Data Shop

Data Shop, a department of Cityscape, *presents short articles or notes on the uses of data in housing and urban research. Through this department, the Office of Policy Development and Research introduces readers to new and overlooked data sources and to improved techniques in using well-known data. The emphasis is on sources and methods that analysts can use in their own work. Researchers often run into knotty data problems involving data interpretation or manipulation that must be solved before a project can proceed, but they seldom get to focus in detail on the solutions to such problems. If you have an idea for an applied, data-centric note of no more than 3,000 words, please send a one-paragraph abstract to* david.a.vandenbroucke@hud.gov *for consideration.*

The Smart Location Database: A Nationwide Data Resource Characterizing the Built Environment and Destination Accessibility at the Neighborhood Scale

Kevin Ramsey
BERK Consulting

Alexander Bell
Renaissance Planning Group

Abstract

A large body of research has demonstrated that land use and urban form can have a measurable effect on the daily transportation habits of urban and suburban residents. These findings can help to inform travel demand studies and evaluations of the likely effects of land use decisions on residents' transportation choices and costs. Developing reliable data can be expensive and time consuming, however. The goal of the U.S. Environmental Protection Agency's Smart Location Database (SLD) is to summarize relevant built environment and destination accessibility variables for every census block group in the nation and to share them publicly in support of planning and research studies nationwide. This article describes the variables available in the SLD and the novel approaches we developed to calculate these variables using available private and

Abstract (continued)

public data sources. Of particular note are several measures of accessibility to destinations via transit developed through an analysis of more than 220 public transit data feeds available from agencies across the United States. The article concludes with a case study describing one current use of the SLD: evaluating potential employment facility locations.

Introduction

During the past two decades, the planning profession has seen an explosion of interest in the roles that land use and urban design play in shaping the transportation habits, health, and livelihood of urban and suburban residents. Researchers in the fields of transportation planning and public health have begun to isolate and measure the relationships between the built environment in which we live and work and our propensity to choose walking, transit, or driving to meet our everyday transportation needs. These studies tend to focus on neighborhood characteristics such as the density of development, mixing of land uses, connectivity of street networks, availability of transit, and accessibility to destinations via car, transit, or foot. A 2010 meta-analysis of this literature reviewed more than 200 different studies (Ewing and Cervero, 2010). Findings from this body of research are being used to inform traffic impact analyses (Ewing et al., 2011; Gulden, Goates, and Ewing, 2013), land use scenario-planning studies (Bartholomew and Ewing, 2008), environmental impact analyses (Ramsey and Poresky, 2013), health impact assessments (de Nazelle et al., 2011), and estimates of transportation cost burdens associated with living in a particular place (Haas et al., 2008). These kinds of studies enable planners and community advocates to quantify the potential benefits of local land use decisions such as encouraging compact and mixed-use development, allowing for more jobs and housing to be in walkable and transit-rich neighborhoods, and reducing the amount of new low-density development occurring at the outer suburban fringe.

Developing data that summarize built environment characteristics unfortunately can be expensive and time consuming. Moreover, each time a new community wants to conduct a planning study, the same general kinds of data must be identified, gathered, and processed. We wondered, therefore, if an economy of scale could be achieved by developing data about the built environment at the block group scale for the entire United States. These data would necessarily rely on national sources or widely used data standards. Therefore, the results could be inferior to locally derived metrics that rely on detailed land use data available only at the local scale. We hypothesized, however, that a nationwide study could produce data that are sufficient for many local and regional studies that would not otherwise move forward because they lack resources. We also hypothesized that making nationally consistent data freely available could spur the development of third party planning analysis tools that significantly reduce barriers to entry for communities seeking to analyze the potential effects of land use decisions.

It is not surprising that summarizing neighborhood-scale built environment characteristics using only nationally available data involves significant challenges. Most notable among these challenges is the lack of a publically available database that describes the location and use of privately owned land parcels. Therefore, an analysis seeking to summarize the density of commercial development, mix of land uses, or availability of destinations must derive these metrics from proxies such as job counts broken down by employment sector. Similar challenges affect the ability to accurately model pedestrian mobility and transit service. Our study is the first attempt to navigate such challenges to reliably summarize neighborhood-scale built environment conditions for the entire United States.

Developing the Smart Location Database

The U.S. Environmental Protection Agency's (EPA's) Smart Location Database (SLD) includes more than 90 variables summarizing conditions for every census block group in the United States. It is broken into 10 topic areas: administrative, area, demographics, employment, density, diversity (of land use), design, transit service, destination accessibility via automobile and via transit, and regional summaries. All administrative, demographic, and employment variables came directly from the 2010 U.S. census, 2006–2010 American Community Survey (ACS) 5-year estimates,[1] or 2010 Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics.[2] Other variables required additional analysis. A full listing of all variables, including data sources, is available in the appendix.

It is beyond the scope of this article to fully describe the method of derivation for every variable in the SLD. Such information is available in the "Smart Location Database User Guide" (Ramsey and Bell, 2013). In this article, we describe our derivation approach and some key challenges we navigated on the way.

Density

One common drawback of calculating the density of population, housing, or employment using only census data is that any given block group may contain both developed and undeveloped land area. Block groups may also include parks or other areas protected from development activity. As a result, average block group density may differ substantially from the actual density of development experienced by residents and visitors. To address this problem, we obtained data from the Protected

[1] ACS 5-year estimates are derived from survey data collected during a 5-year period to ensure a large enough sample size in smaller geographic units. Therefore, the data released reflect conditions during a 5-year period. The census also releases margins of error for all ACS data values. These margins of error can be quite large for individual block groups, in some circumstances. Users of the demographic data available in the SLD are encouraged to consult the ACS to assess the accuracy of estimates. Most variables in the SLD are not derived from ACS data.

[2] LEHD data summarize employment at the census block level for all U.S. states except Massachusetts. Massachusetts data were provided by the Metropolitan Area Planning Council. LEHD data are developed by synthesizing state unemployment insurance earnings data and the Quarterly Census of Employment and Wages data with additional administration data from the census. Because employment data originate from individual states, some inconsistencies arise in regards to how employment locations are assessed. For instance, employment in some school districts is allocated to the school district headquarters instead of the individual school locations, which contributes to data quality problems that affect a number of SLD variables. In general, some individual block group employment estimates may be inaccurate. The broad patterns of employment depicted in these data, however, appear consistent with known conditions. More information about census LEHD data and its limitations is available at http://lehd.did.census.gov/data/.

Area Database of the United States (USGS GAP, 2012) and NAVTEQ[3] to identify land areas that are parklands, privately owned conservation easements, and other public lands that are protected from private development activity. We used geographic information systems to overlay these areas with block group polygons to isolate the unprotected areas of each block group. We then determined the unprotected land area of each block group and used this value to calculate each activity density variable. The result, we believe, is an improved estimation of actual density of activity.

Diversity

Land use diversity refers to the relative mix of different land uses. For this national study we used housing counts and employment counts broken down by job sector as proxies for different land uses. Although this approach enabled us to calculate a variety of different entropy metrics (see the appendix), it had a few notable limitations. First, counts summarized at the block group scale provide no information about how different activities are spatially distributed within each block group. For instance, a very large block group in an area of low-density development may include a variety of different activities. Those activities, however, may be spatially separated within the block group area. As a result, any given part of the block group might have very little diversity when examined in detail, even though the diversity value for the block group as a whole is quite high.

Another problem emerges in higher density urban areas, where block groups may be quite small. In this case, it is possible for a block group dominated by office jobs or residential uses to have a very low value for land use diversity even though a variety of different and complementary land uses exist directly across the street in a neighboring block group. Our methods of calculation did not consider activities outside the boundaries of any given block group. A more sophisticated spatial analysis approach could be used to partially address this latter limitation. For instance, it could be possible to estimate the mix of land uses in all block groups that intersect a 0.25-mile radius of each block group centroid.

Despite these limitations, a few diversity metrics calculated for the SLD have proven to be correlated with outcomes of interest such as workplace-based walk trips and vehicle travel. Therefore, we are optimistic they will prove to be at least somewhat useful in their current form.

Design

Our urban design variables all measure some aspect of street connectivity based on a detailed analysis of street network data from NAVTEQ. Highly connected street networks enable travelers to reach nearby destinations more efficiently. Although design metrics are most commonly used to assess the pedestrian environment, we were challenged by the lack of data about the presence or quality of sidewalks. Our solution involved analyzing the roadway link attribute information. Using attributes such as speed class, direction of travel (one- or two-way), and auto or pedestrian restrictions enabled us to classify each roadway link as auto-oriented, multimodal, or pedestrian-oriented.

[3] NAVTEQ is a geographic data provider that undertakes independent data collection rather than relying on government maps and data sources. It is the primary data source for many portable global positioning system devices and navigation systems. The parks and street network data used in this analysis were released in 2011.

This classification enabled us to develop a variety of metrics that, collectively, summarize the relative connectivity of the street network from the perspectives of both automobile and pedestrian travel.

Transit Service

One of the most challenging aspects of this study was collecting uniform information about transit service for communities across the country. The data gathered to calculate these variables fall into two categories. First, we obtained the locations of all fixed-guideway transit stations from the Center for Transit Oriented Development (CTOD, 2011), including all rail lines, streetcars, ferries, trolleys, and some bus rapid-transit systems. Second, we obtained transit-service data (stop locations, routes, and schedules) for 228 local and regional transit agencies[4] across the nation, which was possible only because these agencies all publicly shared their data in the common general transit feed specification (GTFS)[5] format. In general, most large transit agencies in transit-rich regions share their data in GTFS format, but many smaller transit agencies do not; therefore, transit-service data from these agencies are missing from the SLD.[6] Ramsey and Bell (2013) provided a full listing of agencies included in the analysis, organized by metropolitan region served.

Destination Accessibility

Destination accessibility refers to the ease of reaching activities (jobs or workforce) from a given location. We calculated auto accessibility values for all block groups in the United States and transit accessibility values for areas served by transit agencies that share GTFS data.

The accessibility concept requires an understanding of travel times between block group locations. Destinations within a given travel budget are considered “accessible” from the origin, and activities at each accessible destination are discounted according to the time it takes to reach them. We used the NAVTEQ streets data (NAVSTREETS) to assess drive times from each block group centroid in the country to all potential block group destination centroids via street network, capping the search for destinations at 45 minutes.[7] Speed of travel was determined by NAVTEQ’s “Speed Category” field. Therefore, drive times estimate freeflow speeds on each roadway, with no attention to congestion effects. Although this analysis was a data-intensive undertaking, it was relatively straightforward in application.

The transit accessibility analysis was carried out in a similar fashion to the auto analysis, using the NAVSTREETS network to model walk times and GTFS schedules to assess the in-vehicle portions of transit trips to find the shortest travel times between block group origin-destination (OD) pairs.

[4] A full listing of transit agencies with data reflected in the SLD is available in Ramsey and Bell (2013), appendix A, at https://edg.epa.gov/data/Public/OP/SLD/SLD_UserGuide.pdf.

[5] More information about GTFS is available at https://developers.google.com/transit/gtfs/. Agencies can post raw GTFS files for public download on the GTFS data exchange (http://www.gtfs-data-exchange.com/). A full listing of agencies that do and do not share their data in GTFS format is available at City-Go-Round (http://www.citygoround.org/agencies/).

[6] An analysis of data from the National Transit Database showed that transit agencies with GTFS data reflected in the SLD account for 88 percent of all transit ridership in the United States. See Ramsey and Bell (2013) for details.

[7] Block group-weighted centroids—point locations that approximate the center of population within a block group—were obtained from the U.S. census.

Transit analyses are inherently more complex, however, because of the interplay of the transit and pedestrian networks, and we faced more numerous and more daunting challenges in running this analysis.

To evaluate transit accessibility, we (1) estimated walk access and egress times between a block group centroid and accessible transit stops, (2) imposed a standard wait time to board a transit vehicle, (3) enforced limitations on how far a traveler would walk and how long he or she might wait to make a transfer, and (4) assessed the competitiveness of walking from one block group to another as an alternative to taking transit for very short trips. Although these topics would merit more detailed attention in a regional study, we accepted some general rules for our nationwide analysis, imposing a constant 5-minute wait time to board the first transit vehicle, allowing for up to 10 minutes to wait for a transfer (5 of which may be used to walk to the transfer opportunity), and limiting the analysis of travel itineraries to a maximum of one transfer. We also faced two major challenges in assessing travel times between block groups for the entire country.

First, we needed to consolidate all GTFS files into a single table of stop locations and stop events (a stop event is a scheduled boarding or alighting opportunity associated with a particular stop location and a specific transit vehicle trip). To keep this table manageable in size and scope, we combined information only on routes and stops that operate on weekdays during the evening peak period (defined as 5 to 7 p.m.). This approach enabled us to analyze interactions among transit properties that produce separate datasets, even though they have overlapping service areas. Although such an exercise is straightforward for one or two GTFS datasets, working with several hundred is more challenging. We created scripts to pull the relevant data out of each GTFS directory and assemble them in a consolidated table with unique identifiers for stops, trips, agencies, and so on. We encountered issues, however, that required manual intervention throughout, the most common of which were related to how various agencies specified calendar dates in the GTFS tables.[8]

After assembling a single nationwide transit schedule, we set out to determine the shortest travel time between transit stop locations, allowing for 45 minutes of in-vehicle travel time. Whereas most GTFS analyses are built with a definite start time of the trip in mind, we needed to analyze all potential itineraries in the evening peak period to identify the transit trip (or combination of two trips) that provided the shortest travel time between two stop events. This approach often resulted in numerous redundant itineraries for a single stop-event pair, generating very large datasets. We ultimately analyzed about 12.5 trillion itineraries between stop events (about 620 gigabytes of data). We pared these data down by relating each stop event to its location and finding the shortest travel time between stop locations. Finally, we cross-referenced the stop locations OD matrix with walk access and egress times to obtain the shortest transit travel times between block group pairs.

The transit accessibility analysis was conducted for the evening peak period. Several examples of places, however, are served only by morning peak-period service toward downtown and evening peak-period service away from downtown. To emulate the morning peak travel period using only

[8] In some instances, agencies have a service code for all regular weekday service. In other cases, service codes specify each day of the week, and, in other cases, service codes are specified by date, such that each calendar date is addressed differently. For the latter cases, we took services on Wednesdays (whether coded by day of week or by calendar date) to be "typical weekdays."

the evening peak-period data and analysis, we assumed that if a traveler could go from downtown to a suburban residential area in the evening peak, they could also go from the residential area to downtown in the morning peak. Therefore, we also analyzed travel times from destination block group to origin block group to ensure that our overall estimates of transit accessibility were not biased by the expected directionality of service in the evening peak.

Accessing the Smart Location Database

The SLD is a free resource available to the public via download, web service, or interactive map viewer. Data for the entire nation can be downloaded in tabular (.dbf), shapefile, or Esri geodatabase formats. Users who want to download data for only a single state, metropolitan region, or locality can do so by using EPA's Clip N Ship tool. Information about all access options is available on the SLD website: http://epa.gov/smartgrowth/smartlocationdatabase.htm.

Data Currency and Suitability

The SLD reflects housing, population, and employment conditions in 2010, street network conditions in 2011, and transit-service conditions in late 2012. It provides a consistent and generally reliable snapshot of built environment and accessibility characteristics for neighborhoods across the United States. The SLD is not suitable for studies that require knowing the very latest conditions in a given neighborhood, particularly in regions that are experiencing rapid changes because of new construction, migration, or transit-service alterations. EPA hopes to update this database regularly—at least every decennial census. Such plans are contingent on the continued availability of funding, however. Methodologies for all variable calculations are published on line to enable others to develop their own updates.

Example Application: Federal Facility Siting

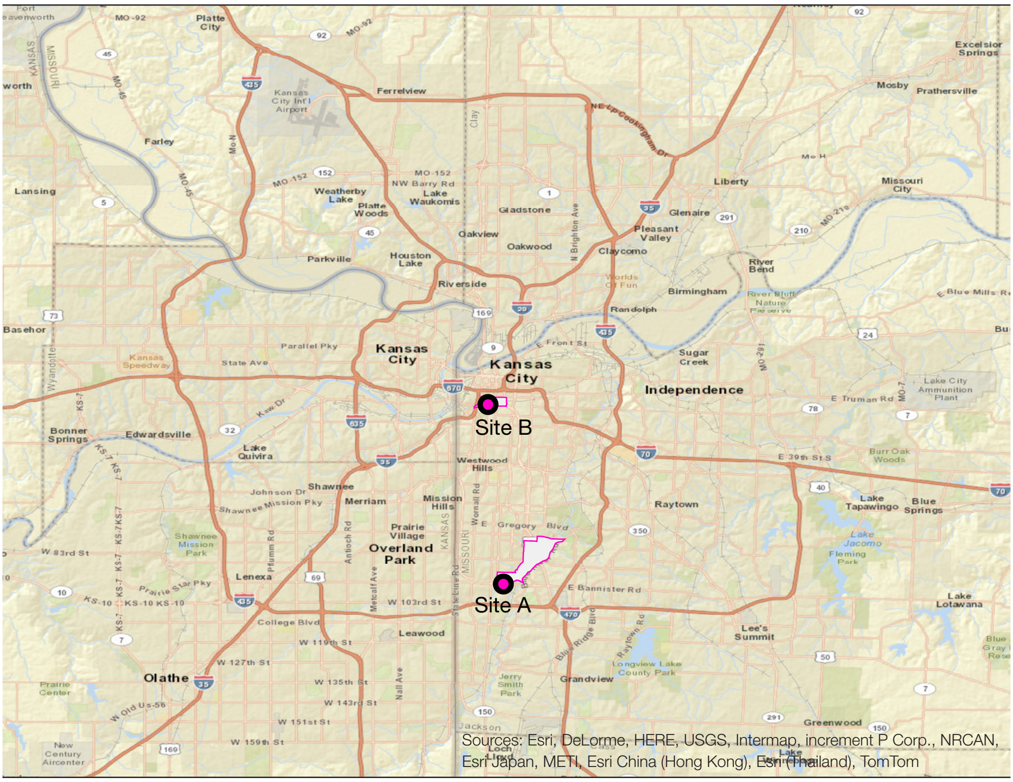
The White House Executive Order, "Federal Leadership in Environmental, Energy, and Economic Performance,"[9] called on federal agencies to consider the sustainability of locations in facility site-selection decisions. To implement this order, the U.S. General Services Administration is drawing on the SLD to develop several new key performance indicators for comparing the sustainability characteristics of existing and proposed facility locations, primarily in terms of workers' commute travel. These indicators measure worksite neighborhood characteristics at the block group scale and can be accessed through an interactive mapping tool available to facility managers. Some indicators are pulled directly from the SLD, and others are modeled based on the results of a nationwide study to measure the effect of workplace neighborhood characteristics (as measured by the SLD) on workers' travel behavior.

[9] Executive Order Number 13514 (2009), "Federal Leadership in Environmental, Energy, and Economic Performance." See http://www.whitehouse.gov/assets/documents/2009fedleader_eo_rel.pdf.

GSA recently used these performance indicators to estimate the environmental benefits of its recent decision to move an existing facility 12 miles south of Kansas City, Missouri, to a new facility near the downtown core of the city. Exhibit 1 shows the location of the two facility sites.

Exhibit 1

Kansas City, Missouri Facility Locations Analyzed for Performance Comparison



Sources: Esri; U.S. Census Bureau; U.S. Geological Survey

This comparison of facility performance (exhibit 2) revealed that Site B (the new facility near downtown Kansas City) performs better than both Site A and the regional average with regard to all indicators. In particular, noncommute vehicle miles traveled (VMT), comprising day trips such as lunch and errands, are estimated to be less than one-half of those of the suburban facility, in large part because of the greater density and diversity of employment in the surrounding neighborhood (indicators that more destinations are within walking distance). Overall, Site B is estimated to generate 22 percent less VMT and associated greenhouse gas emissions per worker than Site A generates.

Exhibit 2

Location Efficiency Comparison of Kansas City, Missouri Facility Sites and Regional Benchmarks

Performance Indicator	Site A	Site B	Regional Benchmarks		
			Average	Highest Performing	Lowest Performing
Transportation (daily per worker)					
Total VMT	26.46	20.69	24.36	17.85	34.81
Commute VMT	20.20	18.06	19.27	16.26	25.41
Noncommute VMT	6.26	2.62	5.09	1.59	9.39
Transportation-related greenhouse gas emissions per worker (lbs) compared with regional average	+ 512	– 888	0	– 1,575	+ 2,535
Monetized impact or benefits to workforce (annual per worker) compared with regional average					
Mobility cost ($)	+ 191	– 332	0	– 589	+ 949
Fuel cost ($)	+ 85	– 147	0	– 261	+ 419
Highway safety cost ($)	+ 99	– 172	0	– 306	+ 493
Efficiency and reliability cost ($)	+ 18	– 32	0	– 56	+ 91
Neighborhood characteristics					
Proximity to nearest transit stop (miles)	0.39	0.19	NA	0.13	No transit within 0.75 mile
Employment within 0.50 mile of fixed-guideway transit station (%)	0	83	11	100	0
Accessibility by workforce via transit (% of regional maximum)	10	37	10	32	0

lbs = pounds. NA = not available. VMT = vehicle miles traveled.

Appendix

Variables Included in the Smart Location Database (1 of 9)

Field	Description	Data Source(s)	Coverage
Administrative			
GEOID10	CBG 12-digit FIPS code.	2010 census TIGER/Line	Entire United States
TRACTCE10	Census tract FIPS code in which CBG resides.	2010 census TIGER/Line	Entire United States
CFIPS	County FIPS code.	2010 census TIGER/Line	Entire United States
SFIPS	State FIPS code.	2010 census TIGER/Line	Entire United States
CSA	CSA code.	2010 census	Entire United States
CSA_Name	Name of CSA in which CBG resides.	2010 census	Entire United States
CBSA	FIPS code for CBSA in which CBG resides.	2010 census	Entire United States
CBSA_Name	Name of CBSA in which CBG resides.	2010 census	Entire United States

Appendix

Variables Included in the Smart Location Database (2 of 9)

Field	Description	Data Source(s)	Coverage
CBSA-wide statistics (same value for all block groups within the same CBSA [metropolitan area])			
CBSA_Pop	Total population in CBSA.	2010 census	Entire United States
CBSA_Emp	Total employment in CBSA.	Census LEHD, 2010	Entire United States (except PR)
CBSA_Wrk	Total number of workers that live in CBSA.	Census LEHD, 2010	Entire United States (except PR)
Area			
Ac_Tot	Total geometric area of the CBG.	2010 census TIGER/Line	Entire United States
Ac_Unpr	Total land area in acres that is not protected from development (that is, not a park or conservation area).	2010 census; NAVTEQ parks; PAD-US	Entire United States
Ac_Water	Total water area in acres.	2010 census; NAVTEQ water and oceans	Entire United States
Ac_Land	Total land area in acres.	2010 census; NAVTEQ water and oceans	Entire United States
Demographics			
CountHU	Housing units, 2010.	2010 census	Entire United States
HH	Households (occupied housing units), 2010.	2010 census	Entire United States
TotPop	Population, 2010.	2010 census	Entire United States
P_WrkAge	Percentage of population that is working age, 2010.	2010 census	Entire United States
AutoOwn0	Number of households in CBG that own zero automobiles, 2010.	ACS; 2010 census	Entire United States
Pct_AO0	Percentage of zero-car households in CBG.	ACS	Entire United States
AutoOwn1	Number of households in CBG that own one automobile, 2010.	ACS; 2010 census	Entire United States
Pct_AO1	Percentage of one-car households in CBG.	ACS	Entire United States
AutoOwn2p	Number of households in CBG that own two or more automobiles, 2010.	ACS; 2010 census	Entire United States
Pct_AO2p	Percentage of two-plus-car households in CBG.	ACS	Entire United States
Workers	Number of workers in CBG (home location), 2010.	Census LEHD, 2010	Entire United States (except PR)
R_LowWageWk	Number of workers earning $1,250 per month or less (home location), 2010.	Census LEHD, 2010	Entire United States (except PR)

Appendix

Variables Included in the Smart Location Database (3 of 9)

Field	Description	Data Source(s)	Coverage
R_MedWageWk	Number of workers earning more than $1,250 per month but less than $3,333 per month (home location), 2010.	Census LEHD, 2010	Entire United States (except PR)
R_HiWageWk	Number of workers earning $3,333 per month or more (home location), 2010.	Census LEHD, 2010	Entire United States (except PR)
R_PctLowWage	Percentage of R_LowWageWk of Workers in a CBG (home location), 2010.	Census LEHD, 2010	Entire United States (except PR)
Employment			
TotEmp	Total employment, 2010.	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E5_Ret10	Retail jobs within a five-tier employment classification scheme (LEHD: CNS07).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E5_Off10	Office jobs within a five-tier employment classification scheme (LEHD: CNS09 + CNS10 + CNS11 + CNS13 + CNS20).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E5_Ind10	Industrial jobs within a five-tier employment classification scheme (LEHD: CNS01 + CNS02 + CNS03 + CNS04 + CNS05 + CNS06 + CNS08).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E5_Svc10	Service jobs within a five-tier employment classification scheme (LEHD: CNS12 + CNS14 + CNS15 + CNS16 + CNS19).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E5_Ent10	Entertainment jobs within a five-tier employment classification scheme (LEHD: CNS17 + CNS18).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E8_Ret10	Retail jobs within an eight-tier employment classification scheme (LEHD: CNS07).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E8_Off10	Office jobs within an eight-tier employment classification scheme (LEHD: CNS09 + CNS10 + CNS11 + CNS13).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E8_Ind10	Industrial jobs within an eight-tier employment classification scheme (LEHD: CNS01 + CNS02 + CNS03 + CNS04 + CNS05 + CNS06 + CNS08).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)

Appendix

Variables Included in the Smart Location Database (4 of 9)

Field	Description	Data Source(s)	Coverage
E8_Svc10	Service jobs within an eight-tier employment classification scheme (LEHD: CNS12 + CNS14 + CNS19).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E8_Ent10	Entertainment jobs within an eight-tier employment classification scheme (LEHD: CNS17 + CNS18).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E8_Ed10	Education jobs within an eight-tier employment classification scheme (LEHD: CNS15).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E8_Hlth10	Healthcare jobs within an eight-tier employment classification scheme (LEHD: CNS16).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E8_Pub10	Public administration jobs within an eight-tier employment classification scheme (LEHD: CNS20).	Census LEHD, 2010; InfoUSA, 2011 (MA only)	Entire United States (except PR)
E_LowWageWk	Number of workers earning $1,250 per month or less (work location), 2010.	Census LEHD, 2010	Entire United States (except MA and PR)
E_MedWageWk	Number of workers earning more than $1,250 per month but less than $3,333 per month (work location), 2010.	Census LEHD, 2010	Entire United States (except MA and PR)
E_HiWageWk	Number of workers earning $3,333 per month or more (work location), 2010.	Census LEHD, 2010	Entire United States (except MA and PR)
E_PctLowWage	Percentage of LowWageWk of Workers in a CBG (work location), 2010.	Census LEHD, 2010	Entire United States (except MA and PR)
D1—Density			
D1a	Gross residential density (HU/acre) on unprotected land.	Derived from other SLD variables	Entire United States
D1b	Gross population density (people/acre) on unprotected land.	Derived from other SLD variables	Entire United States
D1c	Gross employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1c5_Ret10	Gross retail (five-tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1c5_Off10	Gross office (five -tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)

Appendix

Variables Included in the Smart Location Database (5 of 9)

Field	Description	Data Source(s)	Coverage
D1c5_Ind10	Gross industrial (five -tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1c5_Svc10	Gross service (five-tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1c5_Ent10	Gross entertainment (five-tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1c8_Ret10	Gross retail (eight-tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1c8_Off10	Gross office (eight-tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1c8_Ind10	Gross industrial (eight-tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1c8_Svc10	Gross service (eight-tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1c8_Ent10	Gross entertainment (eight-tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1c8_Ed10	Gross education (eight-tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1c8_Hlth10	Gross healthcare (eight-tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1c8_Pub10	Gross retail (eight-tier) employment density (jobs/acre) on unprotected land.	Derived from other SLD variables	Entire United States (except PR)
D1d	Gross activity density (employment + HUs) on unprotected land.	Derived from other SLD variables	Entire United States (PR does not reflect employment)
D1_Flag	Flag indicating that density metrics are based on total CBG land acreage rather than unprotected acreage.	Derived from other SLD variables	Entire United States (PR does not reflect employment)

Appendix

Variables Included in the Smart Location Database (6 of 9)

Field	Description	Data Source(s)	Coverage
D2—Diversity			
D2a_JpHH	Jobs per household.	Derived from other SLD variables	Entire United States (except PR)
D2b_E5Mix	Five-tier employment entropy (denominator set to observed employment types in the CBG).	Derived from other SLD variables	Entire United States (except PR)
D2b_E5MixA	Five-tier employment entropy (denominator set to the static five employment types in the CBG).	Derived from other SLD variables	Entire United States (except PR)
D2b_E8Mix	Eight-tier employment entropy (denominator set to observed employment types in the CBG).	Derived from other SLD variables	Entire United States (except PR)
D2b_E8MixA	Eight-tier employment entropy (denominator set to the static eight employment types in the CBG).	Derived from other SLD variables	Entire United States (except PR)
D2a_EpHHm	Employment and household entropy.	Derived from other SLD variables	Entire United States (except PR)
D2c_TrpMx1	Employment and household entropy (based on vehicle trip production and trip attractions including all five employment categories).	Derived from other SLD variables	Entire United States (except PR)
D2c_TrpMx2	Employment and household entropy calculations, based on trips production and trip attractions including four of the five employment categories (excluding industrial).	Derived from other SLD variables	Entire United States (except PR)
D2c_TripEq	Trip productions and trip attractions equilibrium index; the closer to 1, the more balanced the trip making.	Derived from other SLD variables	Entire United States (except PR)
D2r_JobPop	Regional diversity. Standard calculation based on population and total employment: deviation of CBG ratio of jobs/population from regional average ratio of jobs/population.	Derived from other SLD variables	Entire United States (except PR)
D2r_WrkEmp	Household workers per job, as compared with the region: deviation of CBG ratio of household workers/job from regional average ratio of household workers/job.	Derived from other SLD variables	Entire United States (except PR)
D2a_WrkEmp	Household workers per job, by CBG.	Derived from other SLD variables	Entire United States (except PR)
D2c_WrEmIx	Household workers per job equilibrium index; the closer to one the more balanced the resident workers and jobs in the CBG.	Derived from other SLD variables	Entire United States (except PR)

Appendix

Variables Included in the Smart Location Database (7 of 9)

Field	Description	Data Source(s)	Coverage
D3—Design			
D3a	Total road network density.	NAVSTREETS	Entire United States
D3aao	Network density in terms of facility miles of auto-oriented links per square mile.	NAVSTREETS	Entire United States
D3amm	Network density in terms of facility miles of multimodal links per square mile.	NAVSTREETS	Entire United States
D3apo	Network density in terms of facility miles of pedestrian-oriented links per square mile.	NAVSTREETS	Entire United States
D3b	Street intersection density (weighted, auto-oriented intersections eliminated).	NAVSTREETS	Entire United States
D3bao	Intersection density in terms of auto-oriented intersections per square mile.	NAVSTREETS	Entire United States
D3bmm3	Intersection density in terms of multimodal intersections having three legs per square mile.	NAVSTREETS	Entire United States
D3bmm4	Intersection density in terms of multimodal intersections having four or more legs per square mile.	NAVSTREETS	Entire United States
D3bpo3	Intersection density in terms of pedestrian-oriented intersections having three legs per square mile.	NAVSTREETS	Entire United States
D3bpo4	Intersection density in terms of pedestrian-oriented intersections having four or more legs per square mile.	NAVSTREETS	Entire United States
D4—Transit			
D4a	Distance from population weighted centroid to nearest transit stop (meters).	GTFS; TOD Database 2012	Participating GTFS transit-service areas/ TOD database locations
D4b025	Proportion of CBG employment within 0.25 mile of fixed-guideway transit stop.	TOD Database 2012; SLD unprotected area polygons	Entire United States
D4b050	Proportion of CBG employment within 0.5 mile of fixed-guideway transit stop.	TOD Database 2012; SLD unprotected area polygons	Entire United States

Appendix

Variables Included in the Smart Location Database (8 of 9)

Field	Description	Data Source(s)	Coverage
D4c	Aggregate frequency of transit service within 0.25 mile of block group boundary per hour during evening peak period.	GTFS	Participating GTFS transit-service areas
D4d	Aggregate frequency of transit service (D4c) per square mile.	Derived from other SLD variables	Participating GTFS transit-service areas
D5—Destination accessibility			
D5ar	Jobs within 45 minutes auto travel time, time decay (network travel time) weighted.	NAVSTREETS	Entire United States (except PR)
D5ae	Working-age population within 45 minutes auto travel time, time decay (network travel time) weighted.	NAVSTREETS	Entire United States
D5br	Jobs within 45-minute transit commute, distance decay (walk network travel time, GTFS schedules) weighted.	NAVSTREEETS; GTFS	Participating GTFS transit-service areas (except PR)
D5be	Working-age population within 45-minute transit commute, time decay (walk network travel time, GTFS schedules) weighted.	NAVSTREETS; GTFS	Participating GTFS transit-service areas
D5cr	Proportional accessibility to regional destinations—auto: employment accessibility expressed as a ratio of total MSA accessibility.	Derived from other SLD variables	Entire United States (except PR)
D5cri	Regional centrality index—auto: CBG D5cr score relative to maximum CBSA D5cr score.	Derived from other SLD variables	Entire United States
D5ce	Proportional accessibility to regional destinations—auto: working-age population accessibility expressed as a ratio of total CBSA accessibility.	Derived from other SLD variables	Entire United States
D5cei	Regional centrality index—auto: CBG D5ce score relative to max CBSA D5ce score.	Derived from other SLD variables	Entire United States
D5dr	Proportional accessibility of regional destinations—transit: employment accessibility expressed as a ratio of total MSA accessibility.	Derived from other SLD variables	Participating GTFS transit-service areas
D5dri	Regional centrality index—transit: CBG D5dr score relative to maximum CBSA D5dr score.	Derived from other SLD variables	Participating GTFS transit-service areas

Appendix

Variables Included in the Smart Location Database (9 of 9)

Field	Description	Data Source(s)	Coverage
D5de	Proportional accessibility of regional destinations—transit: working-age population accessibility expressed as a ratio of total MSA accessibility.	Derived from other SLD variables	Participating GTFS transit-service areas
D5dei	Regional centrality index—transit: CBG D5de score relative to maximum CBSA D5de score.	Derived from other SLD variables	Participating GTFS transit-service areas

ACS = 2010 American Community Survey 5-Year estimates. CBG = census block group. CBSA = core based statistical area. CSA = combined statistical area. FIPS = Federal Information Processing Standard. GTFS = general transit feed specification. HU = housing units. LEHD = Longitudinal Employer-Household Dynamics. MA = Massachusetts. MSA = metropolitan statistical area. PAD-US = Protected Areas Database of the United States. PR = Puerto Rico. SLD = Smart Location Database. TOD = transit-oriented development.

Acknowledgments

The authors acknowledge Amy Smith, Jerry Walters, Jennifer Ziebarth, and Gustavo Jimenez at Fehr and Peers Transportation Consultants; and the U.S. General Services Administration Urban Development/Good Neighbor Program, the work and support of which contributed to the facility-siting case study and performance indicators.

Authors

Kevin Ramsey is a senior associate at BERK Consulting. He managed the development of the Smart Location Database while participating in a fellowship at the Environmental Protection Agency.

Alexander Bell is a senior planner at Renaissance Planning Group. He conducted much of the data analysis required to develop the Smart Location Database.

References

Bartholomew, Keith, and Reid Ewing. 2008. "Land Use—Transportation Scenarios and Future Vehicle Travel and Land Consumption: A Meta-Analysis," *Journal of the American Planning Association* 75: 13–27.

Center for Transit Oriented Development (CTOD). 2011. "TOD Database." http://toddata.cnt.org/.

de Nazelle, Audrey, Mark J. Nieuwenhuijsen, Josep M. Antó, Michael Brauer, David Briggs, Charlotte Braun-Fahrlander, Nick Cavill, Ashley R. Cooper, Hélène Desqueyroux, Scott Fruin, Gerard Hoek, Luc Int Panis, Nicole Janssen, Michael Jerrett, Michael Joffe, Zorana Jovanovic Andersen, Elise van Kempen, Simon Kingham, Nadine Kubesch, Kevin M. Leyden, Julian D. Marshall, Jaume Matamala, Giorgos Mellios, Michelle Mendez, Hala Nassif, David Ogilvie, Rosana Peiró, Katherine Pérez, Ari Rabl, Martina Ragettli, Daniel Rodríguez, David Rojas, Pablo Ruiz, James F. Sallis, Jeroen Terwoert, Jean-François Toussaint, Jouni Tuomisto, Moniek Zuurbier, and Erik Lebret. 2011. "Improving Health Through Policies That Promote Active Travel: A Review of Evidence To Support Integrated Health Impact Assessment," *Environment International* 37 (4): 766–777. DOI: http://dx.doi.org/10.1016/j.envint.2011.02.003.

Ewing, Reid, and Robert Cervero. 2010. "Travel and the Built Environment," *Journal of the American Planning Association* 76 (3): 265–294.

Ewing, Reid, Michael Greenwald, Ming Zhang, Jerry Walters, Robert Cervero, Lawrence Frank, Senait Kassa, and John Thomas. 2011. "Traffic Generated by Mixed-Use Developments—Six-Region Study Using Consistent Built Environmental Measures," *Journal of Urban Planning and Development* 137 (3): 248–261. DOI: http://dx.doi.org/10.1061/(ASCE)UP.1943-5444.0000068.

Gulden, Jeff, J. P. Goates, and Reid Ewing. 2013. "Mixed-Use Development Trip Generation Model," *Transportation Research Record* 2344: 98–106. DOI: http://dx.doi.org/10.3141/2344-11.

Haas, Peter, Carrie Makarewicz, Albert Benedict, and Scott Bernstein. 2008. "Estimating Transportation Costs by Characteristics of Neighborhood and Household," *Transportation Research Record* 2077: 62–70. DOI: http://dx.doi.org/10.3141/2077-09.

Ramsey, Kevin, and Alexander Bell. 2013. "Smart Location Database User Guide." http://www.epa.gov/smartgrowth/pdf/sld_userguide.pdf.

Ramsey, Kevin, and Aaron Poresky. 2013. "A Place-Based Tool for Assessing Cumulative Impervious Surface Outcomes of Proposed Development Scenarios," *URISA Journal* 25: 25–38.

U.S. Geological Survey, Gap Analysis Program (USGS GAP). 2012. "Protected Areas Database of the United States (PADUS), version 1.3, Combined Feature Class." http://gapanalysis.usgs.gov/padus/data/.