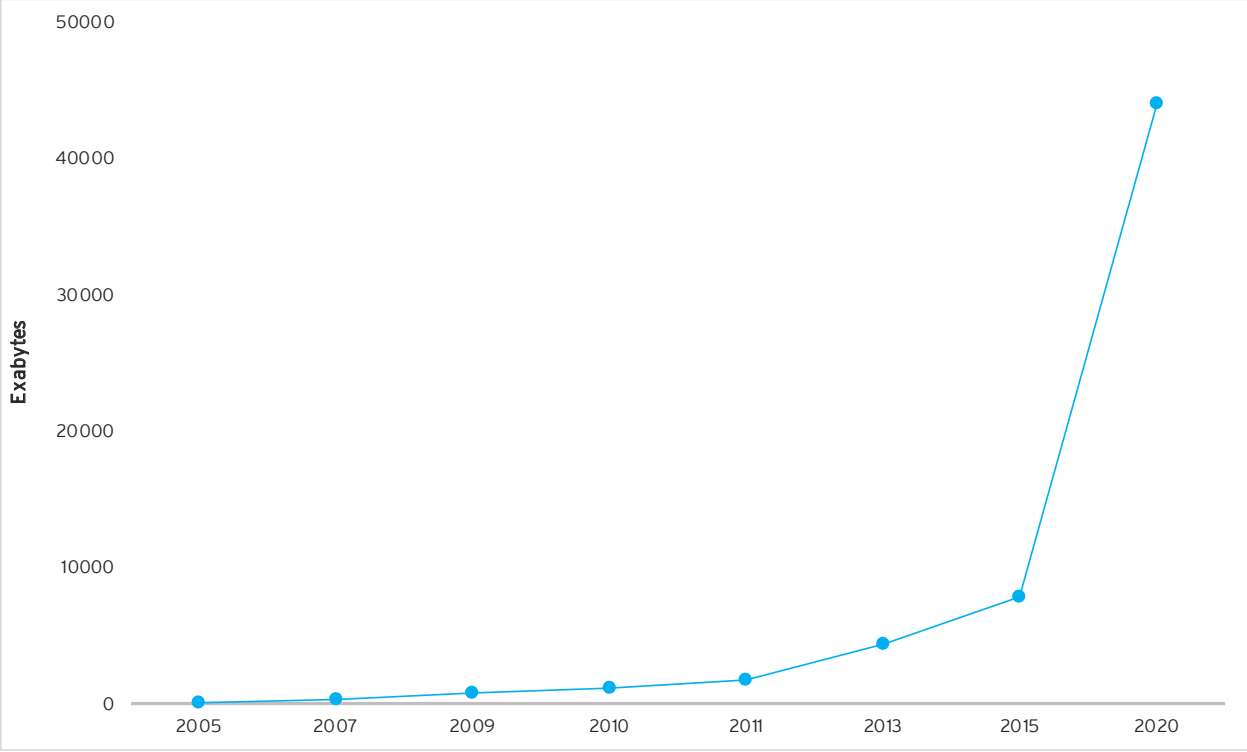
1 Introduction

The rapid transition to a digital economy is omnipresent in American life. Digital equipment and high-volume data exchange are core components of American industry, from retail to manufacturing. As a result, those industries require most workers to use computer equipment regularly. Outside the workplace, households invest more each year in personal digital equipment—most notably smartphones and wearables—to enrich their daily lives. The built environment is undergoing a similar transformation, as Internet-connected sensors become commonplace in homes, businesses, infrastructure, and vehicles.

As a result, the American economy is now a living, breathing digital ecosystem—and awash in data. As Figure 1 shows, data is the new digital exhaust of the modern economy. The digitization of businesses' product inventories and sales, tracking of all freight fleets across the country, monitoring of in-home energy and Internet use—all of this leaves a trail of digital breadcrumbs of how the economy functions at the most microscopic scale. Much of this data is geospatial and structured in some way, creating enormous analytical opportunities for those who have access to it. Yet, working with this data is challenging: most databases are created in isolation and are unrelated to one another; the sheer volume of data can tax even the largest storage and processing systems; and much of the data is exclusively private or expensive to procure.

These challenges and opportunities are especially important for public agencies responsible for managing how people move around metropolitan areas. As agencies from the federal to local level consider the next stage of planning and investment in local transportation networks and land development, modernizing regulatory

Figure 1: Rate of data creation



Source: EMC Digital Universe Study

approaches to data procurement and use will affect their ability to advance broader objectives around equality of opportunity and environmental sustainability. This white paper is an early step in designing such a data-focused playbook.

This report begins by assessing current regulatory frameworks related to transportation and land use data. Next, it catalogs emerging data sources and describes the challenges to integrating those sources within current frameworks. It concludes with implications for future policy design and research efforts.

2 Current regulatory environment

Data is vital to how we plan, invest, and evaluate local transportation networks and the metropolitan areas they connect. Yet as it stands, there are enormous regulatory gaps at all levels of government regarding data collection and use.

The federal government certainly does not lack data-centric programs. The U.S. Department of Transportation (USDOT) creates multiple data sets for public consumption, manages centralized data warehousing, and even includes qualitative tool sets to improve local data collection and management. The U.S. Department of Housing and Urban Development (HUD) makes available its hyperlocal housing and investment data, and offers online mapping tools to use it. The Obama administration demonstrated a clear commitment to getting as much public data online as possible, especially geospatial and spending data. Similar efforts were developed to promote interagency data collaboration, including coordinated spatial data infrastructure. (Appendix A includes a full assessment of the federal government's approach to data collection, management, and distribution.)

The issue is that federal transportation and land use data—while certainly voluminous—is not granular and timely enough to provide clear insights on local travel and land use habits. At the federal level, there simply is no regulatory or financial commitment to better understand how people move at the local level, especially on a daily basis.

The only survey of daily travel habits is conducted every seven years and does not speak to specific origin-destination flows. Annual travel demand data only looks at large-scale commuting flows—which themselves represent a small share of all trips—while new road use data primarily tracks congestion. There is no single aggregated set of zoning data, parcel maps, or establishment types at the street scale. Even the public road network and relevant speeds are not made available. Outside unquestionably impressive geological satellite data, the federal government leaves the responsibility of charting movements in local communities to local governments.

At the state level, most governments vary considerably in their approach to local transportation and land use data collection and quality standards. Based on preliminary Brookings research, states are inconsistent in making available localized transportation investment data. State departments of transportation are likely to have expensive subscriptions to geo-specific traffic flows, but there is little to no transparency as to whether they share this data with localities (or if their contracts would permit such sharing). Certain states use project selection criteria to aggregate more information, such as Virginia's new SMART SCALE system, but this is the exception and not the rule.[1] Most states operate central geographic information system (GIS) offices, but the data they house and its quality are variable. Nor is there a centralized, regularly updated source for statewide (or metropolitan) travel surveys.[2] Due to the breadth and variety of state approaches, this is an area for extensive primary research in the future that would likely require significant resources.

The gaps within the federal and state approaches create an information vacuum for local governments within each metropolitan area. Outside federal and state reporting requirements, municipal, county, and regional governments decide what they want to collect, how much they're willing to invest in data systems, and how that data interrelates with other planning and investment policies.

There are two major implications, especially for local transportation and land use agencies.

First, there is a lack of consistent requirements from the national level, plus a wide range of state approaches. This means that it is up to each region's set of local governments to decide on their own what data to procure and how to put it into action. While such flexibilities can be advantageous, this is effectively isolating and leads to a lack of standards. As a result, there is no public entity or nonprofit that maintains a central record of the databases used by local governments in the transportation and land use fields. Operationally, this isolation makes it difficult to leverage economies of scale when signing contracts with private data providers—each deal is likely separate. It also complicates local governments' ability to find best practices as they try to innovate, since the list of peer agencies to network with is long and a search for similar approaches would be wide-ranging. Boston and Massachusetts can already see this emerging pattern. Variable approaches to real-time data collection and shared mobility services—to list just a couple—are already showing wide-ranging capabilities within the region and state.[3] Put another way, the lack of consistent statewide or local standards is already leading to divergent outcomes in Massachusetts.

Second, there is enormous potential for local information gaps. For those localities unwilling to wade into emerging transportation demand data sets, the alternatives are costly travel surveys that are often outdated before they are even finalized. But these modern data sets are the key to understanding how closely neighborhoods integrate, known as origin-destination (OD) flows. Missing and inconsistent data also restrict regions' ability to conduct benchmarking exercises, which are a vital tool to understand why other places may generate better or worse outcomes based on a given focus area. Similarly, the lack of standards and federal requirements for multi-agency collaboration promotes disconnect among experts from different sectors. For example, there's no certainty that every municipal transportation department or metropolitan planning organization (MPO) data team knows about industrial and household income growth patterns, nor that their peers in budget offices understand coming maintenance needs on local roads and sidewalks. Past work in Massachusetts confirms that improved data sharing is a vital step for the region, showing that information asymmetry exists between and within regions.[4] This isolation only reinforces path dependencies within these separate public agencies, effectively creating incentives to work with data streams already in place. This is an issue because innovative, multidisciplinary approaches to local transportation such as accessibility planning will require entirely new approaches to data.[5]

> "With limited capital investment resources, a taxpaying public demanding results from their money, and shifting travel models, antiquated data could get in the way of optimized decisionmaking."

Combined, the disconnect between regions and the information gaps at the local level only increase the likelihood of planning inefficiencies and questionable investments. Antiquated data or data that considers only one component of a market dynamic—for example, traffic levels on a road but not where those trips started or will end—makes it difficult for transportation and land use planners to understand how the built environment and their proposed changes to it directly affect travel behavior. The emergence of new travel models, most notably shared vehicles, will quickly upend assumptions about how people move. To best manage the next iteration of local construction, zoning decisions, and other major developmental policies, planners need current data with high levels of geographic granularity. With limited capital investment resources, a taxpaying public demanding results from their money, and shifting travel models, antiquated data could get in the way of optimized decisionmaking.

Emerging data

3

While current data regulation is outdated, the emergence of new data sets is an area of intense opportunity to fundamentally rethink how the country plans, builds, governs, and even finances local transportation networks and the built environment where they operate. This section presents results and key conclusions from an original market scan.

Table 1 offers a summary of private, public, and open-source data sources that are available, and an expanded version is in Appendix B.[6] To confirm the list and discover additional data providers, the Brookings team interviewed public officials at all levels of government (including Boston city and Massachusetts state) and private sector practitioners (including major data providers and noteworthy local implementers).

While not intended to be exhaustive, the wide range of sources and the depth of micro-level detail within the data are breathtaking. The evolution of data inputs is moving at a speed difficult to manage. On the positive side, local agencies and their civic-focused colleagues from the non-profit and for-profit sectors have more ways to understand how people move and interact with the built environment than at any point in human history. On the flip side, the data choices and range of capabilities could be overwhelming.

Key lessons emerge from this summary table, the expanded Appendix B, and comprehensive interviews with key stakeholders:

Table 1: Summary of data sources related to transportation and land use

Data category	Example	Typical geography	Frequently used by governments	Used in research
Private				
Telecommunications	AT&T	Domestic**	No	Yes
TNCs and rental companies	Lyft	Global	No	Beginning
Social media	Facebook	Global	No	Yes
Analytics	Inrix	Global	Yes	Yes
Location-based apps*	Yelp	Global	No	Yes
Location-based intelligence*	Telenav	Global	No	Yes
Fitness/recreation	Fitbit	Global	No	Beginning
Navigation sensors	Google Maps	Global	No	No
Financial transactions	MasterCard	Global	No	No
Industry/labor	Dun & Bradstreet	Domestic	Yes	No
Real estate listings	Trulia	Domestic	No	Beginning
Real estate intelligence	CoStar-Loopnet	Domestic	No	Beginning
Land use parcels	Digital Map Products	Domestic	No	Beginning
E-commerce	Amazon	Global	No	No
Demographics	Claritas	Domestic	No	Yes
Public				
Departments of Transportation	Fleet-based tracking	Jurisdiction	No	Yes
Taxicab commissions	OD data	Jurisdiction	Yes	No
Planning agencies	Parcel data	Jurisdiction	Yes	Yes
Environmental planning	Sensors	Jurisdiction	No	Yes
Budgets/tax	IRS	National	No	Yes
Police	Accident locations	National	Yes	Yes
Demographics	U.S. Census	National	Yes	Yes
Open Source				
Mapping	OpenStreetMap	Global	No	Yes

*Location-based service; note that many other data categories also collect location-based data.
**Telecommunications firms tend to be incorporated along country borders, but not always, and may be owned by the same parent firm (meaning the parent houses data from multiple countries).
Source: The Brookings Institution, primary in-house research

1. **There is a distinct shift in who the primary data collector/provider is.** Historically, the public sector took the leading role and invested significant resources into the collection, analysis, and dissemination of data relevant to transportation and land use planning. However, the private sector is quickly becoming a leading source of cutting-edge data that has utility in public decisionmaking. While it would be difficult to expect a local government to have the same level of financial transaction detail as MasterCard, for example, it requires a major culture shift to recognize that Google—via its mapping division—now knows more about where people move on a daily basis than their peers in local government who build the roads, rails, and sidewalks that facilitate such travel.

2. **There is a distinct shift in the types of data available.** Traditionally, transportation decisionmaking depended on supply-side data collected by public agencies. For instance, travel journals provided data on the travel patterns of a sample set of commuters, while observational surveys and traffic counters provided an idea of traffic flows in specific neighborhoods or on specific road segments. Collecting and maintaining such data is expensive, labor-intensive, and might come with longitudinal limitations if methods aren't comparable across different time periods. Now, cellphones, GPS trackers, and other navigation devices offer real-time demand-side data. This is hyper-localized, real-time data at a level of detail that simply did not exist before. Further, while earlier data sets focused more on vehicular transportation, emerging data sources can offer comparable information regarding bicycle, pedestrian, and mass transit travel—plus multimodal trips.

3. **Many publicly produced data sets are not yet part of regular planning processes.** Local, regional, and state agencies do not exhaust local resources such as tracking public vehicle fleets—consider service vehicles, not necessarily mass transit—nor fully integrate federal resources such as tax records and establishment data. Multiple interviewees confirmed the lack of digitized parcel data, both within regions and across markets in different corners of the country.

4. **Government programs are playing catch-up to the rate of data innovation.** For the data types listed as "private" in Table 1, the general trend is that public agencies do not use these private sources. Not only is this a missed opportunity, but it also symbolizes the public sector's disconnect from the innovative geospatial data creation going on within the private sector. And with few government programs using private data, the exchange of best practices is restricted.

5. **Despite the challenges, integrating these new data sets through public-private data sharing can offer clear advantages.** By capitalizing on data already being collected by private actors and finding ways to draw insights for public use, agencies can cut costs and improve the automation of data collection and management. In other words, there is no need to replicate the efforts of private data collectors if an effective, validated, and secure sharing platform can exist.

These takeaways confirm that the public sector and its peers can better target investment of limited public resources to maximize socioeconomic impact for constituents. For instance, insurance data and geo-located police data can guide traffic safety improvements, especially in accident-prone zones. Mobile phone data can be used to understand where distracted driving is a problem and to implement measures to deter such behavior.[7] Geotagged photo data can illustrate the use of popular public spaces by

locals and tourists alike, enabling greater return on investment from public space programming. Data from exercise and activity apps such as Fitbit and Runkeeper can help identify recreational hot spots that attract people and those that don't. These applied techniques would be impossible without a new approach to data integration.

4 Challenges

While the promise of using emerging data sets in transportation planning is exciting, interviews with key stakeholders revealed clear-cut challenges. The flowchart in Figure 1 summarizes the opportunities discussed in the prior sections and the challenges described in this section.

Governmental capacity and standards

As transportation, land use, and related public agencies learn to work with emerging data sources, they often lack capacity in three crucial ways: skilled personnel, data infrastructure, and cross-agency collaboration. Most of the emerging data sets fall in the realm of big data, which requires a different set of skills and data infrastructure than traditional data collection and analysis.

The lack of skilled personnel plays out on three levels. 1) There is limited fiscal capacity to hire data scientists, especially at the state and local levels; governmental salaries struggle to compete with those offered by the private sector. 2) At the managerial level, there is a lack of experience in managing data scientists and using the insights generated in confident decisionmaking. The resulting training gaps in staff and managerial positions limits an agency's ability to obtain data in a usable format, analyze it, and effectively deploy it in decisionmaking. 3) Individuals with data science experience may chafe at the highly regulated processes related to public sector project development. Data scientists looking for a more nimble, startup-like business atmosphere have plenty of opportunities to work outside of government.

The secure storage and management of large data sets require substantial investment in the necessary data infrastructure such as servers/cloud storage. However, many agencies demonstrate conservative approaches to hardware and software, holding on to current products longer than private sector peers and acting more risk averse to testing new products.

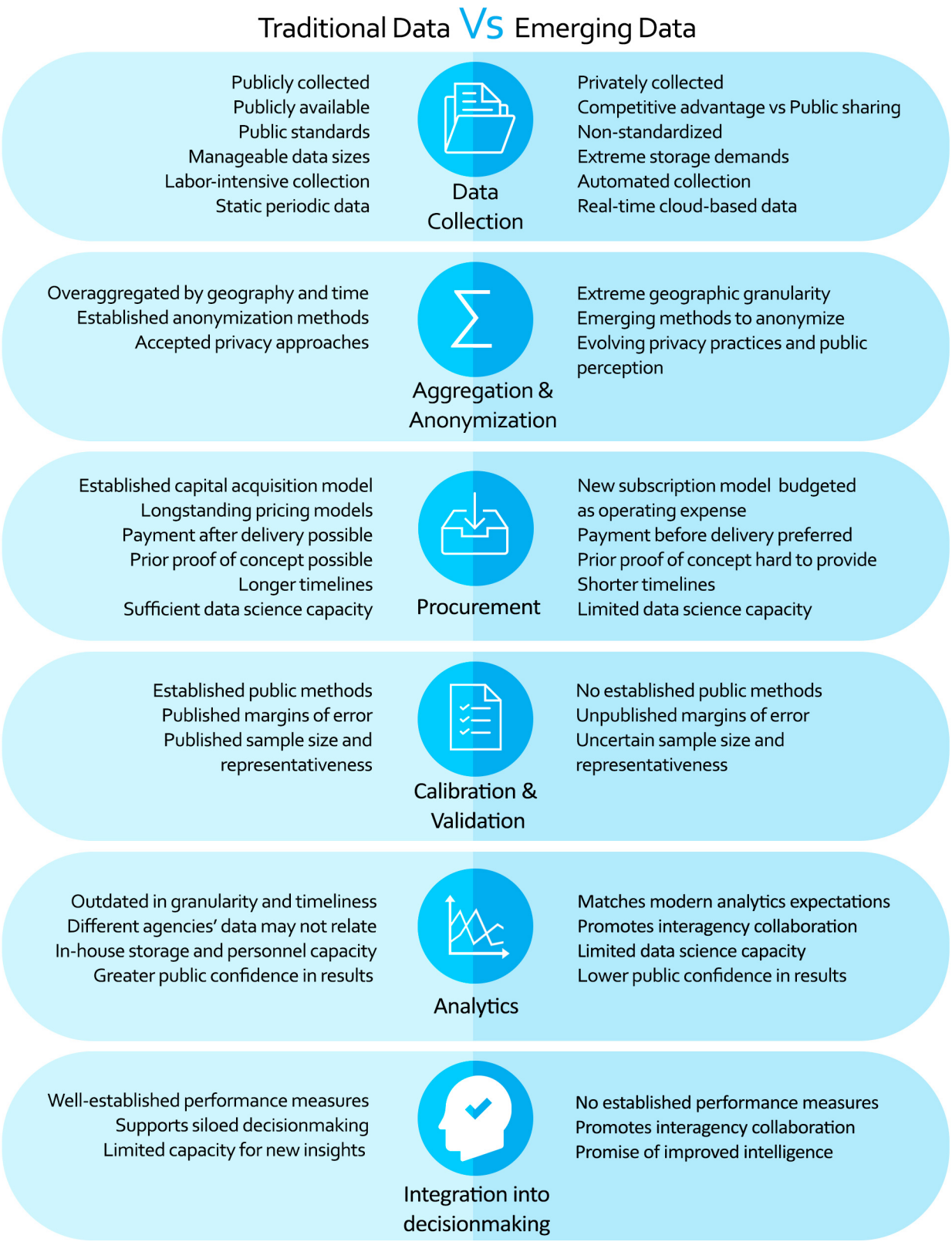
The effective use of emerging data sets requires strong cross-agency collaboration, including the ability to scale costs across multiple budgets. This can run counter to the independent functioning of many agencies related to the built environment.

Even with ample staff and physical resources, a lack of data standards for emerging data sets introduces different challenges. Without an established set of performance measures, private data providers and their public sector counterparts must often design tailored metrics for each contract. This involves considerable time and introduces risk around designing poor measures. A lack of standards also complicates benchmarking among different local, regional, and even national economies. Public officials like to make comparisons to their peers to understand various dimensions of performance, but that's difficult when peers do not generate the same performance measures or even use the same data inputs.

Procurement practices

Procurement practices are the gateway through which public and private interaction occurs around service provision. Agencies traditionally use an acquisition model of procurement—often called capital procurement—in which physical assets are acquired, inventoried, and tracked. The terms around procuring private sector data, especially databases updated daily and with large volumes, and cloud computing are a bit different

Figure 2. Comparison of data sources: Opportunities and Challenges



Source: The Brookings Institution

from standard goods and services procurements. They work on a subscription model, and agencies often do not have modernized procurement processes in place for such a model. This is reflected directly in agency budgets and standard contract provisions. While hardware is a capital expenditure, a subscription is categorized as an ongoing operating expense. Budgeting officials may simply not commit enough resources to operations-related procurements. As such, data providers do not fit neatly into this system—whether using local funds or pass-through federal funds with capital restrictions. This can result in chronic underspending on data and data infrastructure in proportion to the total scale of infrastructure project spending. Data providers also face an additional layer of uncertainty—when agencies face a budget cut, non-personnel operating expenses can be slashed first.

Another hurdle in procurement is having the right point of contact within the public agency. Private data providers are often directed to the information technology (IT) department during the procurement process, which may have little to no expertise in transportation or land use. It also can further complicate the procurement process, as some agencies may be structured for data subscriptions to come from IT budgets. At the same time, the procurement specialists within built environment agencies may have little to no experience with their in-house data storage and processing capabilities.

In the same vein, another procurement problem is a classic Catch-22. When a public agency releases a request for proposal (RFP), it is required to ask for something specific, rather than allowing the private sector to indicate what it is capable of offering. In other words, to request an innovative statement of work through existing procurement practices, a public agency needs to clearly know and articulate what it wants. But here's the catch: limited data science expertise and outdated procurement practices can make it difficult for an agency to know what it wants before it can buy it.

A further obstacle in procurement is many agencies' need for a proof of concept or trial before a long-term contract is signed. Such requirements place upfront costs on private providers to create small-scale projects, and they must wait to recoup their costs. For instance, NextBus provided a free trial to the Massachusetts Bay Transportation Authority before the state agency signed a final contract, but this is a luxury that some companies cannot afford. Many private data providers today are startups, and the longer timelines and payment-after-delivery associated with governmental projects can be challenging when their budgets and investor reports operate on tight quarterly cycles.

In terms of the RFP process, state requirements for state-specific businesses and percentage clauses in RFPs pose obstacles to startups based on their site of incorporation. The same requirements also limit agencies' ability to contract with leading innovators. In some cases, only one firm may provide a specific data set or software. There is a need to find the right balance between fostering innovation while serving local interests.

Balancing private competitive advantage and the public good

Generally, there are two types of private sector data sources: companies that are in the business of collecting and selling data and analytics, and companies that generate data but hesitate to monetize or share it. Good examples of the former from Appendix B include location-based intelligence, real estate intelligence, and analytics companies such as Inrix, Zendrive, StreetLight, NextBus, and Claritas. Examples of the latter include cellphone, mapping, and other location-based services companies. It is much easier to procure data from the first type of company, since it aligns directly with their business model.

The second type of company poses more difficulties when it comes to public sector collaboration. Often, their original data is fundamental to competitive advantage within their given marketplace. Simply put, private data helps their business thrive. Many may have a desire to support the public good—especially when it aligns with increasing their bottom line—but there are natural conflicts with data transparency. For instance, private ride-hailing companies are an intriguing alternative to marginally expensive paratransit services.[8] From a governmental perspective, handing responsibilities to a private sector colleague will require clear data sharing around pickup and drop-off points, routes, and service cost to ensure the safety and affordability of this service to users. But, from the private sector perspective, sharing this level of detail might jeopardize not only the individual's privacy but also the firm's business practices and intellectual property. Balancing these factors is a distinct challenge: without clear guidelines, justifying the enormous risks around privacy concerns for small profit margins might not be worth it for many private sector organizations.

This asymmetrical data issue will only grow in importance as new transportation service models grow in stature. Again drawing on the ride-hailing example, how does a local government adequately regulate this new industry without working with the same data as the regulated firm? While the ride-hailing industry presents a stark reminder of this given the lack of prior regulation, it is a generalizable problem across privately operated, government-regulated services. As new models of service provision such as autonomous vehicles appear on the horizon, the ability of public agencies to perform their regulatory roles is called into question unless all sides agree on common data-sharing principles.

Geographic scale and cost

Procuring and storing big data can require significant financial resources. To construct a truly multimodal data set that captures a representative sample of travel patterns and the related land uses, it is necessary to combine multiple data sets from different sources. While it can be prohibitively expensive for a small municipality or county to purchase data sets from many providers, a larger regional government or private company might be able to amortize the cost of data set procurement across local jurisdictions or clients, respectively, if there is a replicable, scalable business model. From the regional governmental perspective, the business model would require both personnel coordination and cost-sharing. From a private sector perspective, profits would depend on creating a process for national or global data collection and analytics that can then be replicated for smaller geographies at low marginal cost. In both cases, it would require a larger collection of cities and counties to work with a single data aggregator rather than creating one-on-one relationships between each city and provider. This is a new model that is yet to be tested, but holds promise.

> Generally, there are two types of private sector data sources: companies that are in the business of collecting and selling data and analytics, and companies that generate data but hesitate to monetize or share it.

Ensuring privacy

For public agencies at all levels and data-focused companies, protecting privacy is paramount. To ensure that personally identifiable information is protected, data is often anonymized and aggregated. The scale at which the anonymization and aggregation happen is a critical consideration: too much detail could raise severe privacy concerns, but too much aggregation could restrict new insights. As the case with any new business process, the public sector users and private sector providers must be willing to experiment to discover the optimal balance. However, the aversion to expose any personal information effectively works as friction against many possible public-private experiments. Both sides must find a way to safely navigate such friction, initiating carefully designed projects and then liberally sharing those insights with peers.

Calibration and validation

As transportation, land use, and similar agencies design ways to use new data sources, they must consider whether the data they procure is accurately calibrated and validated. Long-established data sets benefit from years of collective calibration: providers continually tweak the methodology, users provide feedback to the data provider, and both enable greater adjustments over time. Likewise, longtime users develop internal practices about when to use specific data sets and when not to. In the case of federal data sets from agencies such as the U.S. Census Bureau and the U.S. Bureau of Labor Statistics, published margins of error are instrumental to this validation process.

Emerging data sets don't necessarily offer the same certainty to public sector users. In some cases, the variables or entire data sets available to public sector consumers may change over time, limiting the ability to conduct longitudinal analysis. This presents a clear procurement risk. If agencies begin to depend on these new sources for data, they will need to ensure that the same sources will exist in 12 months or 12 years. In other cases, methods to validate or calibrate emerging data sets may not be made available to the public consumer, introducing new levels or risk and potential trust issues. Some emergent data sets can represent a skewed socioeconomic section of the population and might not have sufficient sample size to offer any meaningful insights. These caveats need thorough analysis and documentation. If these private data sets are to inform billions of dollars in public investment—plus zoning and other land use decisions with equally large impacts on private land values—public

officials need to have confidence in the data they're using. Using new private data sources without clear error-checking feeds or transparent calibration techniques will naturally give public sector consumers pause.

5 Implications

Integrating these new data sets into policy frameworks will require more than simply procuring data—it will call for major adjustments to internal and interagency processes. Below are some key implications as governments try to modernize their approach to transportation and land use data.

Hard-wiring objectives into planning processes

Before initiating major data, software, and hardware procurements, public officials must consider whether they can hard-wire broader objectives into their planning processes. Does the region want to reduce its carbon footprint? Is improving job accessibility a top priority? Which industries does the region want to grow and attract? These kinds of broader objectives—ones that aren't exclusively about transportation—become the guideposts to design performance measures. In turn, those performance measures can become new ways to evaluate future public projects and reforms, ranging from transportation and housing investments to zoning amendments. Once established, data can be procured and put to use around these explicit objectives, related performance measures, and clear evaluation criteria. Clear applications of new data sets can also build internal agency support for the financial resources and procurement reform needed to move from design to practice. Otherwise, procuring data without a clearly defined purpose will increase the risk that data is not used and becomes a wasted investment.

Improving public sector capacity

In the digital era, public agencies at all federalist levels must address issues around staffing and responsibilities. Data science is a relatively fast-growing field, and wage numbers suggest the labor force supply is not keeping up with demand.[9] This puts the onus on the public sector to make a clear value proposition to these in-demand workers as to why it will benefit their careers—and possibly their sense of personal satisfaction—to work in the public sector. Agency and central government leadership will also need to find champions for change and encourage healthy risk-taking among themselves and their managerial colleagues. Similarly, those top-level managers will need to consider redesigning their organizational charts and asset inventories. This includes establishing new centralized data centers or hiring more data scientists (and graphic designers to visualize their work) within existing offices. For some agencies, it will also require procuring new hardware and software to support these new occupations. Still for others, it may involve new private contracts for formerly internal work.

One possible solution to local capacity constraints is a new compact between the federal and subnational agencies. Localities always benefit from centralized data warehouses at the federal level, but expanding those to include emerging data sets will require new policy approaches. On one side, there is potential for the federal government to procure national data sets and make them available to local government peers, such as specific origin-destination (OD) flow data, as well as release more federal proprietary data, such as parcel-level industry data and historical tax records. In other instances, there is potential for local agencies to aggregate-up data sets where they have a shared interest—consider financial transaction samples—and the federal government could support this effort with targeted funding support. There is also the potential for an entirely new, quasi-independent federal agency to house extensive data. Although it would require significant time to design

and gain federal approval—rather than simply leveraging existing governmental agencies—such an independent framework could ease privacy and procurement challenges. Finally, governments increasingly interested in apprenticeships and job training broadly should consider programs that could train and steer emerging data scientists to interested state and local agencies.

Updating procurement policies

Local and state governments will need to update their procurement policies. First, the public sector will have to update its financial practices to accommodate a subscription-based model. There is a need for shorter timelines and alternatives to the payment-after-delivery model that exists today, along with ways to ensure proof of concept without shifting too much financial risk onto the private sector. Second, the public sector will need to make sure it requests the right kinds of data—including the metadata, vintage and future support. There will have to be an assigned person(s) with the subject matter and data expertise to ensure that the public sector can design statements of work that reflect the innovation occurring in the data production space. Having data scientists as an integral part of the workforce will support both goals. Many local and state governments now have chief information officers—sometimes centralized, sometimes stationed within multiple agencies—who manage these roles, and enabling them to foster more interagency collaboration will be crucial.

Addressing privacy concerns

Public officials will need to determine their willingness to work with sensitive data and how they transparently communicate their data collection efforts with the public. A recent effort by Transport for London (TfL) demonstrates how this process can go well.[10] TfL had an opportunity to use their public WiFi network to accurately track how riders moved between and within London Underground stations, but made sure to use extensive signage to inform riders about the tracking program. The groundbreaking results confirm that planning privacy information campaigns upfront is likely worth the extra time. But public communication is just one of many steps necessary to enhance public trust, especially if public agencies intend to promote open data principles.[11]

Privacy issues often grow more complex when working with private data providers. Overall, public sector leaders will need to strike a delicate balance between public concerns (such as the ability to leverage data insights for the public good) and private concerns (such as protecting intellectual property and competitive advantage). Various organizations now use the concept of differential privacy, which aims to maximize the accuracy of queries from statistical databases while minimizing the chances of identifying its records.[12] For transportation data, this means being able to generate a random representative sample while not revealing individual behavior patterns. Another approach is to anonymize at source, which involves variable techniques to introduce anonymity within the original source data flow, or even systems such as Privacy by Design that protect individual privacy from their very foundation. As public agencies consider new approaches to collect sensitive private data, they should review past efforts from similar data-driven industries, including health care and finance.

One experience from Boston encapsulates the potential consequences if these broad implications aren't addressed. In January 2015, Boston signed a deal with Uber to freely procure internal trip data, most notably anonymized OD trip data.[13] However, local reporting documented components to the agreement that limited its overall value to the city and region, including an inability to share information with other agencies and overly generalized geographic data.[14] While the two parties continue to work together, and the state debates new regulations, the suboptimal results demonstrate how a lack of clear objectives for

the data and optimized procurement can limit the effectiveness of a deal that initially looked promising.

At the same time, coalitions of private, public, and civic actors have begun to formally explore methods to share data in a manner that can boost social returns while protecting firms' competitive advantages. The World Bank's Open Transport Partnership, which includes data-sharing agreements with Asian ride-hailing firms such as Easy Taxi and Grab, offers one such model.[15] A collaboration between Transportation for America and Sidewalk Labs around "connected streets" includes an exploration of how to boost public-private collaboration.[16] The Open Algorithms (OPAL) Project represents one effort to develop algorithms that can anonymize private data to advance public good.[17] In each case, advanced planning between the different entities improved the eventual project, leading to a public launch with formal partnerships between all sides. All signs point toward more such efforts emerging in the future.

Increase sharing of best and worst experiences

With so much innovation occurring in the data production and analytical space, it's difficult for any private, public, or civic entity to keep pace. Yet in the case of public sector leaders, an inability to know what projects and reforms their peers have undertaken—and their relevant results—only stretches planning timelines and reduces opportunities for success. Launching a project to collect and share such practices could address this deficiency, although it would need sustained funding considering the high likelihood of continued practice evolution in the next few years. An established association or other national civic organization could serve as a natural home, although it would require long-term funding. A new federal project could also serve the communicator role, but would also face similar long-term funding questions.

Conclusion

The combination of millions of geospatial sensors—whether smartphones in individual pockets or fixed equipment in public spaces—and rapidly developing computing capacity have created new opportunities to answer a long-held fascination: how are people and products moving in space? Emerging data sets based on travel, fitness, and purchasing habits combine with geo-located real estate and infrastructure supply data to enable public agencies to answer this fundamental question with more certainty than at any time in human history. Plotting near-misses on city streets, understanding athletic use of public lands, mapping how many neighborhood residents shop locally: these high-tech inputs are no longer science fiction.

Yet, leveraging all this new data will not be easy. Numerous challenges concerning privacy and procurement make it harder to successfully integrate emerging data sets into current public policy frameworks. Government capacity, both in terms of staffing and financial resources, will need to improve. Data-rich firms will need to feel more comfortable in sharing data with public agencies.

There is reason for optimism, though. The challenges listed in this report are surmountable, and doing so will help develop new forms of public-private partnerships and modernized policy frameworks along the way. Data is better than it has ever been, and governments have an incredible opportunity to institute the data-related reforms that will help them deliver more equitable, sustainable, and efficient communities.

Selected references

Abdulazim, Tamer, Hossam Abdelgawad, Khandker M. Nurul Habib, and Baher Abdulhai. "Framework for Automating Travel Activity Inference Using Land Use Data." Transportation Research Record: Journal of the Transportation Research Board 2526 (January 1, 2015): 136–42. doi:10.3141/2526-15.

Andrienko, G., N. Andrienko, P. Bak, S. Kisilevich, and D. Keim. "Analysis of Community-Contributed Space- and Time-Referenced Data (example of Flickr and Panoramio Photos)." In 2009 IEEE Symposium on Visual Analytics Science and Technology, 213–14, 2009. doi:10.1109/VAST.2009.5333472.

Aubrecht, Christoph, Joachim Ungar, and Sergio Freire. "Exploring the Potential of Volunteered Geographic Information for Modeling Spatio-Temporal Characteristics of Urban Population: A Case Study for Lisbon Metro Using Foursquare Check-in Data," 57–60. Department of Civil Engineering of the University of Coimbra and e-GEO, Research Center in Geography and Regional Planning of the Faculty of Social Sciences and Humanities of the Nova University of Lisbon, 2011. http://upcommons.upc.edu/handle/2099/15418.

Bowes, David R., and Keith R. Ihlanfeldt. "Identifying the Impacts of Rail Transit Stations on Residential Property Values." Journal of Urban Economics 50, no. 1 (July 1, 2001): 1–25. doi:10.1006/juec.2001.2214.

Connington, James. "Mapped: UK Areas That Rely Most upon Credit Cards." Telegraph, April 28, 2016. http://www.telegraph.co.uk/personal-banking/credit-cards/mapped-uk-areas-that-rely-most-upon-credit-cards/.

Cortés, Cristián E., Jaime Gibson, Antonio Gschwender, Marcela Munizaga, and Mauricio Zúñiga. "Commercial Bus Speed Diagnosis Based on GPS-Monitored Data." Transportation Research Part C: Emerging Technologies 19, no. 4 (August 2011): 695–707. doi:10.1016/j.trc.2010.12.008.

Crandall, David J., Lars Backstrom, Daniel Huttenlocher, and Jon Kleinberg. "Mapping the World's Photos." In Proceedings of the 18th International Conference on World Wide Web, 761–70. WWW '09. New York: ACM, 2009. doi:10.1145/1526709.1526812.

D'Agostino, Michelangelo. "Mapping 35 Million Credit Cards on Top of Census Data with R." Strongly Typed, February 25, 2013. http://www.braintreepayments.com/blog/vaulted-credit-card-maps-with-r/.

De Nadai, Marco, Jacopo Staiano, Roberto Larcher, Nicu Sebe, Daniele Quercia, and Bruno Lepri. "The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective." Proceedings of the 26th International ACM Conference on World Wide Web (WWW), 2016, 2016, 413–23. doi:10.1145/2872427.2883084.

Faouzi, Nour-Eddin El, Henry Leung, and Ajeesh Kurian. "Data Fusion in Intelligent Transportation Systems: Progress and Challenges—A Survey." Special Issue on Intelligent Transportation Systems, Information Fusion 12, no. 1 (January 2011): 4–10. doi:10.1016/j.inffus.2010.06.001.

Grauwin, Sebastian, Stanislav Sobolevsky, Simon Moritz, István Gódor, and Carlo Ratti. "Towards a Comparative Science of Cities: Using Mobile Traffic Records in New York, London, and Hong Kong." In Computational Approaches for Urban Environments, edited by Marco Helbich, Jamal Jokar Arsanjani, and Michael Leitner, 363–87. Geotechnologies and the Environment 13. Springer International Publishing, 2015. doi:10.1007/978-3-319-11469-9_15.

Guo, Diansheng, Xi Zhu, Hai Jin, Peng Gao, and Clio Andris. "Discovering Spatial Patterns in Origin-Destination Mobility Data." Transactions in GIS 16, no. 3 (June 1, 2012): 411–29. doi:10.1111/j.1467-9671.2012.01344.x.

Haklay, Mordechai. "How Good Is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets." Environment and Planning B: Planning and Design 37, no. 4 (August 1, 2010): 682–703. doi:10.1068/b35097.

Iqbal, Md. Shahadat, Charisma F. Choudhury, Pu Wang, and Marta C. González. "Development of Origin-desti-

nation Matrices Using Mobile Phone Call Data." Transportation Research Part C: Emerging Technologies 40 (March 2014): 63–74. doi:10.1016/j.trc.2014.01.002.

Jin, Peter J., Fan Yang, Meredith Cebelak, Bin Ran, and C. Michael Walton. "Urban Travel Demand Analysis for Austin, Texas, Using Location-Based Social Networking Data," 2013. https://trid.trb.org/view.aspx?id=1241465.

Kung, Kevin S., Kael Greco, Stanislav Sobolevsky, and Carlo Ratti. "Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data." PLOS ONE 9, no. 6 (June 16, 2014): e96180. doi:10.1371/journal.pone.0096180.

Li, Xiaolong, Gang Pan, Zhaohui Wu, Guande Qi, Shijian Li, Daqing Zhang, Wangsheng Zhang, and Zonghui Wang. "Prediction of Urban Human Mobility Using Large-Scale Taxi Traces and Its Applications." Frontiers of Computer Science in China 6, no. 1 (February 1, 2012): 111–21. doi:10.1007/s11704-011-1192-6.

Lin, Wei-Hua, and Jian Zeng. "Experimental Study of Real-Time Bus Arrival Time Prediction with GPS Data." Transportation Research Record: Journal of the Transportation Research Board 1666 (January 1, 1999): 101–09. doi:10.3141/1666-12.

Mamei, M., A. Rosi, and F. Zambonelli. "Automatic Analysis of Geotagged Photos for Intelligent Tourist Services." In 2010 Sixth International Conference on Intelligent Environments, 146–51, 2010. doi:10.1109/IE.2010.34.

Mayer, Michelle L. "Disparities in Geographic Access to Pediatric Subspecialty Care." Maternal and Child Health Journal 12, no. 5 (September 2008): 624–32. doi:10.1007/s10995-007-0275-3.

Beshah, Tibebe, and Shawndra Hill. "Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors on Accident Severity in Ethiopia." In AAAI Spring Symposium: Artificial Intelligence for Development, 2010.

Noulas, A., C. Mascolo, and E. Frias-Martinez. "Exploiting Foursquare and Cellular Data to Infer User Activity in Urban Environments." In 2013 IEEE 14th International Conference on Mobile Data Management, 167–76, 2013. doi:10.1109/MDM.2013.27.

O'Brien, Oliver, James Cheshire, and Michael Batty. "Mining Bicycle Sharing Data for Generating Insights into Sustainable Transport Systems." Journal of Transport Geography 34 (January 2014): 262–73. doi:10.1016/j.jtrangeo.2013.06.007.

Stedman, Richard, Tom Beckley, Sara Wallace, and Marke Ambard. "A Picture and 1000 Words: Using Resident-Employed Photography to Understand Attachment to High Amenity Places." Journal of Leisure Research 36, no. 4 (Fourth Quarter 2004): 580–606.

Stragier, Jeroen, Peter Mechant, and Lieven De Marez. "Studying Physical Activity Using Social Media: An Analysis of the Added Value of RunKeeper Tweets." International Journal of Interactive Communication Systems and Technologies (IJICST) 3, no. 2 (July 1, 2013): 16–28. doi:10.4018/ijicst.2013070102.

Traunmueller, Martin, Giovanni Quattrone, and Licia Capra. "Mining Mobile Phone Data to Investigate Urban Crime Theories at Scale." In Social Informatics, edited by Luca Maria Aiello and Daniel McFarland, 396–411. Lecture Notes in Computer Science. Springer International Publishing, 2014. http://link.springer.com/chapter/10.1007/978-3-319-13734-6_29.

Zhou, Pengfei, Yuanqing Zheng, and Mo Li. "How Long to Wait? Predicting Bus Arrival Time with Mobile Phone Based Participatory Sensing." In Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, 379–92. MobiSys '12. New York: ACM, 2012. doi:10.1145/2307636.2307671.

Housing rule—would also gut many of HUD's data collection programs. This bill follows in the same spirit of other commentary by federal Republican legislators since their ascension to control of the House in 2011 and of the entire Congress in 2017, including consistent recommendations to eliminate vital Census data products.[37] Congress and federal agencies also continue to generally resist requiring state and local data collection if it is not accompanied by federal funding support, what are often referred to as unfunded mandates. If Congress and agency leadership are not willing to fund data programs, this approach can birth an unvirtuous cycle in which less funding makes more data programs look like unfunded mandates.

Endnotes

1. For more background on Virginia's SMART SCALE, see the project website at http://vasmartscale.org/ (accessed May 2017).

2. The University of Minnesota's Metropolitan Travel Survey Archive collected local surveys, but has not been updated in years. The Inter-university Consortium for Political and Social Research housed at the University of Michigan also maintains a mixed database of travel surveys, but is not exhaustive.

3. For more information, see: Tony Dutzik, ed., "Fast Forward: The Technology Revolution in Transportation and What It Means for Massachusetts" (Boston: Transportation for Massachusetts, 2016).

4. Ibid.

5. Jeffrey Gutman and Adie Tomer, "Developing a Common Narrative on Urban Accessibility: Overview" (Washington: Brookings Institution, 2016).

6. More sources than those listed in Appendix B are certainly available, and the list will only expand as more innovation occurs.

7. "Cellular Census: Explorations in Urban Data Collection" can be found at http://www.ucl.ac.uk/~ucftb48/Cellular_Census.pdf (accessed May 2017).

8. Joseph Kane, Adie Tomer, and Robert Puentes, "How Lyft and Uber Can Improve Transit Agency Budgets" (Washington: Brookings Institution, 2016).

9. Gil Press, "The Supply and Demand of Data Scientists: What The Surveys Say," Forbes, April 30, 2015.

10. James O'Malley, "Here's What TfL Learned From Tracking Your Phone on the Tube," Gizmodo UK, February 13, 2017.

11. Ben Green et al., Open Data Privacy Playbook (Cambridge, Mass.: Berkman Klein Center for Internet and Society at Harvard University, 2017).

12. Cynthia Dwork and Aaron Roth, "The Algorithmic Foundations of Differential Privacy," Foundations and Trends® in Theoretical Computer Science 9, nos. 3–4 (2014), 211-407.

13. Nicole Dungca, "In First, Uber to Share Ride Data with Boston," Boston Globe, January 13, 2015.

14. Adam Vaccaro, "Highly Touted Boston-Uber Partnership Has Not Lived Up to Hype So Far," Boston.com, June 16, 2016.

15. For more background on the Open Transport Partnership and related materials, see the announcement press release at http://www.worldbank.org/en/news/press-release/2016/12/19/the-world-bank-launches-new-open-transport-partnership-to-improve-transportation-through-open-data (accessed May 2017).

16. For more information about the June 2017 meeting involving these issues, see the recap web story at http://t4america.org/2017/06/20/smart-cities-collaborative-rolls-cities-get-nuts-bolts/ (accessed June 2017).

17. The OPAL Project website is http://www.opalproject.org/ (accessed May 2017).

18. Primarily under §1201 and §1202 of the FAST Act; primarily under §1203 of MAP-21.

19. The updated congestion- and greenhouse gas-related performance measurement was made under Docket Number FHWA-2013-0054.

20. More detail on the NPMRDS can be found at https://ops.fhwa.dot.gov/perf_measurement/ (accessed May 2017).

21. More detail on TMIP can be found at https://www.fhwa.dot.gov/planning/tmip/ (accessed May 2017).

22. More detail on HPMS can be found at https://www.fhwa.dot.gov/policyinformation/hpms.cfm (accessed May 2017).

23. Authors' review of relevant data sets.

24. Brookings interview with the USDOT chief financial officer and staff, June 2016.

25. Authors' review of data sets at https://www.bts.gov/ (accessed May 2017).

Appendix A

Federal regulatory approach to transportation and land use data

The federal government takes a relatively laissez-faire approach to data regulation, with most programs related to transportation and land use permitted to design their own requirements around data collection and use. While recent developments show a shift toward more specific collection requirements and quality standards, the progress is fragile and far from a finished product.

Programs administered under the U.S. Department of Transportation (USDOT) demonstrate the breadth of regulatory approaches to data collection and characteristics. On one side, federal surface transportation law uses planning requirements to compel metropolitan planning organizations (MPOs) to generate specific performance measures.[18] This required USDOT to develop new data-specific reporting rules, including requirements around road congestion and carbon emissions.[19] However, USDOT is not overly prescriptive of the data components MPOs must use to inform these performance measures, which confers broad flexibilities for local governments but limits the potential for national benchmarking and comparison. On the other extreme, USDOT makes available to all states and MPOs a common travel performance data set, the National Performance Management Research Data Set (NPMRDS).[20] In this instance, local governments can use geolocated transportation data for free, but they have no control over the included data fields. Somewhat more in the middle, USDOT also offers direct planning support such as the Travel Model Improvement Program (TMIP), whose Travel Analysis Toolbox provides decisionmaking tools for statewide and local transportation planning.[21]

USDOT also works directly with states to collect certain geospatial data, which federal staff aggregate into publicly available data sets. These include the Highway Performance Monitoring System (HPMS), which aggregates traffic counters across the country to assemble a county-level database of driving levels and other road conditions.[22] Vital safety data, most notably through the Crash Outcome Data Evaluation System and the Fatality Analysis Reporting System, offers strong geographic granularity based on state reporting requirements, but does not cover the full extent of vehicular accidents.[23] The agency's in-house financial system, FMIS, does not require states to report their federal spending to a specific geography.[24] Finally, while USDOT spends well over $20 million per year to support the Bureau of Transportation Statistics and its centralized portal, little of this data relates to local transportation habits, nor is it geographically granular enough to support improved planning and analysis.[25] In fact, "local" or "metropolitan" is not an available geography from its drop-down menus.

The U.S. Department of Housing and Urban Development (HUD) is another major agency in this space, and its record is one of more geographically granular data collection and overall data transparency. The long-standing HUDUser portal centralizes much of the agency's hyperlocal data and includes a range of housing data and investments of federal dollars.[26] HUD recently launched the Community Assessment Reporting Tool, which builds on the prior platform with a more accessible mapping format and single webpage aggregation of multiple categories of different funding flows and investments.[27] Simply put, HUD has a long-standing commitment to publicizing its data.

At a more centralized level, the path during the Obama administration was clearly to expand data volume and richness. The 2009 American Recovery and Reinvestment Act included explicit requirements around geospatial reporting of spending data.[28] That data then fed into USAspending.gov (itself mandated by the Federal Funding Accountability and Transparency Act

of 2006) and data.gov (also created in 2009).[29] Based on a 2013 executive order and the Office of Management and Budget's (OMB) Open Data Policy, the federal government committed all new federal data sets to be made in machine-readable formats and be open to the public.[30]

Similar reform efforts continued to be presented in Congress, including introduction of the Geospatial Data Act of 2015 and the OPEN Government Data Act of 2016. Building energy behind these efforts will be the Commission on Evidence-Based Policymaking, which was commissioned by law in 2016.

Aiding efforts to grow data volume and quality are the standards for geospatial data established through the Federal Geographic Data Committee.[31] Chaired by the secretary of the interior and OMB's deputy director for management, this group helps coordinate federal agencies and develops a strategic plan to advance the National Spatial Data Infrastructure, which is a set of broad policies related to geospatial data.[32] This includes the centralized website, GeoPlatform.gov, which creates economies of scale by pooling information around past and upcoming data purchases.[33] GeoPlatform also serves as a central information portal for other multiagency efforts, such as the Multi-Resolution Land Characteristics Consortium (MRLC). MRLC involves 10 distinct federal agencies—including the U.S. Environmental Protection Agency, the U.S. Geological Survey, and NASA—and maybe most notably publishes the National Land Cover Database, an important collection of land use data for federalist use.[34]

Yet even with the federal expansion of data collection writ large and geospatial standards in particular, there is still much that federal law does not require local governments to collect and much data that the federal government collects but does not share. There are no official rules requiring local governments to digitize their land use data, including parcel maps, zoning codes, and building conditions. The same applies to land-based transportation features, whether it be sidewalk quality or digital transit feeds.[35] There still is not a national, publicly available data set of all road locations and speed limits. The U.S. Census Bureau maintains an enormous collection of micro-level establishment data by industry but does not make it available to local governments—forcing them to procure private data when their work demands it. A similar situation applies to invaluable tax data via the IRS.

Maybe the most limiting issue is how the federal government approaches an emerging category of data—local transportation demand data. The issue certainly is not one of type. The federal government already collects detailed travel demand data for commercial aviation passengers and a major five-year survey of goods trade, and these data sets are available to the public. But each of those data sets involves inter-metropolitan flows, not local. Where it does maintain local demand data, such as county-level HPMS data or Census commuting flows, either it's overly aggregated geographically or the data is too narrow in terms of trip type and frequency of data collection. This is especially true for the Census Transportation Planning Products (CTPP).[36] While CTPP is arguably the most widely used place-to-place travel data in the country, it tracks only journey-to-work trips, and the date range typically covers multiple years. The country's only national travel survey—the National Household Travel Survey—is collected too infrequently and without ample geographic granularity. There simply is no regulatory or financial commitment to better understand how people move at the local level on a daily basis.

Finally, there is no guarantee the federal government's modernizing approach to data will continue to be one of openness and higher volume collection. The Local Zoning Decisions Protection Act of 2017—a bill introduced to ostensibly nullify the Affirmatively Furthering Fair

Appendix B: Selected data sources

Category	Source	Data types	Geography	Examples of real-world use	Examples of relevant academic literature
Private data sources					
Telecom	AT&T, Verizon	Origin-destination (OD) travel; GPS-like travel habits; other communication choices (voice, text)	National		Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data Mining Mobile Phone Data to Investigate Urban Crime Theories at Scale How Long to Wait? Predicting Bus Arrival Time with Mobile Phone Based Participatory Sensing Towards a Comparative Science of Cities: Using Mobile Traffic Records in New York, London and Hong Kong The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective
Transportation network and rental companies	Uber, Lyft, Car2Go, Zipcar	OD travel by cab; GPS data	Cities of availability	NYC Taxi & Limousine Commission	
	Bikeshare	OD travel by bike; some GPS data	Cities of availability	District Department of Transportation	Mining Bicycle Sharing Data for Generating Insights into Sustainable Transport Systems
Social media	Pinterest, Instagram, Facebook, Twitter, Flickr		National		Mapping the World's Photos Automatic Analysis of Geotagged Photos for Intelligent Tourist Services Analysis of Community-contributed Space- and Time-referenced data A Picture and 1000 Words: Using Resident-Employed Photography to Understand Attachment to High Amenity Places

Category (Contd.)	Source	Data types	Geography	Examples of real-world use	Examples of relevant academic literature
Location-based service apps	Dark Sky, AccuWeather, Yelp, Foursquare, Pokémon Go	Location	Cities of availability		Exploiting Foursquare and Cellular Data to Infer User Activity in Urban Environments
					Framework for Automating Travel Activity
					Urban Travel Demand Analysis for Austin TX USA using Location-based Social Networking Data
					Exploring the Potential of Volunteered Geographic Information for Modeling Spatio-Temporal Characteristics of Urban Population: A Case Study for Lisbon Metro Using Foursquare Check-In Data
Fitness/ recreation	Fitbit, Nike Run Club, Runkeeper, MapMyRun	Location, speed, time of use, elevation, recreational hot spots	National		Studying Physical Activity Using Social Media: An Analysis of the Added Value of RunKeeper Tweets
Navigation sensors	NextBus, Transit (app)	Real-time location information of transit	Many cities in U.S., one city in Australia	NextBus - city DOTs such as SF	
	Google Maps		National		
	Waze (acquired by Google in 2013), HERE		National		
Financial transactions	Credit card companies (e.g., MasterCard)	Location, time of use, financial, demographics	National		
Location-linked industry/ labor market	Dun & Bradstreet		National		
Real estate listings	Trulia, Zillow, Apartments.com, Housing.com, PadMapper, Craigslist, Airbnb, LoopNet, CommercialSearch, 42Floors, MLS (Multiple Listing Service), Redfin (via MLS), Realtor.com	Real estate inventory details	National/ subnational		

Category (Contd.)	Source	Data types	Geography	Examples of real-world use	Examples of relevant academic literature
Real estate intelligence	CoStar-LoopNet, Xceligent-CommercialSearch, ReisReports, CBRE	Real estate inventory details	National		The WalkUP Wake-Up Call: Boston
Land use parcel data	Digital Map Products	Parcel and property details	National		
E-commerce	Amazon, eBay	Geography of purchase and delivery; vehicle routing	National		
Demographics	Claritas, Nielsen	Socioeconomic indicators for the population	National		Missed Opportunity: Transit and Jobs in Metropolitan America
Transport data aggregation, crowdsourcing and analytics	Inrix, TomTom, Garmin	OD flows by multiple modes, freight	Global		
	StreetLight, Airsage, Zendrive	OD travel by multiple modes	National		
Public data sources					
Departments of Transportation	GPS-based fleet tracking	Location, speed, surrounding traffic	By city/ DOT/ private fleets		Experimental Study of Real-Time Bus Arrival Time Prediction with GPS Data Prediction Model of Bus Arrival Time at Signalized Intersection Using GPS Data Commercial Bus Speed Diagnosis Based on Big Data Intelligence for Eco-Friendly Bus Routing
	Traffic speed detectors/ other sensors		By city/ DOT	NYCDOT Traffic Map	
Federal Highway Administration	Highway performance monitoring system		National		
Taxicab commissions	Taxi Origin-Destination	Location, time of use	By city/DOT		
Planning agencies	Land-use maps, parcel data		By city/ MPO/ state		

Category (Contd.)	Source	Data types	Geography	Examples of real-world use	Examples of relevant academic literature
Environmental planning/ pollution control boards	Air and water quality sensors	Pollution levels	By city/state		
Budgets/tax	IRS	Commuter tax benefits, property tax	National (but require special access)		
	Federal/ state budget office	Tax income	By city/state (but require special access)		
Police	Accident locations		By city/ DOT		
	CCTV surveillance		By city/ DOT		
Demographics	U.S. Census, Bureau of Labor Statistics, et al.	Socioeconomic indicators for the population	National		
Open source data					
Mapping	OpenStreetMap	Location of practically everything in cities	National		How Good Is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets

Acknowledgments

The Brookings Metropolitan Policy Program gratefully acknowledges the Barr Foundation and the Ford Foundation for their support of this research. The Program is also grateful to the Metropolitan Council, a network of business, civic, and philanthropic leaders who act as financial and intellectual partners of the Metro Program.

The authors wish to thank the following interviewees who generously contributed their time and expertise:
Chris Dempsey (Transportation for Massachusetts)
Linda Bailey and Mollie Pelon (National Association of City Transportation Officials)
Laura Schewel (StreetLight Data)
Kevin Webb (Independent consultant)
Andy Taylor (Cubic Transportation Systems)
Tony Dutzik (Frontier Group)
Bob Pishue and Ted Trepanier (Inrix)
Daniel Morgan (U.S. Department of Transportation)
Chris Osgood (City of Boston)
Dean Violagis (CoStar)
Caleb Winter and Robert Kirkman (Oregon Metro)
Jake Sion (Transit App)
Anonymous
Jascha Franklin-Hodge (City of Boston)
David Block-Schachter (Massachusetts Bay Transportation Authority).

The authors also thank the following individuals for their helpful and constructive comments on an earlier draft of this report:
Eric Jaffe (Sidewalk Labs)
Jacob Mason (ITDP)
Robert Puentes (Eno Center for Transportation)

Image Attributions

Cover page icon via Flaticon
Map tile by Stamen Design
Photos by Dan Gold, Thomas Kvistholt and Zach Inglis on Unsplash

About the Metropolitan Policy Program at Brookings

The Metropolitan Policy Program at Brookings delivers research and solutions to help metropolitan leaders build an advanced economy that works for all.
To learn more, visit www.brookings.edu/metro.

The Brookings Institution is a nonprofit organization devoted to independent research and policy solutions. Its mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations for policymakers and the public. The conclusions and recommendations of any Brookings publication are solely those of its author(s), and do not reflect the views of the Institution, its management, or its other scholars.

Brookings is committed to quality, independence, and impact in all of its work. Activities supported by its donors reflect this commitment.

For More Information

Adie Tomer
Fellow
Metropolitan Policy Program at Brookings
atomer@brookings.edu

B | Metropolitan Policy Program
at BROOKINGS

1775 Massachusetts Avenue, NW
Washington, D.C. 20036-2188
telephone 202.797.6139
fax 202.797.2965