MF

METROFREIGHT

Volvo Center of Excellence

Data Challenges in Urban Freight Research: The Case of Warehousing and Distribution

Project Number: 14-4.1b
Years: 2017-2018

FINAL REPORT
May 2018

Principal Investigator
Genevieve Giuliano

Researcher
Sanggyun Kang

MetroFreight Center of Excellence
METRANS Transportation Center
Sol Price School of Public Policy
University of Southern California
Los Angeles, California, USA

Data Challenges in Urban Freight Research: The Case of Warehousing and Distribution

ABSTRACT

We are conducting research on the spatial dynamics of warehouse and distribution activities. This paper addresses the problem of data for examining logistics industry trends over space and time. The research requires fine grain data with respect to both location and attributes that is consistent over time. We explored various data sources and compared their reliability and consistency. Our results make it possible to determine trade-offs between spatial scale, industry sector detail, and accuracy across the available data sources. Our results also suggest caution in using establishment level data.

INTRODUCTION

Urban freight research is limited in many ways by lack of data. Detailed information on truck flows within metro areas, types of commodities moved, origin and destination of shipments and many other things are rare, even for large metropolitan areas. Warehouse location is a growing area of research. In the Los Angeles region, the spatial distribution of warehousing and distribution centers (W&Ds) has changed substantially (Giuliano and Kang, 2018, Dablanc and Ross, 2012). W&Ds provide warehousing and logistics services to the logistics industry and are located along industry supply chains. Hence, their spatial shifts might also imply spatial shifts in freight flows. The first step to evaluate freight flows originating from and destined to W&Ds may be to precisely quantify where the W&Ds are located. Yet basic information at the establishment level is difficult to obtain. In this paper, we examine data sources for warehouse and distribution activities with regard to granularity, comparability, and accuracy.

In the United States (US) only two public datasets with coverage of the entire US are available to identify the location of W&Ds: County Business Patterns (CBP) or ZIP Code Business Patterns (ZBP). However, the geographic units (county or ZIP Code) of these datasets are too large for precise location of any given W&D establishment. Furthermore, there is no public data source that provides establishment-level information on the location, size, and characteristics of a facility. Proprietary datasets may be second-best options, but they are subject to reliability issues and are costly (Giuliano et al. 2015). In this paper, we examine the following research questions: How accurately do secondary datasets identify and trace the spatial distribution of W&Ds? How consistent are the secondary datasets with each other? Do the secondary datasets capture all W&D activity? If not, how large is the gap between the secondary datasets and the actual activity?

We assess the comparability among two datasets available to identify economic activities in the US: ZIP Code Business Patterns (ZBP; public economic census) and National Establishments Time Series (NETS; proprietary business data). We further compare COSTAR (proprietary real estate

listings), which provides the most comprehensive list of active W&D facilities in the leasing market, to the actual warehousing and storage activity captured in high resolution satellite imagery. As a case study, we use several sites in the Los Angeles-Long Beach, CA Combined Statistics Area (CSA).

The organization of this paper is as follows. In the following section, we present the research approach and list five categories of information necessary to precisely identify W&D location and its activity level. In section 3, we explore the characteristics of the three datasets, and in section 4, we examine comparability among the datasets and to actual activity. We present our conclusions in section 5.

RESEARCH APPROACH

Datasets to identify W&D activities are limited with respect to attributes and precision. What information do we need to precisely locate a W&D facility and quantify its activity level? We first assume that we are in an ideal world where we know everything, and we lay out necessary attributes to identify W&D activity in five categories – unit of analysis, size, geographic identification, facility type and freight flow.

We further examine the comparability among secondary datasets – one public (CBP/ZBP) and one proprietary (NETS). By definition, a secondary dataset is collected and maintained by a third-party data analyst, who may be either public or private. A primary dataset is collected and maintained by the user through field work, interview, or survey. W&D activities identified by the CBP/ZBP and NETS datasets are comparable because the two commonly use an establishment as the unit of analysis. The two datasets are different in terms of the granularity of the spatial unit, available attributes, and costs for data access. The CBP/ZBP datasets are publicly available and are used most frequently in US W&D studies. The NETS data provides proprietary business information (name, location, employment, industry sector, and revenue) at the establishment level. The NETS is similar to other business datasets, such as Dun and Bradstreet or InfoUSA. COSTAR, also proprietary, is a commercial and industrial real estate listings database. Its unit of analysis is rentable space, and those facilities privately owned and operated by logistics, retail, or wholesale companies are not listed. Hence, COSTAR is not directly comparable to CBP/ZBP or NETS.

We also examine whether the secondary datasets successfully capture all the actual W&D activity and examine if there is a gap. We define the physical characteristics of a building that may be involved with warehousing activity. We manually capture all buildings that have similar characteristics using high-resolution satellite imagery (LAR-IAC). We then compare results with the COSTAR. We delineate a study area in Southeast Los Angeles and conduct a case study. All analysis is conducted in the Los Angeles CSA. We choose Los Angeles because of high concentration of W&D activity, our familiarity with the region, as well as data availability.

INFORMATION NECESSARY TO LOCATE A W&D AND ITS ACTIVITY

In this section we describe the "ideal" dataset, and then discuss what types of data actually exist.

Unit of Analysis

In an ideal world, we would have information on every location where W&D activity takes place or could take place. This means information on every W&D facility, whether owned or leased, and whether currently operating. A W&D facility is typically a stand-alone facility, but W&D operations also take place as part of other operations, for example retail establishments that include W&D operations. Although ideally, we would like to capture all W&D activity, we focus here on facilities, which should be much easier to identify.

Size

We want a measure of intensity of W&D activity so that we may distinguish between small and large facilities. The floor area (square feet) of a W&D facility is commonly used to estimate warehousing capacity. Capacity also depends on ceiling height and number of bays, but to a lesser extent than square footage.

Geographic Identification

Ideal location information is address or point level information. An address can be converted to X-Y coordinates through geo-coding and mapped into a geographical information system (GIS).

Type of Facility

Many different functions of warehousing facilities are involved along goods supply chains. Conventionally, warehouses have provided functions of receiving, storing, packing and shipping of intermediate or final products. In addition, W&Ds provide cross-docking, package consolidation, and multi-channel/online fulfillment services. Ideally, we would have information on the types of functions or services provided by each W&D facility.

Freight Flow

Finally, we are interested in W&Ds as trip generators. The level of W&D activity will be reflected in the volume of goods originating from/destined to a facility. Disaggregated commodity flow information by dollar value or tonnage is ideal. In addition, the location of shippers and receivers as well as the type of freight vehicles traveling to/from the facility are of interest in order to understand the freight flows resulting from the associated supply chains.

Possible Data Sources

Given our ideal dataset, we examined several different data sources. None of them provide all the data items described above, and all have different advantages and disadvantages. We selected the following for our comparative analysis:

- ZBP: Zip Code Business Patterns, produced annually by the U.S. Census Bureau and publicly available via the U.S. Census website

- NETS: National Establishment Time Series, a proprietary dataset produced by Walls and Associates and updated annually
- COSTAR: A commercial real estate proprietary dataset produced by CoStar Group, Inc, available by subscription and updated monthly
- LAR-IAC: Los Angeles Region Imagery Acquisition Consortium high resolution satellite imagery dataset, 2006 and 2011 vintages currently available at the USC Spatial Sciences Institute

The following section describes each of these data sources.

DATA

ZIP Code Business Patterns

ZIP Code Business Patterns (ZBP) is published by the U.S. Census Bureau as part of the County Business Patterns (CBP) datasets. The ZBP offers two kinds of datasets: one for industry totals and the other for industry details. We do not review the industry total dataset because it does not provide industry sector attributes that we require. The industry details dataset is available annually from 1994 to 2015. Available attributes are the number of establishments by six-digit NAICS industry sector by nine establishment size classes by ZIP Code.[1] An establishment is defined as "a single physical location at which business is conducted, or services or industrial operations are performed."[2] ZBP is one of the very few public datasets with detailed sector information: To identify W&Ds, we use a three-digit NAICS definition (493 Warehousing and Storage). The primary data source of CBP/ZBP is the Business Register, which maintains business information of each establishment with an EIN (Employer Identification Number) and a minimum of one employee in the US, Puerto Rico and Island Areas. ZBP is based on United States Postal Service ZIP Codes.

ZBP has several limitations. First, relative to an establishment or a facility, a ZIP Code is a large spatial unit, particularly in areas that are not densely populated. In the Los Angeles-Long Beach, CA CSA, there are approximately 4,000 census tracts and 940 ZIP Codes.[3] Like census tracts, the size of ZIP Codes is related to population density; tracts within urban areas are smaller than tracts outside urban areas. Second, the actual distribution of W&Ds in a ZIP Code is unknown. With no information on the actual spatial distribution, one must make some type of assumption. Typically location is assigned to the ZIP Code centroid. Third, ZIP Code boundaries are an estimate of the aggregation of physical addresses rather than spatially delineated areas. Fourth, ZIP Codes are not consistent with administrative boundaries (municipality, county, or sometimes state). Fifth, sector-level employment counts are available only at the county or state level.

[1] 1-4, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-999, and 1,000 or more
[2] CBP, Census Bureau (http://www.census.gov/econ/cbp/)
[3] The Los Angeles-Long Beach CSA consists of five counties: Los Angeles, Orange, Riverside, San Bernardino, and Ventura.

Finally, ZBP and CBP are subject to inconsistencies from the conversion of industry codes from Standard Industrial Classification (SIC) to North American Industry Classification System (NAICS) in 1998. The conversion resulted in a significant undercount of warehouse establishments, which was corrected with a NAICS revision in 2002. The effects of the conversion are shown in Table 1 and Figure 1 for the Los Angeles CSA. This problem is not unique to Los Angeles; the same pattern is observed in many other metro areas in New York, Texas, and Florida (not presented here). The ZBP annual data is therefore internally consistent only after 2003.

Table 1 N of W&Ds in LA CSA before and after the industry code revision from SIC to NAICS 1998 and NAICS 2003 (CBP stats)

Year	Los Angeles	Orange	Riverside	San Bernardino	Ventura	Total	Note
1995	436	122	54	66	29	707	SIC
1996	434	140	60	80	32	746	
1997	497	148	78	94	35	852	
1998	220	49	15	34	8	326	NAICS
1999	222	49	20	40	9	340	1998
2000	235	51	20	51	10	367	
2001	240	52	23	55	11	381	
2002	249	47	32	68	9	405	
2003	435	93	61	183	16	788	NAICS
(2003 share)	55.2%	11.8%	7.7%	23.2%	2.0%	100.0%	2003
2004	473	100	70	207	18	868	Revision
2005	463	110	74	211	20	878	
2006	478	128	73	216	20	915	
2007	498	122	84	236	27	967	
2008	492	109	85	248	27	961	
2009	515	105	85	252	33	990	
2010	506	105	80	250	30	971	
2011	491	103	79	235	29	937	
2012	479	94	97	263	24	957	
2013	485	94	100	294	26	999	
(2013 share)	48.5%	9.4%	10.0%	29.4%	2.6%	100.0%	
Change (N) 2003-2013	50	1	39	111	10	211	
Change (%) 2003-2013	11.5%	1.1%	63.9%	60.7%	62.5%	26.8%	

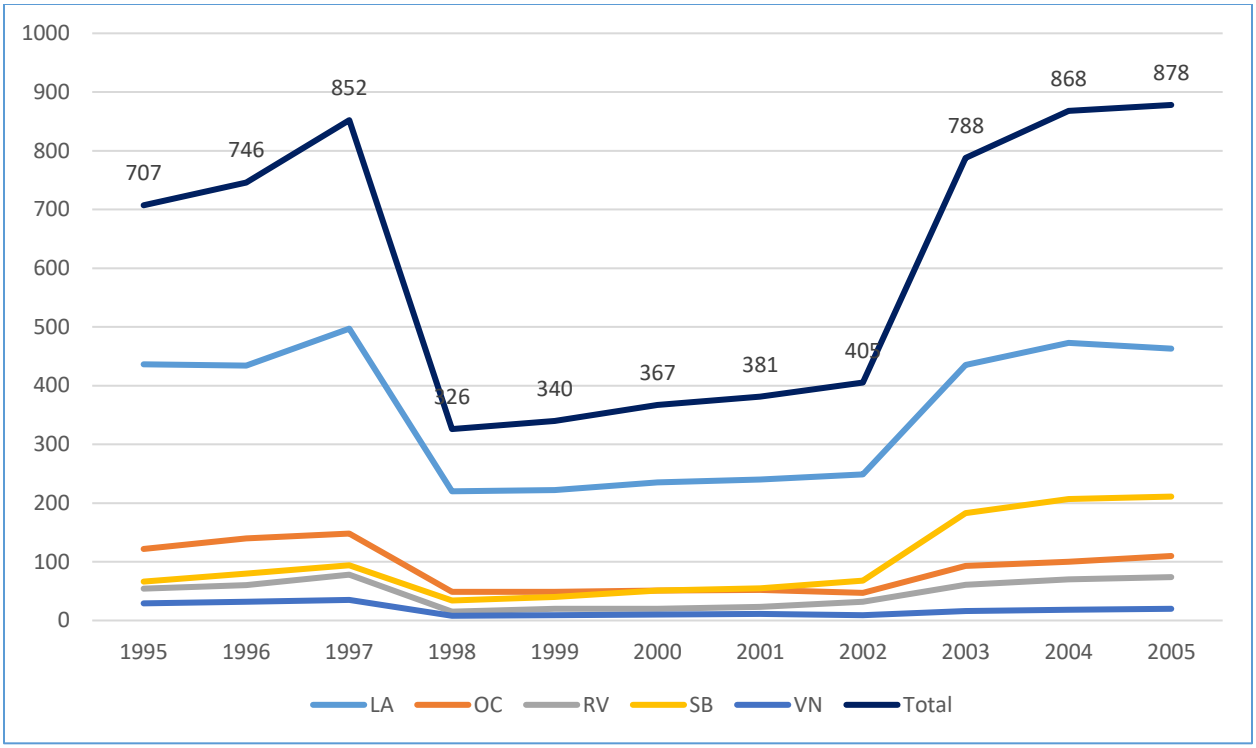

Figure 1 Time-series trends of W&Ds in five counties of the Los Angeles CSA from 1995 to 2005

ZBP or CBP?

W&D location studies in the U.S have primarily used the CBP and ZBP over the last decade. We list five representative studies in Table 2. We make two observations regarding how the recent studies have used the two datasets. First, national studies are all based on the CBP with a county as the unit of analysis (Bowen, 2008; Cidell, 2010; Rivera, et al., 2014), whereas sub-metropolitan studies use the ZBP with a ZIP Code as the unit of analysis (Dablanc and Ross, 2012; Dablanc, et al., 2014; Giuliano and Kang, 2018). Each dataset has its own strength and weakness. The ZBP provides finer geographic details than the CBP. But the ZBP provides very limited information: the number of establishments in 9 size categories by sector by ZIP Code. The CBP has number of establishments in 12 size categories, employment, and annual/quarterly payroll data. Also, the granularity of the spatial units – county and ZIP Code – is different between the two. The trade-off for more data is more spatial aggregation, as counties are generally much larger spatial units than ZIP Codes.[4]

[4] The size difference is greater in the west and Midwest parts of the country, where counties can be very large. For example, the Los Angeles CSA includes just 5 counties.

Table 2 US studies that identified the spatial distribution of W&Ds

Paper	Dataset	Unit of Analysis	Year	Location	Purpose
Bowen (2008)	County Business Patterns	County	1998-2005	143 US metropolitan counties	Factors to explain warehousing expansion
Cidell (2010)	US Economic Census & County Business Patterns	Core-Based Statistical Areas (CBSA)	1986 & 2005	50 CBSAs	Factors to explain warehousing suburbanization
Dablanc and Ross (2012)	ZIP Code Business Patterns	ZIP Code	1998-2008	Atlanta, GA	Examine logistics sprawl and polarization
Dablanc, Ogilvie, and Goodchild (2014)	ZIP Code Business Patterns	ZIP Code	1998-2009	Los Angeles, CA and Seattle, WA	Examine changes in the geographic distribution of warehouses
Rivera, et al. (2014)	County Business Patterns	County	1998 & 2008	US metro areas	Develop methodologies to identify logistics clusters
Giuliano and Kang (2018)	ZIP Code Business Patterns	ZIP Code	2003 & 2013	Los Angeles, San Francisco, San Diego, and Sacramento, CA	Examine changes in the geographic distribution of warehouses

*Studies in chronological order

National Establishment Time-Series

The National Establishment Time-Series (NETS) dataset is available from Walls and Associates and is based on the Dun and Bradstreet (D&B) national establishment database. The NETS is an establishment-level, annual time series database starting in 1990. It provides business information such as location, industry sector (6-digit NAICS), number of employees, and the D&B firm evaluation scores. Establishments are located by address and X-Y coordinates. NETS also provides information on business types (stand-alone, headquarters, or branch/divisions) and corporate hierarchies. Because each observation is based on a permanent unique identifier (DUNS number) and the dataset is structured as an annual panel, firm births, deaths, expansions, and moves can be traced. Various methodologies are used to keep the database up to date and consistent.

The richness of this database makes it an attractive candidate for examining warehouse location changes. Every warehouse establishment is located at the point level and traceable over time. In theory, it is possible to track increases or reductions in employment, as well as linkages among W&D establishments. However, NETS has problems of both internal and

external consistency. Giuliano, et al. (2015) compared NETS to several other data sets, including CBP, for 1993 through 2009 and found large inconsistencies with both establishment and employment counts of the entire economy even at the county level.

We present by county the statistics of W&D establishments in Table 3. Until 2008, the number of establishments increased substantially and shrank in 2009, which might be attributed to the economic recession in 2008 and 2009. W&Ds in Los Angeles accounted for 59% in 2003, but it decreased. Riverside, San Bernardino, and Ventura showed a distinguishable expansion in the number of W&D establishments. No W&D location study has used the NETS dataset.

Table 3 N of W&D in LA CSA (NETS)

Year	Los Angeles	Orange	Riverside	San Bernardino	Ventura	Total
2003	976	265	135	231	60	1,667
(share)	58.5%	15.9%	8.1%	13.9%	3.6%	100.0%
2004	991	260	143	249	61	1,704
2005	994	258	150	244	61	1,707
2006	1,020	262	162	247	68	1,759
2007	1,116	279	189	268	79	1,931
2008	1,246	320	203	291	84	2,144
2009	962	269	191	244	75	1,741
(share)	55.3%	15.5%	11.0%	14.0%	4.3%	100.0%
Change (N) 2003-2009	-14	4	56	13	15	74
Change (%) 2003-2009	-1.4%	1.5%	41.5%	5.6%	25.0%	4.4%

COSTAR

COSTAR is a real estate database that includes listings for multiple commercial and industrial real estate types. Listings classified as truck terminals, warehouses, distribution centers and cold storage facilities as of 2016 were included in the study.

The COSTAR database allows the researcher to differentiate listings based on building characteristics such as total square feet, number of stories, and number of truck docking positions. While these metrics do not directly convey truck trip generation, they can be used as effective proxies. The COSTAR database also provides year of construction or rehabilitation of individual facilities. This allows researchers to track how a particular urban area's pattern of distribution facilities has evolved over time both in terms of which locations were selected and what attributes were preferred. For example, it can be used to track how the average size of a warehouse within a particular urban area has changed over time, or the difference in average size between warehouses in the core vs the periphery of an urban area. COSTAR provides summary data for user-defined geographies (county, city, ZIP Code, or user-defined geography) in terms of rentable space, average rental rates, and vacancy rates.

The COSTAR data has limitations. First, the purpose of COSTAR is to provide information on industrial and commercial real estate markets to brokers. As such, it regularly updates and maintains listings, but does not keep records that are not currently available. For example, when warehouses are demolished or otherwise leave the market, they are no longer tracked. Hence, it is not possible to reconstruct warehousing listings of the past. We retrieved data in early 2016, hence this analysis reflects the statistics as of the time period. Another weakness of the data is that it does not track the industrial profile of the warehouse user. If the warehouse changes lessors and serves a different function, this will not be reflected in the database. For example, some warehousing and distribution usage types such as container freight stations and bonded warehouses produce a much higher frequency of trip generation when compared with traditional storage warehouses. Table 4 presents the number of W&D facilities by county. Approximately 60% of W&Ds are in Los Angeles, followed by San Bernardino and Orange.

Table 4 N of W&Ds in LA CSA (COSTAR) (only covers those warehouses constructed after 1986 that are 30K square feet or larger)

County	Los Angeles	Orange	San Bernardino	Riverside	Ventura	Total
N	3,454	623	1,207	439	122	5,845
%	59%	11%	21%	8%	2%	100%

In summary, Table 5 presents the characteristics of the three datasets we discussed so far in terms of the five types of information to capture warehousing and distribution activity.

Table 5 Available information from public and proprietary datasets

Type of information	ZBP	NETS	COSTAR
Type of data	Public, economic census	Proprietary, business marketing data	Proprietary, commercial real estate listings
Unit attributes	ZIP Code	Establishment	Individual rentable facility
Location attributes	ZIP Code centroids	Address	Address
Size attributes	Establishment size category	Number of Employment	Rentable building area, total lot size
Type attributes	6-digit NAICS: general, refrigerated, farm product, and other	6-digit NAICS: general, refrigerated, farm product, and other	Truck terminals, warehouses, distribution centers and cold storage
Flow attributes	None	None	None

High-Resolution Satellite Imagery (LAR-IAC)

The LAR-IAC (Los Angeles Region Imagery Acquisition Consortium) Program is a consortium of multiple Los Angeles County departments, municipalities, and public agencies. It is a collective effort to collect high-resolution satellite imagery data. The USC Spatial Science Institute

participated in its third phase (LAR-IAC 3 in 2011) and has maintained the dataset. We acquired part of the third phase deliverables (one-foot orthophotos, 2011 dataset) of the study area – an area encompassed by highways I-110, I-710, I-405, and SR 91. This area includes Compton, Carson and Dominguez Hills, CA. See Figure 2. We chose this location because it has the most intense W&D activity in Los Angeles County. We define a warehousing facility as a building that has at least one loading dock (or loading bay) for heavy-duty trucks. We assume that a dedicated loading dock is significant capital investment, which reflects the demand to transport large volumes of freight by heavy-duty truck. Hence, the presence of a loading dock is strongly associated with the likelihood that the building is involved with warehousing activity. We disregard industry sector and land use/zoning ordinances of the parcel but only consider the physical characteristics of a building. Using ArcMAP 10.4.1, we manually captured all buildings that have such characteristics.

Figure 2 Study Area - Southeast Los Angeles

Figure 3 shows the physical characteristics of typical large-scale distribution centers. The building on the left is a local drayage facility, which receives marine containers from the San Pedro Bay Ports. The buildings on the right are a regional distribution center of a cosmetics company (right-side) and a logistics service provider (left-side). In most cases, a warehouse has a one-story structure for the convenient use of platform truck, pallet truck, or forklift. Most W&Ds we captured from the images have many trucks and containers either placed on the loading dock or awaiting to be processed in the parking lot. Some facilities even have trucks queuing around the building on street for loading/unloading.

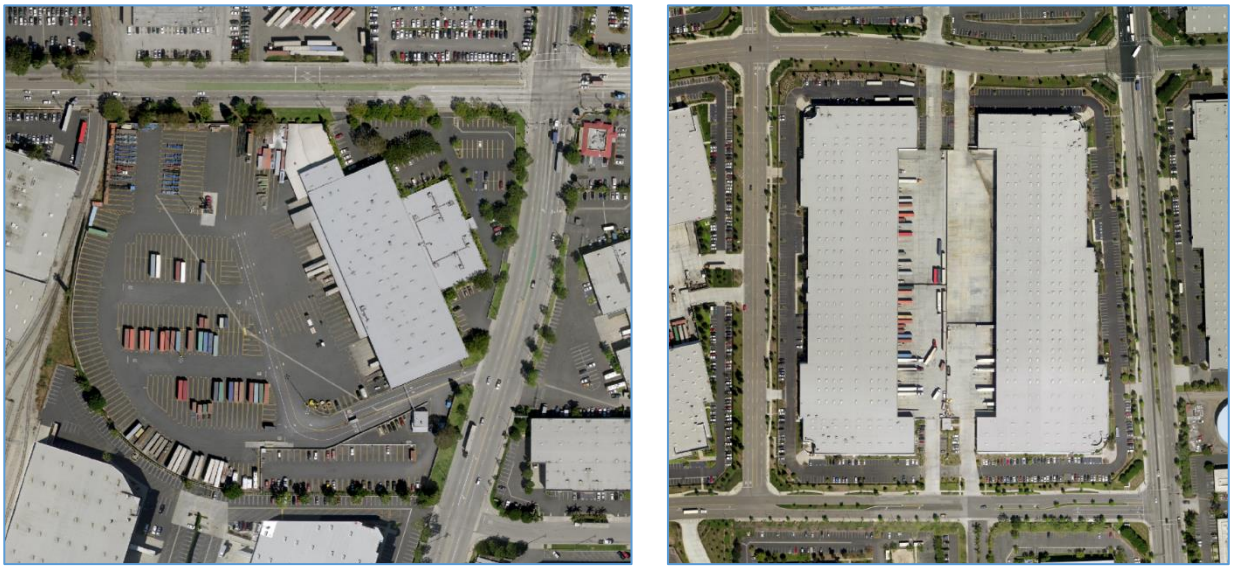

Figure 3 Left: warehouse for local drayage (110,000 sqft); Right: Regional distribution centers (two buildings of 250,000 sqft)

EVALUATION OF EXTERNAL COMPARABILITY

In this section, we compare W&D statistics from public and proprietary datasets, as well as the W&D inventory we identified using satellite imagery. The comparison is made in two parts. First, at the metropolitan level, we compare W&D establishment counts between CBP and NETS in the Los Angeles CSA between 2003 and 2009. The two datasets have a consistent unit of analysis (establishment). Second, we delineate a small area for a case study of W&D activity at the facility level and compare the number of W&D facilities in COSTAR listings to the W&D buildings we identified using LAR-IAC.

Metro Area-Level Comparison: CBP/ZBP and NETS

We present the number of W&Ds in the Los Angeles CSA from 2003 to 2009 in ZBP and NETS. The last year of the NETS dataset we purchased was 2009. As seen in Table 6, the statistics (annual counts and changes in number and percentage) are substantially different between the two datasets. Between the ZBP and NETS, the difference is on the order of one hundred percent: 990 establishments in ZBP, and 1,741 in NETS in 2009. In ZBP, 202 W&Ds were added from 2003 to 2013, while in NETS the net increase is 74. Note also that the changes between some years in NETS is quite large. A detailed comparison between the two datasets at the county level by industry sector is also available in Giuliano et al. (2015). This report also found similar levels of differences between the two datasets.

Table 6 Comparison of the number of W&Ds between ZBP and NETS from 2003 to 2009

Year	ZBP (Establishment)	Percent change	NETS (Establishment)	Percent change
2003	788	-	1,667	-
2004	868	10.2%	1,704	2.2%
2005	878	1.2%	1,707	0.2%
2006	915	4.2%	1,759	3.0%
2007	967	5.7%	1,931	9.8%
2008	961	-0.6%	2,144	11.0%
2009	990	3.0%	1,741	-18.8%
Change (N) 2003-2009	202	25.6%	74	4.4%

How do we explain these very large discrepancies? The ZBP is published by a government agency based on a tax ID (Employer Identification Number). Hence, it is hard to believe that the ZBP is undercounting. It is also hard to believe that NETS is over-counting establishments consistently over multiple years. Other than the explanation that they are counting establishments differently, we found no convincing explanation. For example, the NETS includes sole-proprietor establishments (self-employed or with no employee but the owner), whereas the CBP/ZBP excludes them. However, sole proprietorships cannot be a significant portion of warehouse establishments. It is more likely that establishments are counted differently. In ZBP, an establishment is defined as a single physical location; it is possible that the NETS accounting for branches or divisions inflates the numbers. In any case, it is clear that the data sets are quite different and that there is no a priori reason to judge one more accurate than the other.

COSTAR and LAR-IAC

From the LAR-IAC high resolution satellite imagery, we identified 406 facilities with at least one loading dock in 2011 from the study area (Table 7). Figure 4 shows the spatial distribution of the buildings as well as 84 image tiles (0.5 * 0.5 miles) of the LAR-IAC. As opposed to this, in COSTAR, only 329 facilities were listed in 2011. The 2011 listings are from the 2016 data minus buildings added (built) between 2012 and 2016. The difference in terms of the rentable building area between the two datasets (77 facilities) was approximately 8 million square feet. The mean size of a building is similar. The difference between these data sets is expected, since COSTAR only includes properties in the rental or leasing markets.

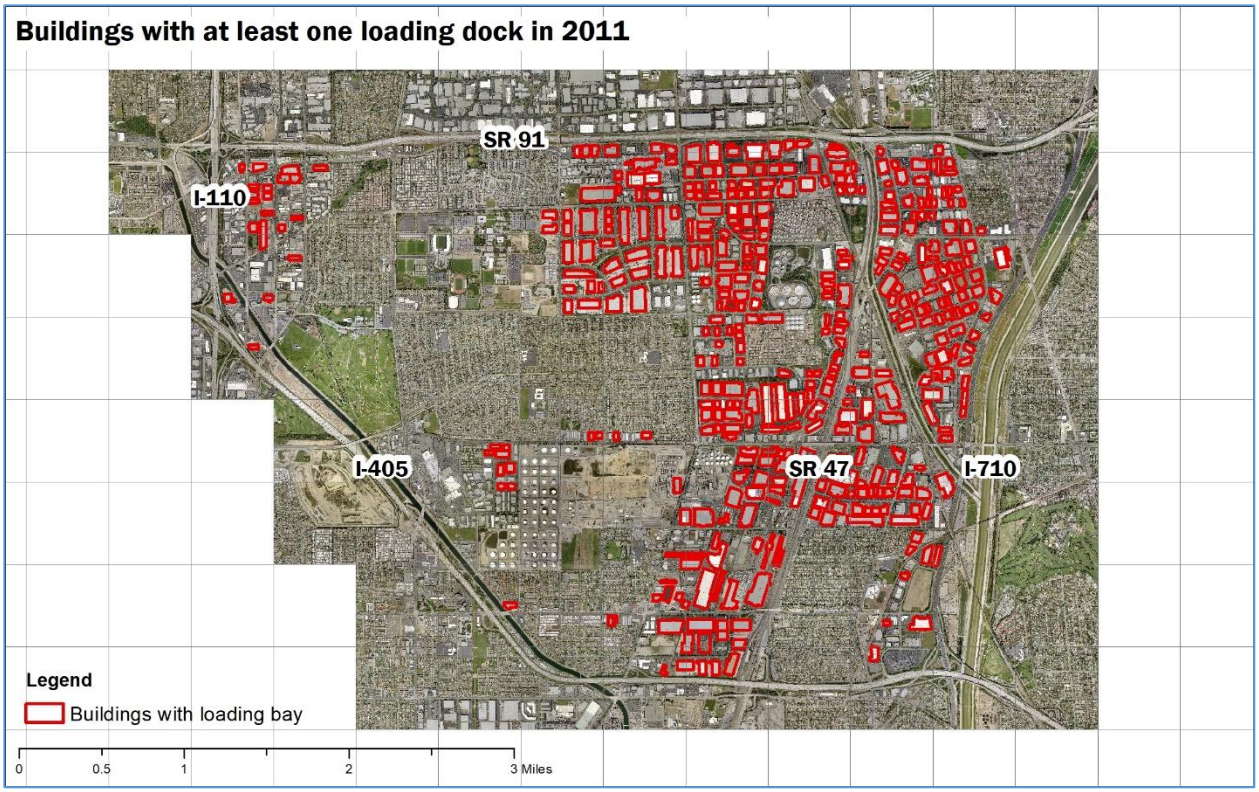

Figure 4 Buildings with a loading dock captured in 84 tiles of high resolution satellite imagery (LAR-IAC, 2011)

Table 7 N and sum of building area of the facilities available in COSTAR and identified in LAR-IAC.

Type	N	Sum	Mean	SD
COSTAR	329	34,055,216	103,511	75,830
LAR-IAC	406	42,115,780	103,733	76,261
Difference	77	8,060,564	Not significant	-

Our definition of a warehouse – a building with a loading dock – is quite a broad term, and it is likely that it will over-count and include those buildings not exclusively functioning as a warehouse. The study area has plentiful light manufacturing, logistics, wholesale trade, and warehousing businesses. Perhaps, for example, a light manufacturing plant might utilize part of its building as storage. In order to receive and ship out intermediate and manufactured goods at a low cost, the building might also have a loading dock. In this case, the business, not necessarily a warehousing establishment, will be classified under the manufacturing sector and perform manufacturing and warehousing functions. We expect that manufacturing, wholesale, and retail trade businesses in this study area are likely to utilize part of their building for warehousing and storage purposes. We further examine the disparity using a more restrictive definition of a warehouse: buildings larger than 100,000 square feet with multiple loading docks.

We identified 155 buildings from LAR-IAC. Of these buildings, 35 buildings were not listed in COSTAR. We used Google Maps to search for the name, business website, and business type of the missing buildings one by one. Figure 5 shows buildings larger than 100,000 square feet as well as those that were not listed in COSTAR. We confirmed that 20 buildings (green highlight in Table 8) were occupied by logistics/freight transportation businesses (sum of the building area: 3.24 million square feet). Those buildings may be private property directly owned and operated by the businesses. If the buildings are not in the rental or leasing market, they are not listed in COSTAR. The other 15 buildings were used by manufacturing, retail trade, and wholesale trade businesses. We could not check further on whether the whole or part of the building is utilized as a warehouse or how much of the space is allocated for warehousing and freight transportation purposes.

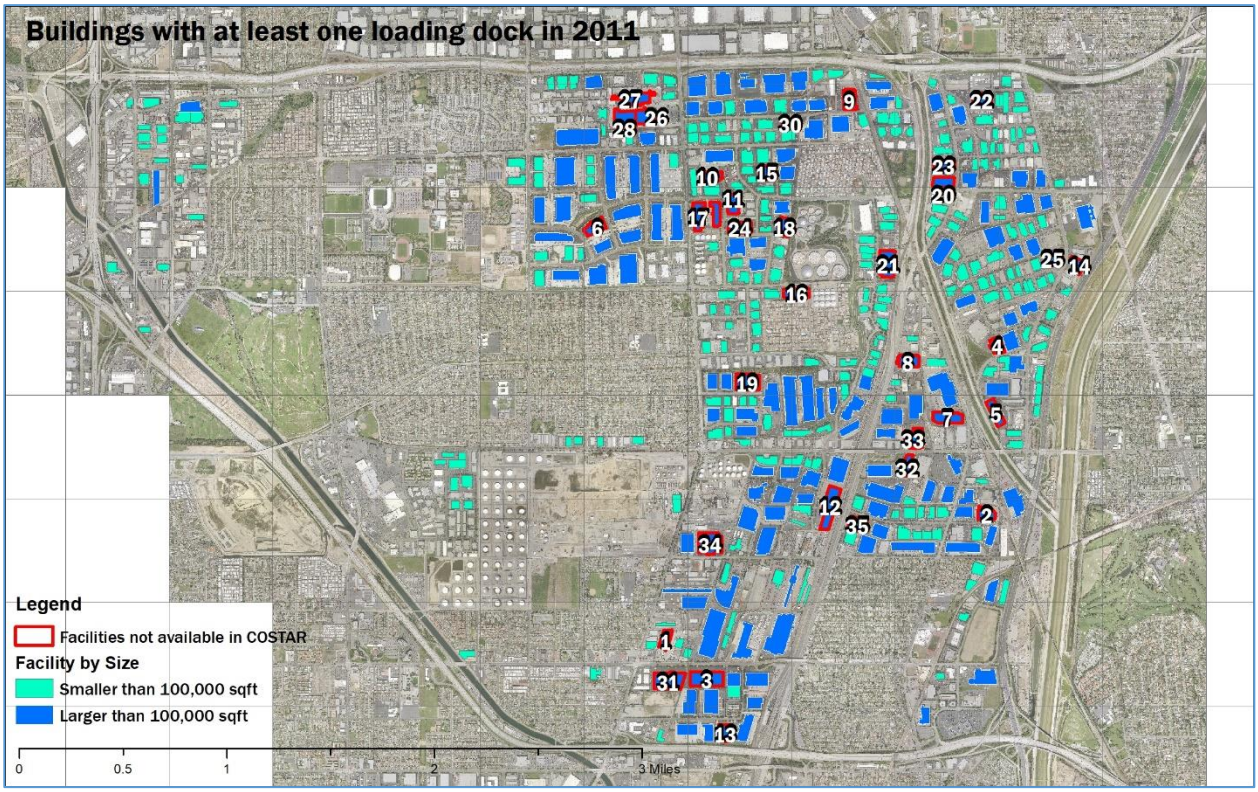

Figure 5 Map of identified buildings with at least one loading dock from LAR-IAC in 2011

Table 8 Characteristics of the buildings with W&D activity, not included in COSTAR

Number in Figure 5	Name of Business	Building size (sqft)	Type of Business *
1	Ability/Try-Modal Transportation Services	108,000	Logistics/Freight Transport
2	Arms Logistics	131,000	Logistics/Freight Transport
3	AZ Container Freight S West Inc	302,000	Logistics/Freight Transport
4	Classic Tents	120,000	Manufacturing
5	De Well Logistics	150,000	Logistics/Freight Transport
6	Dermalogica - LA HQ	146,000	Manufacturing
7	Designers Fountain	201,000	Manufacturing
8	Empire Container Freight Station	152,000	Logistics/Freight Transport
9	FedEx; Carmichael Int.Ntl. Service-APL Logistics	157,000	Logistics/Freight Transport
10	FNS Inc. (Freight Forwarding)	110,000	Logistics/Freight Transport
11	GlobeCon Freight Systems	135,000	Logistics/Freight Transport
12	Great American Seafood Imports	301,000	Wholesale
13	Hellman Worldwide Logistics	111,000	Logistics/Freight Transport
14	Home Decor Center	104,000	Manufacturing
15	IPS Corp (Plumbing/roofing products)	120,000	Manufacturing
16	Jam'n Logistics Inc-California	144,000	Logistics/Freight Transport
17	Kintetsu World Express Inc; Oasis Logistics	171,000	Logistics/Freight Transport
18	Marathon Distribution Services	131,000	Logistics/Freight Transport
19	Massive Marketing	235,000	Manufacturing
20	Matson Global	136,000	Logistics/Freight Transport
21	Nakano Warehouse & Trans Corp	243,000	Logistics/Freight Transport
22	National Retail Systems, Inc.	152,000	Logistics/Freight Transport
23	Ohmstede Industrial	137,000	Manufacturing
24	Pitney Bowes Presort Services	160,000	Logistics/Freight Transport
25	Polywin Corporation	101,000	Manufacturing
26	Ralph Grocery	171,000	Retail Trade
27	Ralph Grocery	231,000	Retail Trade
28	Ralph Grocery	220,000	Retail Trade
29	Santa Monica Seafood Company	117,000	Wholesale Trade
30	Tiger Global Logistics	100,500	Logistics/Freight Transport
31	Toll Global Forwarding	311,000	Logistics/Freight Transport
32	Used Gym Equipment	110,000	Manufacturing
33	Vanguard Logistics Services	203,000	Logistics/Freight Transport
34	Western Tube & Conduit Corp	306,000	Manufacturing
35	Win Win Express	131,000	Logistics/Freight Transport

* Type of business does not necessarily mean the NAICS industry sector associated with the warehousing facility. Rather, we describe the type of business as presented on their website.

CONCLUSIONS

We examined the comparability among multiple datasets available to identify W&D activity in the US: ZIP Code Business Patterns (ZBP, public), National Establishments Times Series (NETS, proprietary), COSTAR (proprietary), and the warehouse inventory identified by ourselves using high resolution satellite imagery (LAR-IAC). As a case study we used the Los Angeles Combined Statistical Area as well as Southeast Los Angeles to assess data comparability. We began with identifying information necessary to precisely locate a warehousing facility and quantifying its activity level. This includes the appropriate unit of analysis, size, location, and type information, as well as accurate freight flow data. No data set has all of these features.

ZBP is a public dataset maintained and published by the U.S. Census Bureau. It is consistent over years from 2003, and with other data sets produced by the Bureau. In our judgement, it is likely the most reliable source of information. Limitations are (1) its spatial unit is large, (2) actual distribution within a ZIP Code is unknown, (3) the boundary of a ZIP Code is not clear and not consistent with administrative boundaries, (5) employment data is limited, and (6) internal inconsistency is due to two revisions of the industry sector definition. NETS is a proprietary business dataset and provides an array of business information (name, address, industry sector, and business type) at the establishment level in a panel data format. However, the dataset suffers from both internal and external inconsistencies.

COSTAR is a commercial and industrial real estate database that maintains listings of real estate property for rent and leasing. It is a proprietary dataset. It provides detailed information of a property, such as rentable building area, year of construction, number of docks, ceiling height, and rail access. It is limited in that no historical inventory is available, and it only includes buildings in the rental and leasing markets. We also constructed an inventory of warehousing buildings in a small study area in Southeast Los Angeles using LAR-IAC. We defined a warehouse as a building with a single or multiple loading docks for heavy-duty trucks. In Table 9, we present the advantages and disadvantages of the three datasets we examined in this report.

Warehouse statistics are substantially different across the datasets. Basic statistics vary considerably even at the metropolitan level. This is a problem. First, our understanding of the spatial distribution of warehousing and distribution will vary significantly depending on the data set we use. This suggests that establishment level data is not yet robust enough to merit highly disaggregate analysis of spatial trends. Absent labor-intensive ground truthing as we performed with the remote sensing data, we have no way to know the extent of inaccuracies in such data sets. Second, unreliable establishment data calls into question the reliability of freight trip generation estimates that are based on them. Third, freight shipment data in the US at the sub-metropolitan level are scarce. The only publicly available shipment data sources are the Commodity Flow Survey (CFS) and the Freight Analysis Framework (FAF), both of which have geographies at the district level. If we were to attempt to disaggregate these flows based on the spatial distribution of establishments, there would be substantial errors in our estimates.

Data scarcity has been one of the most serious constraints to urban freight research. In this environment, securing the reliability and consistency of land use data may be the first step before relying on modeling and simulation approaches. Furthermore, if there exists a good portion of warehousing activity in the non-transportation sector, as seen previously, methodologies to capture the missing activity should be developed.

Table 9 Advantages and disadvantages of the business datasets

Datasets	Advantages	Disadvantages
ZBP	Public access Reliable source Annual time series from 1994 6-digit SIC/NAICS sectors	Large spatial unit No employment information; only by size classes Actual activity distribution unknown Boundaries are an estimate of address aggregation Not consistent with administrative boundaries Subject to inconsistencies from industry code conversions
NETS	Establishment level Address information Annual panel data from 1990 with establishment ID 6-digit SIC/NAICS sectors Detailed business information (employees)	Proprietary data Internally inconsistent over time Externally not comparable with US Census data
CoStar	Rent/lease listings level Address information Detailed physical characteristics of rentable facilities (building area, number of docking bays, building height, rail access)	Proprietary data No historical listings information maintained No lessor/lessee information Facilities not part of the lease market not included

ACKNOWLEDGMENT

This research was supported by the Volvo Research and Educational Foundations through the MetroFreight Center of Excellence, METRANS Transportation Center, University of Southern California. All errors and omissions are the responsibility of the authors.

REFERENCES

Bowen Jr., J., (2008). Moving places: the geography of warehousing in the US. *Journal of Transport Geography*, 16, 379–387.

Cidell, J. (2010). Concentration and decentralization: the new geography of freight distribution in US metropolitan areas, *Journal of Transport Geography*, 18(3), 363-371.

Dablanc, L., Ogilvie, S., & Goodchild, A. (2014). Logistics Sprawl: Differential Warehousing Development Patterns in Los Angeles, California, and Seattle, Washington.

Transportation Research Record: Journal of the Transportation Research Board, (2410), 105-112.

Dablanc, L., & Rakotonarivo, D. (2010). The impacts of logistics sprawl: How does the location of parcel transport terminals affect the energy efficiency of goods' movements in Paris and what can we do about it? *Procedia Social and Behavioral Sciences*, 2, 6087-6096.

Dablanc, L., & Ross, C. (2012). Atlanta: A Mega Logistics Center in the Piedmont Atlantic Megaregion (PAM). *Journal of Transport Geography*, 24, 432-442.

Giuliano, G. Y. Hou, S. Kang, and E-J. Shin. (2015). *Accessibility, Location and Employment Center Growth*. Final Report, METRANS Transportation Center Project 11-06, University of Southern California. Available at: https://www.metrans.org/research/11-06-accessibility-location-and-employment-center-growth

Giuliano, G., & Kang, S. (2018). Spatial dynamics of the logistics industry: Evidence from California. *Journal of Transport Geography*, 66, 248-258.

Rivera, L., Sheffi, Y., and Welsch, R. (2014). Logistics agglomeration in the U.S. *Transportation Research Part A*, 59, 222-238.