MF 4.1b: Data Challenges in Urban Freight Research: The Case of Warehousing and Distribution

Research Brief

Genevieve Giuliano and Sanggyun Kang
MetroFreight Center of Excellence, METRANS Transportation Center
Sol Price School of Public Policy, University of Southern California

Introduction

We explore various data sources and compare their reliability and consistency in examining the location patterns of warehousing and distribution (W&D) activity in Los Angeles, California.

Research Approach

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 describe the "ideal" data set, and then select 4 data sources with contrasting advantages and disadvantages relative to the ideal. We then evaluate the data sources for internal and external consistency.

Information Necessary to Identify W&D Activity

We describe the characteristics of the "ideal" dataset. Then, we discuss what type of data actually exist.

Unit of Analysis

Information on every W&D facility. A W&D is typically a stand-alone facility. Some W&D operation also takes place as part of other retail or manufacturing businesses.

Size

Information on the intensity of W&D activity. Intensity (or capacity) is correlated with ceiling height, number of loading docks, and floor area.

Geographic Identification

Information on address or X-Y coordinates.

Type of Facility

Information on the types of functions or services provided by each facility: receiving, storing, shipment consolidation, packing, shipping, cross-docking, and fulfillment services.

Freight Flow

Information on the volume of goods by commodity type, originating from or destined to a facility; shipper and receiver locations; type of freight vehicles.

Data Sources

ZIP Code Business Patterns (ZBP)

Published by the U.S. Census annually, ZBP provides the number of establishments by six-digit NAICS sector from 1994 to 2015. W&Ds are identified under NAICS 493 Warehousing and Storage. Employment information is available only at the county level (County Business Patterns).

National Establishment Time Series (NETS)

NETS is establishment-level annual panel data, in which various business information is available, such as address, number of employees, and industry sector at the six-digit NAICS level. NETS is based on Dun and Bradstreet data.

CoStar

CoStar is a real estate listings database for real estate brokers. Facilities classified as truck terminals, cold storages, warehouses, and distribution centers are included. It provides rentable building area, number of stories, number of loading docks, and built year.

Los Angeles Region Imagery Acquisition Consortium

The LAR-IAC program is a joint effort to collect high-resolution satellite imagery data. We use the 2011 imagery of Southeast LA with the most intense W&D activity in LA County (Figure 1).
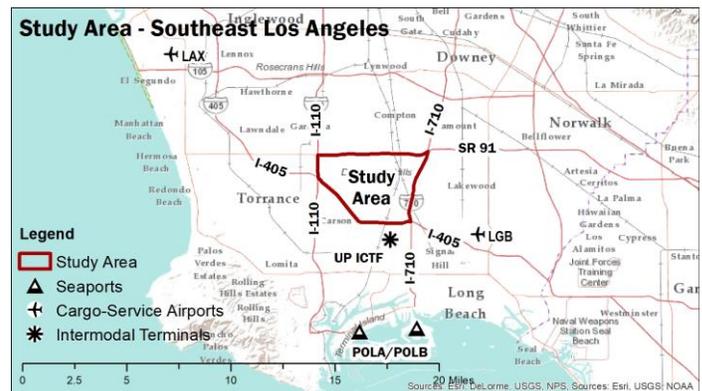
We summarize these characteristics in Table 1.

Figure 1 Study Area – Southeast Los Angeles

Table 1 Available information from public and proprietary datasets

| Datasets	ZBP	NETS	CoStar	LAR-IAC
Type of data	Public, economic census	Proprietary, business marketing data	Proprietary, industrial real estate listings	User-identified W&D samples
Unit attributes	ZIP Code	Establishment	Individual rentable facility	Individual building
Location attributes	ZIP Code centroids	Address	Address	X-Y coordinates
Size attributes	Establishment size category	Number of Employment	Rentable building area, number of loading docks	Building foot print
Type attributes	6-digit NAICS (general, farm product, refrigerated, other)	6-digit NAICS (general, farm product, refrigerated, other)	Truck terminals, warehouses, DCs, and cold storage	None
Flow attributes	None	None	None	None
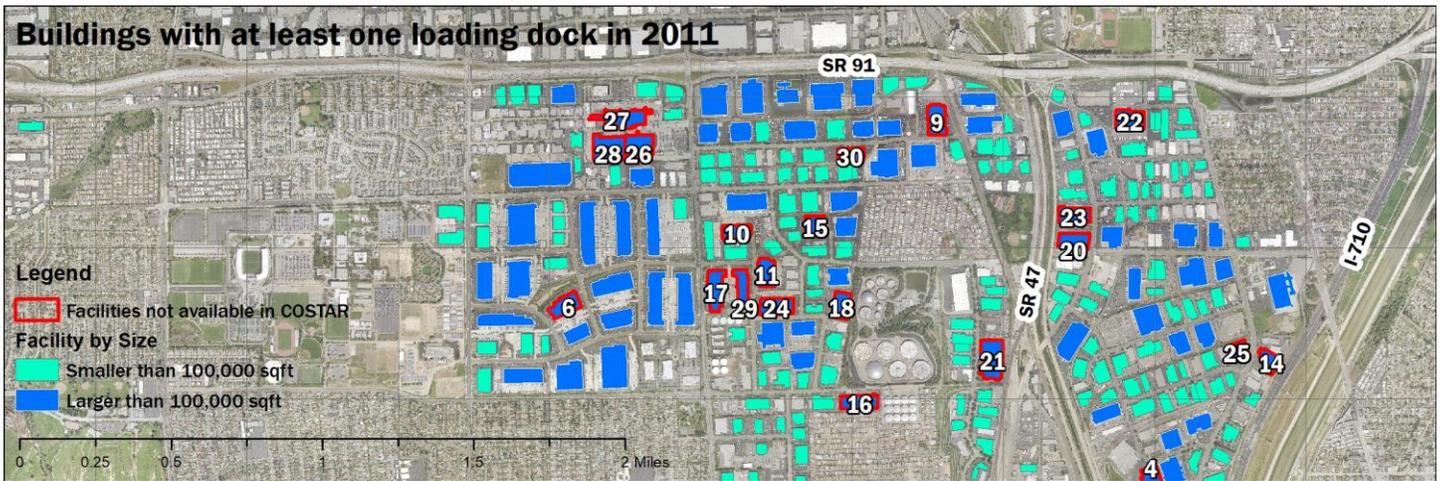
Figure 2 Map of identified buildings with at least one loading dock from LAR-IAC in 2011 (This map shows the northeast part of the study area.)

External Comparability

We compare W&D statistics (number of establishments and annual percent changes) between ZBP and NETS from 2003 to 2009 in Southern California (Los Angeles, Orange, Riverside, San Bernardino, and Ventura County). We also compare the number of W&D facilities in CoStar listings to the buildings we manually identified using high resolution satellite imagery in Southeast Los Angeles.

Metro Area-Level Comparison: ZBP and NETS

The number of W&Ds and annual percent changes are substantially different between the two data sets, even at the regional level (Table 2).

Table 2 Comparison of the number of W&Ds between ZBP and NETS

Year	ZBP	% change	NETS	% 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 2003-09	202	25.6%	74	4.4%

CoStar vs. LAR-IAC

We used the LAR-IAC imagery and identified buildings that have at least one loading dock for heavy-duty trucks. We compare results with CoStar in Table 3.

Table 3 Statistics all buildings in CoStar and identified in LAR-IAC

Type	Number	Rentable building area (ft²)		
		Sum	Mean	Std. Dev.
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 (P<0.05)	

The study area has plentiful wholesale trade, logistics, light manufacturing, and warehousing businesses. For example, a light manufacturing plant might utilize part of its building as storage and is likely to have a loading dock. The plant will be classified under the manufacturing sector, while the business will perform manufacturing and warehousing functions. We expect many buildings in this area have a similar business model.

We identified 155 buildings larger than 100,000 square feet (Figure 2). 35 out of the 155 buildings were not listed in CoStar. We used Google Maps and business websites to identify the type of business of the missing buildings. We confirmed that 20 buildings were occupied by logistics-freight transport businesses. The other 15 buildings were used by retail and wholesale trade as well as manufacturing, businesses.

Conclusions

We examined the comparability among multiple data sets available to identify W&D activity in Los Angeles, CA: ZBP, NETS, CoStar, and those identified by ourselves using LAR-IAC.

We conclude that there is no comprehensive and reliable data set for examining warehousing and distribution activity. Warehouse statistics are substantially different across the datasets even at the metropolitan level. One of the possible explanations is the difference in how W&Ds are counted. ZBP uses Tax ID, whereas NETS uses a unique business ID (DUNS Number) assigned to every establishment – a business at a single physical location. It is possible that NETS accounting for branches or divisions inflates the numbers. We do not have concrete evidence, and there is no a priori reason to judge one more accurate than the other in terms of measuring the size of W&D activity.

This is a problem. First, our understanding of the spatial distribution of W&Ds will vary significantly depending on the dataset we use. Second, unreliable land use data calls into question the reliability of freight trip generation estimates that are based on them. Perhaps more seriously, our results call into question our ability to conduct highly disaggregate employment or establishment studies with existing data sources.

Securing the reliability and consistency of land use data may be the first step before relying on modeling and simulation approaches. If there exists a good portion of warehousing activity in the non-transportation sector, methodologies to capture the missing activity should be developed.