Parcel Level Land Use Data Acquisition and Analysis for Measuring Non-Auto Accessibility

Report # 6 in the series

Access to Destinations Study

Report # 2008-19

Sponsored by:

University of Minnesota
Center for Transportation Studies

Minnesota Department of Transportation

Hennepin

Metropolitan Council

Technical Report Documentation Page

1. Report No.	2.	3. Recipients Accession No.	
MN/RC 2008-19			
4. Title and Subtitle Access to Destinations: Parcel Level Land Use Data Acquisition and Analysis for Measuring Non-Auto Accessibility		5. Report Date July 2008 6.	
7. Author(s) Jessica Horning, Ahmed El-Geneidy, Kevin J. Krizek		8. Performing Organization Report No.	
9. Performing Organization Name and Address Hubert H. Humphrey Institute of Public Affairs University of Minnesota 301 19th Ave South Minneapolis, Minnesota 55455		10. Project/Task/Work Unit No. 11. Contract (C) or Grant (G) No. (c) 89261 (wo) 31	
12. Sponsoring Organization Name and Address Minnesota Department of Transportation Research Services Section 395 John Ireland Boulevard St. Paul, Minnesota 55155		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code	

15. Supplementary Notes
http://www.lrrb.org/PDF/200819.pdf
Report #6 in the series: Access to Destinations Study

16. Abstract (Limit: 200 words)

This research aids in tackling one important part of accessibility metrics—measuring land use. It introduces complementary strategies to effectively measure a variety of different destination types at a highly detailed scale of resolution using secondary data. The research describes ways to overcome common data hurdles and demonstrates how existing data in one metropolitan area in the U.S. –the Twin Cities of Minneapolis and St. Paul –can be exploited to aid in measuring accessibility at an extremely fine unit of analysis (i.e., the parcel).

Establishment-level data containing attribute information on location, sales, employees, and industry classification was purchased from Dun & Bradstreet, Inc. The research process involved cleaning and tailoring the parcel dataset for the 7-county metro area and integrating various GIS datasets with other secondary data sources. These data were merged with parcel-level land use data from the Metropolitan Council. The establishment-level data were then recoded into destination categories using the 2 to 6-digit classifications of the North American Industry Classification System (NAICS). The development of important components of this research is illustrated with a sample application. The report concludes by describing how such data could be used in calculating more robust measures of accessibility.

17. Document Analysis/Descriptors	18. Availability Statement		
Accessibility, parcel level data, destinations, GIS systems, locational attributes, NAICS codes	No restrictions. Document available from: National Technical Information Services, Springfield, Virginia 22161		
19. Security Class (this report) Unclassified	20. Security Class (this page) Unclassified	21. No. of Pages 43	22. Price

Access to Destinations:

Parcel Level Land Use Data Acquisition and Analysis for Measuring Non-Auto Accessibility

Report # 6 in the series

Final Report

Prepared by:

Jessica Horning
Ahmed El-Geneidy
Kevin J. Krizek

Active Communities/Transportation Research Group
Humphrey Institute
University of Minnesota

July 2008

Published by:

Minnesota Department of Transportation
Research Services Section
395 John Ireland Boulevard, MS 330
St. Paul, Minnesota 55155-1899

ACKNOWLEDGEMENTS

The authors thank Prof. Mohamed Mokbel, assistant professor computer science, University of Minnesota, and Mohamed Khalifa, graduate student computer science, University of Minnesota, for their help in joining the data in the Oracle database system. Also, the authors would like to thank the TAP members for their advice and support throughout the project.

TABLE OF CONTENTS

CHAPTER I: Introduction 1

CHAPTER II: LITERATURE REVIEW 2

CHAPTER III: DATA 3
Business Data 3
Land Use Data 6
Parcel Data 6

CHAPTER IV: METHODOLOGY 8
Classifying Destinations 8
Parcel Matching 9
Data Cleaning 9
Oracle Database Matching 9
Geocoding Matching 10
Grid Cell Matching 11
Matching Summary 12
Final Note 13

REFERENCES 14

APPENDIX A: Reclassified Destination Categories and Specific Destinations
APPENDIX B: NAICS Reclassification SPSS Syntax
APPENDIX C: Instructions for Matching Dun & Bradstreet and Parcel Point Data Using Oracle Database
APPENDIX D: Working with D&B – Parcel Matches Made by CS Student
APPENDIX E: Matching D&B Businesses to Parcel Points Using the Adapted Geocoding Method
APPENDIX F: Grid Cell Matching Methodology

LIST OF TABLES

Table 1: Comparison of Dun & Bradstreet and InfoUSA Business Datasets 4

LIST OF FIGURES

Figure 1: Accuracy of InfoUSA Business Locations 5
Figure 2: Accuracy of Geocoded Dun & Bradstreet Business Locations 6
Figure 3: Grid Cell Matching Example 11
Figure 4: Business/Parcel Match Summary 12
Figure 5: Sample of Matched Businesses with Parcels 13

EXECUTIVE SUMMARY

A defining aspect of any measure of accessibility lies in knowing attributes of the destinations to which people are traveling. The traditional emphasis on employment accessibility is understandable, given its link to other important aspects of urban structure, such as choice of residential location, and also to outcomes hypothesized to be related to urban structure, such as social exclusion. However, access to *other* types of destinations, such as retail, recreation or medical care is just as important. These *other* destinations strongly influence various dimensions of travel behavior such as trip frequency, destination choice, mode choice, and trip or tour complexity. The problem is that the existing practice and literature is void in describing strategies for effectively measuring a broad array of activities at a high degree of resolution.

Broadening the scope of accessibility to include a full array of destinations has been long talked about in land use-transportation discussions. Doing so is also a key to adequately capturing accessibility metrics for non-auto modes such as walking and cycling. To date, however, such discussions have been short on execution. Issues including, but certainly not limited to, lack of reliable data, computational power or knowledge of non-motorized travel behavior have prevented widespread application of such measures.

This research aids in tackling one important part of accessibility metrics—measuring land use. It introduces complementary strategies to effectively measure a variety of different destination types at a highly detailed scale of resolution using secondary data. The research describes ways to overcome common data hurdles and demonstrates how existing data in one metropolitan area in the U.S. –the Twin Cities of Minneapolis and St. Paul –can be exploited to aid in measuring accessibility at an extremely fine unit of analysis (i.e., the parcel).

Establishment-level data containing attribute information on location, sales, employees, and industry classification was purchased from Dun & Bradstreet, Inc. It was then necessary to clean, tailor, and in some cases modify different methodologies to match with secondary datasets. A secondary goal was to identify which matching method is most effective and efficient and describe such. The process involved cleaning and tailoring the parcel dataset for the 7-county metro area and integrating various GIS datasets with other secondary data sources. In all, data were available for 135,928 businesses within the region. These data were merged with parcel-level land use data from the Metropolitan Council, the Twin Cities' regional planning agency. The establishment-level data were then recoded into destination categories using the 2 to 6-digit classifications of the North American Industry Classification System (NAICS). The final outcome of this process is a valuable set of parcel-level land use data with information on employment counts and sales volumes. The development of important components of this research is illustrated with a sample applications. The report concludes by describing how such data could be used in calculating more robust measures of accessibility.

CHAPTER I: INTRODUCTION

The goal of the Access to Destinations initiative is to generate a three dimensional multi-modal accessibility matrix for the 7-county Twin Cities Metropolitan region. There are three components needed to calculate the accessibility indices included in this matrix: knowing travel times, knowing the types of activities people travel to (e.g., land uses), and knowing how much land uses that are “closer” should be valued over "further away" land uses (e.g., distance decay functions). All are important. However, data and methods used to create accessibility measures for the automobile have been around for some time and are relatively proven. The protocol for data preparation and subsequent analysis is relatively straightforward. The same is not true for the data and methods necessary to capture smaller levels of analysis—analysis that appropriate to accurately capture walking and cycling. The goal of this Parcel Level Land Use project is to develop a detailed dataset of land use activities in the 7-county Metro Area. This dataset will include destinations that are influenced by all modes (auto, transit, walking, and bicycling) and will serve as the second component for accessibility calculations in the Access to Destinations initiative.

The end product of the Parcel Level Land Use project is a geographic information system (GIS) layer for the 7-county Twin Cities Metropolitan region that reliably represents the types of land use activities that people travel to at the neighborhood and regional levels of analysis. This layer includes detailed data that captures the small-scale land uses that are likely to be more influenced by transit, cycling, or walking. It was generated by cleaning and tailoring the parcel dataset for the 7-county metro area and integrating various GIS datasets with other secondary data sources. The findings of this analysis will be integrated as part of the Access to Destinations initiative to generate accessibility measures to various destinations in the region using various modes.

CHAPTER II: LITERATURE REVIEW

The concept of accessibility, has been around in the transportation planning field for more than 40 years. Improving accessibility is a common element in the goals section in almost all transportation plans in the US (Handy, 2002). However, the term accessibility is often misused and confused with other terms such as mobility. There are various methods to measure accessibility in a region. For example (Baradaran & Ramjerdi, 2001) identify five different ways for measuring accessibility, while Handy and Niemeier (1997) named only three of these five as potential measures for planners to use. The isochronic or cumulative opportunity measure is one of the basic and early measures discussed in the literature (Vickerman, 1974; Wachs & Kumagai, 1973). This approach counts the number of potential opportunities that can be reached within a predetermined travel time. The gravity-based measure discussed in Hansen (1959) is still the most widely used general method for measuring accessibility, although it is more complex in calculations and has some points of weaknesses. The most complex and data intensive is the utility-based measure. Several researchers use this method since it adheres to travel behavior theories (Ben-Akiva & Lermand, 1977; Neuburger, 1971). High levels of accessibility to various activities in a city can be present, yet the amount of time available in a day that people can spend to reach these activities might be low. This leads to the constraints-based measure or people-based measure of accessibility (Wu & Miller, 2002). A fifth measure is the composite accessibility measure. A composite measure is suggested by (Harvey Miller, 1999) where he combines space-time and utility-based measures in one measure. This approach introduces a higher level of complexity where time constraints are superimposed. The composite accessibility measure requires more data than utility-based measures and is even more complex in terms of calculations and accordingly generalizing it for usage is not an easy task.

The common factor between these methods of measuring accessibility is the level of attractiveness at the destination. Most researchers use accessibility to jobs and accessibility to resident workers as a sample due to the simplicity of explaining and conducting such analysis. Yet accessibility to detailed retail destinations, for example, can be rarely found in the planning literature due to the lack of appropriate data to generate such analysis. Even if such information is available accessibility measures are calculated at a very small level of analysis and not a regional one. This research will help in advancing the area of measurement of accessibility through providing a detailed land use layer that can be used with various distance decay functions and appropriate knowledge of travel behavior research to generate a better accessibility measure for the Twin Cities region.

CHAPTER III: DATA

The goal of the Access to Destinations initiative is to generate a three-dimensional accessibility matrix for the Twin Cities region. This matrix will contain the various modes (auto, transit, walk, bike) as its rows and activity types (e.g., shopping, entertainment, services) as its columns. The third dimension or the z-value will represent time, so that accessibility for each cell can be presented over time. Current research conducted by the Access to Destinations research team is mainly involved in estimating travel time for various modes of transportation and generating a set of accessibility measures. Future projects will include research that focuses on distance decay (the diminishing effect of attractiveness of activities) and forecasting land use change over time.

The main objective of this research project is to generate a detailed parcel level land use dataset in a GIS environment. To accomplish this task, commercial data (including retail and non-retail establishments) as well as information about residential and recreational activities were prepared from secondary data sources and included as GIS layers. These layers enable expanding the accessibility matrix in a way that can accommodate measuring accessibility to detailed land use (grocery shopping, doctors, movie theaters, barber shops,etc) rather than generalized ones (commercial, employment, entertainment, recreation,etc). The GIS layers will also help in developing accessibility measures at the neighborhood level of analysis, for various modes of transportation including auto and non-auto.

Business Data

Two commercial datasets (including retail and non-retail establishments) were considered for the creating the GIS layer. The first was the InfoUSA dataset, which is included as a part of the ArcGIS Business Analyst software extension. The second was the Dun and Bradstreet business listing. The research team acquired the 2005 business listings for the 7-county metro area from both sources and compared the two datasets for accuracy, thoroughness, and suitability for the parcel level land use project. Table 1 shows a comparison of the business data provided by Dun & Bradstreet and InfoUSA. Overall, the Dun & Bradstreet dataset contained over 40,000 more businesses than InfoUSA. Dun & Bradstreet also provided more detailed sales and employment figures for each company that were presented as figures for each individual business site as well as for each company as a whole (for chains with multiple locations). InfoUSA, on the other hand, provided only one sales and employment field that sometimes contained figures for an entire company, not just the specific site.

Table 1: Comparison of Dun & Bradstreet and InfoUSA Business Datasets

	Dun & Bradstreet (D&B)	InfoUSA (ESRI Business Analyst)
# of Records	135,928	93,840
Data Format	Tables downloaded from online database based on zip code	Shapefile included with ESRI's ArcGIS Business Analyst extension software
Attributes	Street & mailing addresses	--
	8-digit SICS (extra 2-digits created by D&B)	6-digit SICS
	6-digit NAICS	8-digit NAICS (extra 2-digits created by InfoUSA)
	Sales (individual site & all sites)	Sales (sometimes shows sales for whole company, not individual site)
	Employees (individual site & all sites)	Employees (sometimes shows employment for whole company, not individual site)
Archived Data	Available for 1995 and 2000	?

Historical business data for the years 1995 and 2000 is available through Dun & Bradstreet, while it is unclear if archived data is available through InfoUSA or ESRI. This was an important consideration when choosing which business database to use, since a complete 3-D accessibility matrix will require accurate historic business and land use data.

Another important consideration was the format that both datasets were made available in and the ease of accurately matching these data to the appropriate parcels. InfoUSA business data was provided to the research team in the form of geocoded points and did not include the actual address for each business as an attribute in the associated data table. Figure 1 shows a sample of the business points provided by InfoUSA and compares the geocoded locations of InfoUSA businesses to the actual locations of these businesses. It is clear from this map that the majority of these points either do not fall within parcel or fall within an incorrect parcel. This fact made linking the InfoUSA data to the parcels problematic, since address information for these businesses needed to correct these geocoding errors must be looked up separately for each individual business using a secondary dataset.

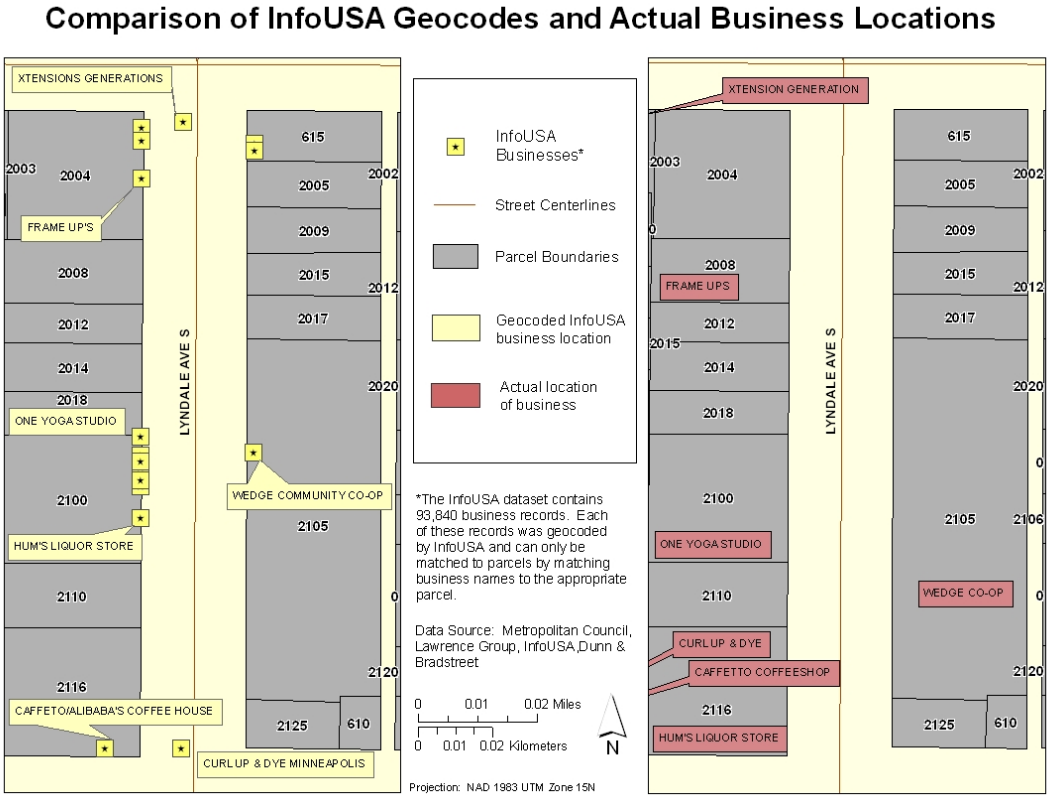

Figure 1: Accuracy of InfoUSA Business Locations

The Dun & Bradstreet data, on the other hand, was downloaded from Dun & Bradstreet's website in .dbf table format. In order to work with these records in GIS and match them to parcels, each record in this dataset had to first be geocoded by the Access research team. Figure 2 shows the actual and geocoded locations of a sample of Dun & Bradstreet businesses. As was the case with the InfoUSA points, it is clear from observing the map the majority of the Dun & Bradstreet points do not fall within the parcels they should be associated to. Accordingly a different method has to be used to enable matching the parcel information with the business dataset. Since the Dun & Bradstreet dataset contained address information for each business and appeared to be a more thorough listing of businesses and attributes, the research team utilized only this dataset for the remainder of the analysis.

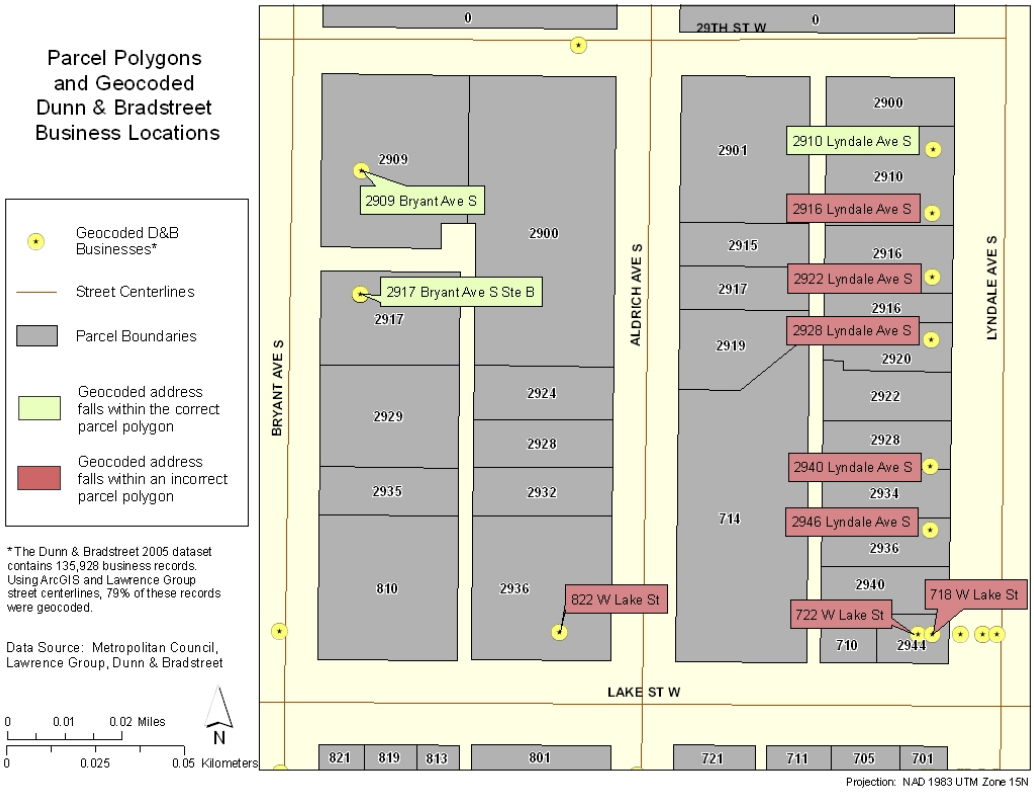

Figure 2: Accuracy of Geocoded Dun & Bradstreet Business Locations

Land Use Data

In addition to business destinations, the Access to Destinations project will create accessibility measures for certain land uses destinations (i.e., parks). For this study, the research team acquired a generalized land use GIS layer from MetroGIS's "Datafinder" website. The 2005 Generalized Land Use dataset encompasses the seven county Twin Cities Metropolitan Area. The dataset was developed by the Metropolitan Council, and was interpreted from 2005 air photos, with additional assistance from county parcel data and assessor's information, reverse directories, and Internet information. The primary land use classifications that will be used in the Access to Destinations project are "airport" and "parks".

Parcel Data

The 2006 Regional Parcel Dataset for the seven county Twin Cities Metropolitan Area was also obtained from MetroGIS's "Datafinder" website. This dataset is a compilation of tax parcel polygon and point layers from the counties of Anoka, Carver, Dakota, Hennepin, Ramsey, Scott and Washington. A standard set of attribute fields is included for each county and the attributes are the same for the polygon and points layers. The polygon layer contains one record for each real estate/tax parcel polygon within each county's parcel dataset. The points layer, on the other hand, is intended to provide information in situations where multiple tax parcels are represented by a single polygon. This layer is especially useful in cases such as malls or condominiums, where multiple units (sometimes with unique addresses) are located within the same building or parcel

polygon. Records for these units/addresses may not show up in the polygon dataset, but exist in the point dataset. The regional parcel point dataset consists of 1,007,484 records, while the polygon dataset consists of only 933,718. For this reason, the parcel point dataset was used to match the Dun & Bradstreet businesses to parcels.

CHAPTER IV: METHODOLOGY

Classifying Destinations

The first task that is necessary for any project dealing with accessibility is to identify attractors and define the destinations that you want to calculate accessibility to. In order to identify more detailed categories of destinations than those that have been used in previous studies, the research team utilized North American Industrial Classification System (NAICS) codes. The U.S., Canada, and Mexico developed NAICS jointly to provide new comparability in statistics about business activity across North America. NAICS replaced the Standard Industrial Classification system in 1997 and is a preferable way to identify destinations because it provides a more accurate and detailed classification system for businesses within the growing service sector of the economy. Each business record in the Dun & Bradstreet dataset contains a 6-digit NAICS code. This code can be used to examine business patterns at different levels of aggregation. The first two digits of the code represent the business's "industry sector" (i.e. retail trade), the first three digits represent its "subsector" (i.e. food and beverage stores), four digits represents its "group" (i.e. specialty food stores), five digits represents its "industry" (i.e. meat markets), and six digits represents its "US national industry".

The level of detail of the NAICS codes included in the Dun & Bradstreet data is too detailed to be used for general accessibility measures. For example the 6 digit NAICS code can be used to differentiate between different kinds of beauty salons and hair dressers, while only using the first 3-digit can easily identify that it is a beauty related business. Thus, the data needed to be reclassified based on the type of desired destinations for our study. This was done both by using different levels of disaggregation to identify different types of destinations and by reclassifying some NAICS codes so that appropriate businesses are grouped together. For example, grocery stores can be identified at the 2-digit level, but coffee shops can only be identified at the 6-digit level. Similarly, some activities that most people would classify as the same type of destination are categorized as different industries by NAICS. For example, fitness centers are included in the Arts & Entertainment industry, but are included in the Education industry if they offer yoga instruction.

The research team reclassified the Dun and Bradstreet businesses into seventeen larger destination categories: Shopping; Food, Groceries, and Restaurants; Fitness and Recreation; Entertainment; Education; Health Care; Post Offices; Financial Services; Other Personal Services; Professional Services; Transportation; Information; Administration and Support; Wholesale Retail; Real Estate and Rental Services; Other; and Unclassified. Each destination category is composed of multiple reclassified specific destinations. For example, the destination category "Food, Groceries, Restaurants" is composed of grocery store, convenience store, restaurant, and bar/nightclub destinations. Appendix A contains a full list of the specific destination reclassifications created for this study and the original NAICS classifications that are included in each reclassification. The research team imported the Dun & Bradstreet dataset into SPSS statistical software and created a syntax file to automate the NAICS reclassification process. Appendix B contains the syntax used to reclassify the 2005 Dun & Bradstreet data.

Parcel Matching

The second and largest task necessary for examining accessibility at the parcel level is to match destinations to the appropriate points and polygons within the metropolitan parcel dataset. As mentioned previously, an accurate street address for each business is required in order to accurately match it to the correct parcel. As a result, only the Dun & Bradstreet business data could be used for this study because InfoUSA does not release such information with its extracted shapefiles. In addition, since simply geocoding the Dun & Bradstreet businesses and spatially joining them to the parcel polygon dataset did not result in satisfactory matches, several other methodologies had to be tested and used to match the two datasets.

The primary methodology used to match the parcel and Dun & Bradstreet datasets involved cleaning the address fields in both datasets and matching them using Oracle database software and a programming script developed by a University of Minnesota Computer Science graduate research assistant. This type of matching ensures a higher level of accuracy and the generation of a more reliable dataset that can be used in generating accessibility measures.

Data Cleaning

Several issues are well known to be present in address data. These issues include problems with abbreviation and spelling mistakes associated with typing errors. A list of commonly used abbreviations was obtained from the US Postal Service website to assist in standardizing all the abbreviations in both the parcel and Dun & Bradstreet tables. Many uncommon but consistent abbreviations also existed within the Dun & Bradstreet address data that needed to be identified and changed before the addresses could be matched to any other dataset. For example, many street names were consistently written with the vowels removed so that “Lexington Ave.” appeared as “Lxngtn Ave.”. Many businesses also had only the name of a shopping mall, a downtown office building, or an airport in their address field. These businesses needed to be identified and given the actual street address for the building they were located in. Businesses located within the Mall of America presented a particular problem because they possess unique “street addresses” based on the street system within the mall, but the mall itself is a single parcel with an address based on the street system outside the mall. To match these businesses, the research team queried the Dun & Bradstreet database for addresses with Mall of America street names (i.e. West Market, South Gardens) and replaced 279 business’s addresses with the Mall of America parcel address.

Businesses with address information that could not possibly be matched to a parcel within the 7-County Metro Area were removed from the dataset before matching was attempted. These records included 815 businesses with P.O. Boxes listed as addresses, 8,377 businesses with zip codes outside of the 7-County Metro Area, and 7,000 businesses with missing zip code, building number, street name, or other address data. After removing these records and cleaning the data, 119,736 businesses remained in the Dun & Bradstreet dataset to be matched with a parcel.

Oracle Database Matching

After cleaning the data, it was given to a Computer Science graduate research assistant to match to the parcel point dataset using Oracle database software. The process

developed by the research assistant standardized the address data from both datasets, replaced inconsistent abbreviations, compared street names to an alternate name table created from the Lawrence Group street centerline database, and created a field containing the unique "pin" number of the parcel point each Dun & Bradstreet business was matched to (if a match was found). Appendix C contains step-by-step instructions how to match the two datasets using this method and recommendations for how to use this method to match the 1995 and 2000 Dun & Bradstreet datasets to the parcel dataset.

This process resulted in 67,185 exact matches and 19,492 semi-matches. An exact match means that the building number, street name, and zip code of the business and matched parcel point are identical. Semi-matched businesses do not have an exact match within the parcel dataset, so instead were matched to the closest parcel address. Semi-matched businesses and parcels have identical street names and zip codes, but are matched to the closest building number on the same side of the street (i.e. even building numbers are matched to the closest even parcel building number and odd building numbers are matched to the closest odd parcel building number). The result of this process is a text file with three columns of data (the DUNS ID, parcel pin, and match score). Appendix D contains instructions for spot checking this table and converting it into a GIS shapefile.

Geocoding Matching

By using the Oracle database, the research team was able to match 86,677 Dun & Bradstreet businesses to parcel points. Although this was over a 60% success rate, we decided to experiment with other matching methodologies to see if an easier, more effective method could be found. The second method that the research team used to match businesses to the parcel dataset was an adapted geocoding method. The first step in this process was to geocode in ArcGIS a table containing the ID numbers and address data for all of the Dun & Bradstreet businesses that were not matched to a parcel using the Oracle database method. This regular geocoding process results in a set of points that are located near the correct parcel, but do not generally fall within its boundaries, as illustrated in Figure 2. To correct for this difference in precision between the two datasets, the research team exported the table containing the text address data for each of the 7 County Metro parcel points and geocoded this address data in ArcGIS. This results in a parcel point shapefile containing the same unique ID and address information as the original parcel point file provided by the Metropolitan Council, but with a slightly different set of XY coordinates (because the original points are generally polygon centroids, whereas the second shapefile is a set of geocoded points). The key attribute of the resulting geocoded parcel point file is that it has an identical spatial error to the geocoded Dun & Bradstreet businesses. Because of this identical error, the research team was able to spatially join the geocoded Dun & Bradstreet and geocoded parcel point datasets in ArcGIS and declare any business points that intersected (had identical XY coordinates as) a parcel point a match. Using this method, the 17,631 additional business/parcel matches were made. Appendix E contains instructions for replicating this approach.

Grid Cell Matching

The remaining Dun & Bradstreet businesses that were not matched using the Oracle database method or the adapted geocoding method were used to test a final matching methodology using grid cells. For this approach the research team created a 300 by 300 meter grid covering the entire 7-County Metropolitan Area. Each grid cell was assigned a unique ID and each parcel point and geocoded business was assigned a grid cell ID based on the grid cell it fell within. Two tables containing the unique ID, grid cell ID, and address of each business and parcel were returned to the computer science research assistant, and by adapting the original Oracle matching method businesses were matched to the parcel point with the closest matching street address that was also within the same or an adjacent grid cell as the geocoded business. Figure 3 shows an example of the grid cell matching system. Using this methodology, a business in grid cell number 55279 with the address 1600 Grand Avenue would be matched to the parcel point with the closest street address (i.e. 1612 Grand Avenue) that is located in the same or an adjacent grid cell (54927, 54926, 54925, 55280, 55279, 55278, 55633, 55632, 55631). Appendix F contains instructions for creating the grid and assigning grid cell IDs to points.

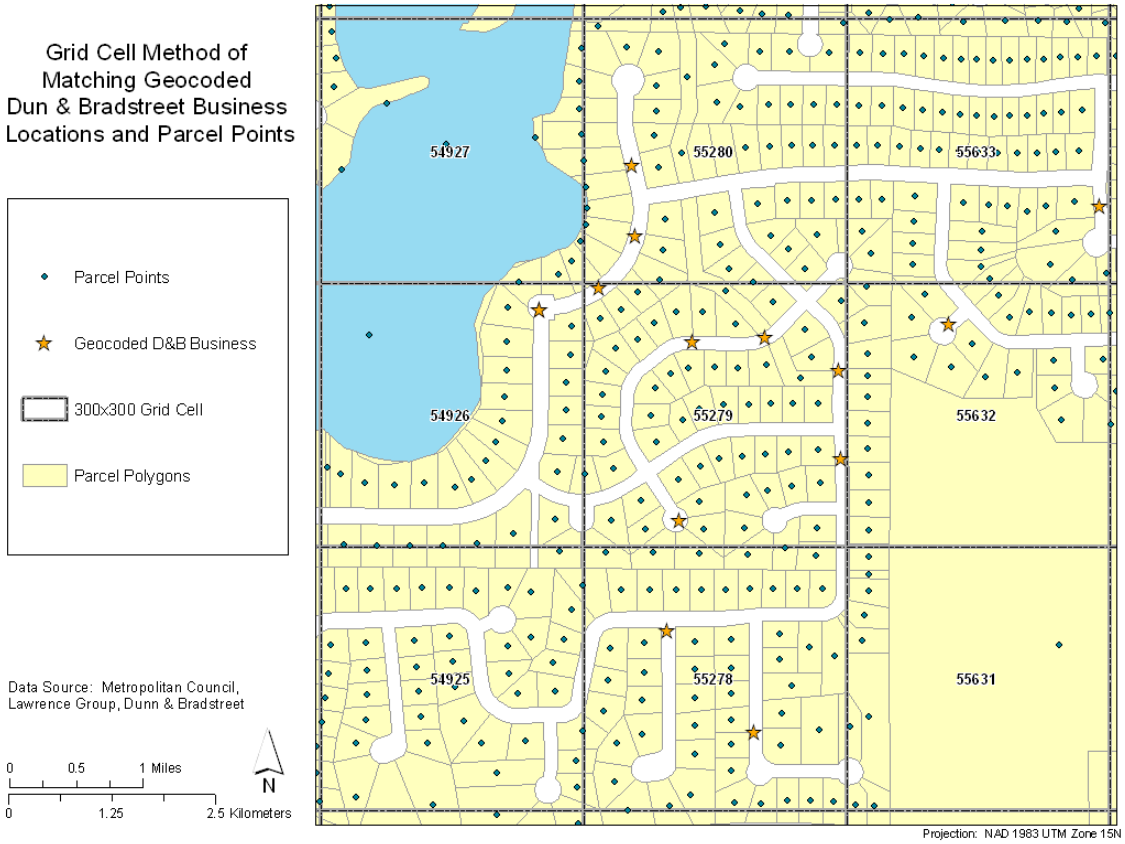

Figure 3: Grid Cell Matching Example

The grid cell methodology resulted in an additional 5,738 business/parcel semi-matches. Using this method also revealed that it is impossible to match some businesses to the closest parcel accurately because of gaps in the parcel dataset. These businesses have complete address data included in the Dun & Bradstreet dataset, but when geocoded fall in areas that have only parcels with "address unassigned" in their address fields. This occurred with 111 businesses in the 2005 datasets. These businesses were removed from the remainder of this analysis.

Matching Summary

The methodologies described above resulted in a total of 119,735 business/parcel matches. After removing the 16,192 businesses that were removed from the analysis due to bad and/or missing address data, only 9,411 businesses remained unmatched. This is a success rate of 92%. Figure 4 shows a detailed breakdown of how many businesses were matched using each method discussed above.

REMOVED	815	P.O. Boxes
	8,377	Not in 7-County Area
	7,000	Bad / Missing Address Data
	16,192	
MATCHED	17,631	Matched with geocoded parcel points
	279	Matched to Mall of America
	67,185	Exact matches made by CS student
	25,230	Semi-matches made by CS student
	110,325	
UNMATCHED	9,300	Unable to find match
	111	All parcels within 300 meters are "Address Unassigned"
	9,411	
TOTAL	135,928	92% Matched

Figure 4: Business/Parcel Match Summary

Figure 5 shows a sample of the matched business. It is important to note that the gray color represents parcels with no matching businesses. One of the goals of the Access to Destinations project is to create a three dimensional matrix that allows accessibility to be analyzed over time. To create this matrix at the parcel level of detail, parcel level land use layers will need to be created for multiple years. This report discusses the methodologies used to match the 2005 Dun & Bradstreet dataset to the 2005 parcel dataset. Multiple methodologies were used for matching this dataset in order to identify which matching method is most effective and efficient. In the future when working with archived and new Dun & Bradstreet and parcel datasets we recommend using a slightly different approach. First, businesses that existed in 2005 and have already been matched should be selected out of each dataset by querying the DUNS ID number fields in Access. (Make sure that the business's address has not changed, then set these aside.) The remaining data should be cleaned extensively following the steps described in this report and records with bad and/or missing address data should be removed. The remaining

business records should be given to the Computer Science student to be matched in Oracle. Any remaining unmatched businesses should be geocoded, assigned grid cell IDs and then returned to the Computer Science student for final matching. This process will be the most efficient and result in the largest number of accurate matches.

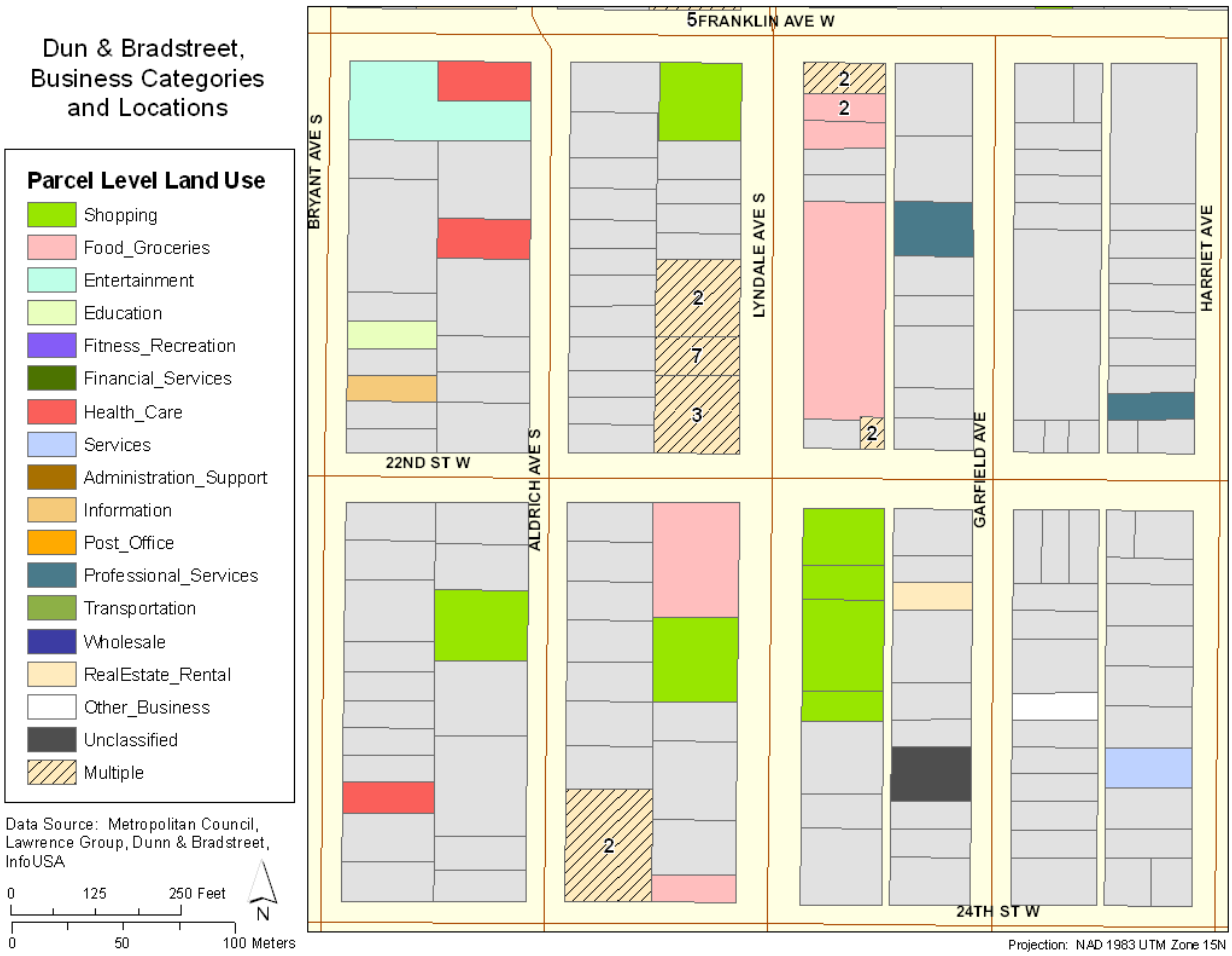

Figure 5: Sample of Matched Businesses with Parcels

Final Note

It is important to note that when performing a semi-match, businesses are linked to the nearest parcel on a prescribed side of the street. The nearest parcel might be a non-business establishment; it is therefore important to head caution when using the results of such matching for extremely detailed analysis. Better parcel and business datasets can lead to higher accuracy in reaching more perfect matches. In a best case scenario all matched data should be achieved during the first stage in the methodology using the geocoding process. Based on our knowledge and research conducted in the Twin Cities region, the parcel level data obtained from the Metropolitan Council is the best available data in terms of parcel accuracy at this scale. Also, the Dun and Bradstreet revealed much higher accuracy compared to other sources such as InfoUSA data.

REFERENCES

Baradaran, S., & Ramjerdi, F. (2001). Performance of accessibility measures in Europe. *Journal of Transportation Statistics, 4*(2/3), 31-48.

Ben-Akiva, M., & Lermand, S. (1977). *Disaggregate travel and mobility choice models and measures of accessibility.* Paper presented at the Proceedings of the Third International Conference on Behavioral Travel Modeling, Tanenda, Australia.

Handy, S. (2002). *Accessibility- vs. mobility-enhancing strategies for addressing automobile dependence in the U.S.* Paper presented at the European Conference of Ministers of Transport.

Handy, S. L., & Niemeier, D. A. (1997). Measuring accessibility: An exploration of issues and alternatives. *Environment and Planning A, 29*(7), 1175-1194.

Hansen, W. (1959). How accessibility shape land use. *Journal of the American Institute of Planners, 25*(2), 73-76.

Miller, H. (1999). *GIS software for measuring space-time accessibility in transportation planning analysis.* Paper presented at the International Workshop on Geographic Information Systems for Transportation and Intelligent Transportation Systems, Hong Kong.

Neuburger, H. (1971). User benefits in the evaluation of transport and land use plans. *Journal of Transportation Economics and Policy, 5,* 52-75.

Vickerman, R. W. (1974). Accessibility, attraction and potential: A review of some concepts and their use in determining mobility. *Environment and Planning A, 6,* 675-691.

Wachs, M., & Kumagai, T. (1973). Physical accessibility as a social indicator. *Socioeconomic Planning Science, 7,* 327-456.

Wu, Y., & Miller, H. (2002). Computational tools for measuring space-time accessibility within transportation networks with dynamic flow (special issue on accessibility). *Journal of Transportation and Statistics, 4*(2/3), 1-14.

APPENDIX A:
Reclassified Destination Categories and Specific Destinations

Destination Category	Reclassified Specific Destination	Original NAICS Code & Classification
Shopping	441100 - Motor Vehicle Dealers, Parts, and Maintenance	441110 New Car Dealers 441120 Used Car Dealers 441210 Recreational Vehicle Dealers 441221 Motorcycle Dealers 441222 Boat Dealers 441229 All Other Motor Vehicle Dealers 441310 Automotive Parts and Accessories Stores 441320 Tire Dealers 532111 Passenger Car Rental 532112 Passenger Car Leasing 532120 Truck, Utility Trailer, and RV (Recreational Vehicle) Rental and Leasing 811111 General Automotive Repair 811112 Automotive Exhaust System Repair 811113 Automotive Transmission Repair 811118 Other Automotive Mechanical and Electrical Repair and Maintenance 811121 Automotive Body, Paint, and Interior Repair and Maintenance 811122 Automotive Glass Replacement Shops 811191 Automotive Oil Change and Lubrication Shops 811192 Car Washes 811198 All Other Automotive Repair and Maintenance
	442000 - Furniture and Home Furnishing Stores	442110 Furniture Stores 442210 Floor Covering Stores 442291 Window Treatment Stores 442299 All Other Home Furnishings Stores
	443000 - Electronics, Appliance, and Camera Stores	443111 Household Appliance Stores 443112 Radio, Television, and Other Electronics Stores 443120 Computer and Software Stores 443130 Camera and Photographic Supplies Stores
	444100 - Hardware Stores and Home Centers	444110 Home Centers 444130 Hardware Stores
	444101 - Specialized Building Material, Equipment, and Supply Stores	444120 Paint and Wallpaper Stores 444190 Other Building Material Dealers 444210 Outdoor Power Equipment Stores

Destination Category	Reclassified Specific Destination	Original NAICS Code & Classification	
Shopping *(continued)*	444101 - Specialized Building Material, Equipment, and Supply Stores *(continued)*	453930	Manufactured (Mobile) Home Dealers
	444220 - Nursery, Garden Center, and Farm Supply Stores	444220	Nursery, Garden Center, and Farm Supply Stores
	446120 - Beauty Supply and Perfume Stores	446120	Cosmetics, Beauty Supplies, and Perfume Stores
	447190 - Gas Station	447190	Other Gasoline Stations
		447110	Gasoline Stations with Convenience Stores
	448100 - Clothing, Shoe, and Accessory Stores	448110	Men's Clothing Stores
		448120	Women's Clothing Stores
		448130	Children's and Infants' Clothing Stores
		448140	Family Clothing Stores
		448210	Shoe Stores
		448150	Clothing Accessories Stores
		448190	Other Clothing Stores
		448310	Jewelry Stores
		448320	Luggage and Leather Goods Stores
	451100 - Sports, Hobby, and Musical Instrument Stores	451110	Sporting Goods Stores
		451120	Hobby, Toy, and Game Stores
		451130	Sewing, Needlework, and Piece Goods Stores
		451140	Musical Instrument and Supplies Stores
		451220	Prerecorded Tape, Compact Disc, and Record Stores
	451210 - Book Stores and Newsstands (new)	451211	Book Stores
		451212	News Dealers and Newsstands
	452100 - Department Stores	452111	Department Stores (except Discount Department Stores)
		452112	Discount Department Stores
	453000 - Miscellaneous Store Retailers	452990	All Other General Merchandise Stores
		453110	Florists
		453210	Office Supplies and Stationery Stores
		453220	Gift, Novelty, and Souvenir Stores
		453910	Pet and Pet Supplies Stores
		453920	Art Dealers
		453991	Tobacco Stores
		812113	Nail Salons
		812191	Diet and Weight Reducing Centers
		812199	Other Personal Care Services
		453998	All Other Miscellaneous Store Retailers (except Tobacco Stores)
	532000 - Rental Goods	532210	Consumer Electronics and Appliances Rental

Destination Category	Reclassified Specific Destination	Original NAICS Code & Classification	
Shopping *(continued)*	532000 - Rental Goods *(continued)*	532220	Formal Wear and Costume Rental
		532230	Video Tape and Disc Rental
		532291	Home Health Equipment Rental
		532292	Recreational Goods Rental
		532299	All Other Consumer Goods Rental
		532310	General Rental Centers
	453310 - Used Merchandise Stores	453310	Used Merchandise Stores
	454000 - Nonstore Retailers	454111	Electronic Shopping
		454112	Electronic Auctions
		454113	Mail-Order Houses
		454210	Vending Machine Operators
		454311	Heating Oil Dealers
		454312	Liquefied Petroleum Gas (Bottled Gas) Dealers
		454319	Other Fuel Dealers
		454390	Other Direct Selling Establishments
Food, Groceries, and Restaurants	445110 - Supermarkets and Grocery Stores	445110	Supermarkets and Other Grocery (except Convenience) Stores
	445121 - Convenience Stores (including those in gas stations)	445120	Convenience Stores
	452910 - Warehouse Clubs and Supercenters	452910	Warehouse Clubs and Supercenters
	445200 - Specialty Food Stores	445210	Meat Markets
		445220	Fish and Seafood Markets
		445230	Fruit and Vegetable Markets
		445291	Baked Goods Stores
		311811	Retail Bakeries
		445292	Confectionery and Nut Stores
		445299	All Other Specialty Food Stores
		445310	Beer, Wine, and Liquor Stores
	722100 - Full-Service Restaurant	722110	Full-Service Restaurants
		722211	Limited-Service Restaurants
		722212	Cafeterias
	722213 - Snack and Nonalcoholic Beverage Bars (i.e. coffee shops)	722213	Snack and Nonalcoholic Beverage Bars
	722300 - Other Food Service	722310	Food Service Contractors
		722320	Caterers
		722330	Mobile Food Services
	722410 - Bars/Nightclubs	722410	Drinking Places (Alcoholic Beverages)
Fitness and Recreation	713940 – Fitness and Recreation Centers	713940	Fitness and Recreational Sports Centers
		611620	Sports and Recreation Instruction
Parks	using land use data		
Entertainment	512130 – Movie Theaters	512131	Motion Picture Theaters (except Drive-Ins)
		512132	Drive-In Motion Picture Theaters

Destination Category	Reclassified Specific Destination	Original NAICS Code & Classification	
Entertainment *(continued)*	711000 – Performing Arts and/or Spectator Sports	711110	Theater Companies and Dinner Theaters
		711120	Dance Companies
		711130	Musical Groups and Artists
		711190	Other Performing Arts Companies
		711211	Sports Teams and Clubs
		711212	Racetracks
		711219	Other Spectator Sports
	712000 – Amusement and Recreation	712110	Museums
		712120	Historical Sites
		712130	Zoos and Botanical Gardens
		713110	Amusement and Theme Parks
		713120	Amusement Arcades
		713210	Casinos (except Casino Hotels)
		713290	Other Gambling Industries
		713950	Bowling Centers
		713990	All Other Amusement and Recreation Industries
		721120	Casino Hotels
	713900 - Outdoor Recreation (Camps, Golf, etc.)	713910	Golf Courses and Country Clubs
		713920	Skiing Facilities
		713930	Marinas
		721211	RV (Recreational Vehicle) Parks and Campgrounds
		721214	Recreational and Vacation Camps (except Campgrounds)
Health Care	446110 - Pharmacies and Drug Stores	446110	Pharmacies and Drug Stores
	446190 - Health, Medical, and Personal Care Stores	446130	Optical Goods Stores
		446191	Food (Health) Supplement Stores
		446199	All Other Health and Personal Care Stores
	621111 – Doctors' Offices	621111	Offices of Physicians (except Mental Health Specialists)
	621210 – Dentists' Offices	621210	Offices of Dentists
	621100 – Specialized Health Care Facility/Office	621112	Offices of Physicians, Mental Health Specialists
		621310	Offices of Chiropractors
		621320	Offices of Optometrists
		621330	Offices of Mental Health Practitioners (except Physicians)
		621340	Offices of Physical, Occupational and Speech Therapists, and Audiologists
		621391	Offices of Podiatrists
		621399	Offices of All Other Miscellaneous Health Practitioners
		621410	Family Planning Centers

Destination Category	Reclassified Specific Destination	Original NAICS Code & Classification
Health Care ***(continued)***	621100 – Specialized Health Care Facility/Office *(continued)*	621420 Outpatient Mental Health and Substance Abuse Centers 621491 HMO Medical Centers 621492 Kidney Dialysis Centers 621493 Freestanding Ambulatory Surgical and Emergency Centers 621498 All Other Outpatient Care Centers 622210 Psychiatric and Substance Abuse Hospitals 622310 Specialty (except Psychiatric and Substance Abuse) Hospitals
	621500 – Health Services	621511 Medical Laboratories 621512 Diagnostic Imaging Centers 621610 Home Health Care Services 621910 Ambulance Services 621991 Blood and Organ Banks 621999 All Other Miscellaneous Ambulatory Health Care Services
	622110 – General Hospital	622110 General Medical and Surgical Hospitals
	623000 – Residential Care Facilities	623110 Nursing Care Facilities 623210 Residential Mental Retardation Facilities 623220 Residential Mental Health and Substance Abuse Facilities 623311 Continuing Care Retirement Communities 623312 Homes for the Elderly 623990 Other Residential Care Facilities
Library	519120 – Libraries	519120 Libraries and Archives
Post Office	491110 - Postal Service	491110 Postal Service
Financial Services	522000 – Banking	522110 Commercial Banking 522120 Savings Institutions 522130 Credit Unions 522190 Other Depository Credit Intermediation
	521000 – Other Financial/Insurance	*All other Financial and Insurance (520000) classifications*
Education	624410 – Child Care	624410 Child Day Care Services
	611110 - Elementary and Secondary Schools	611110 Elementary and Secondary Schools
	611200 - Postsecondary Schools and Technical Training	611210 Junior Colleges 611310 Colleges, Universities, and Professional Schools 611410 Business and Secretarial Schools 611420 Computer Training 611430 Professional and Management Development Training 611511 Cosmetology and Barber Schools 611512 Flight Training

Destination Category	Reclassified Specific Destination	Original NAICS Code & Classification	
Education *(continued)*	611600 - Other Schools and Instruction	611513	Apprenticeship Training
		611519	Other Technical and Trade School
		611610	Fine Arts Schools
		611630	Language Schools
		611691	Exam Preparation and Tutoring
		611692	Automobile Driving Schools
		611699	All Other Miscellaneous Schools and Instruction
Personal Services	624000 – Social Services	624110	Child and Youth Services
		624120	Services for the Elderly and Persons with Disabilities
		624190	Other Individual and Family Services
		624210	Community Food Services
		624221	Temporary Shelters
		624229	Other Community Housing Services
		624230	Emergency and Other Relief Services
		624310	Vocational Rehabilitation Services
	721110 - Hotels and Traveler Accommodation	721110	Hotels (except Casino Hotels) and Motels
		721191	Bed-and-Breakfast Inns
		721199	All Other Traveler Accommodation
	721310 - Rooming and Boarding Houses	721310	Rooming and Boarding Houses
	811210 - Electronics Repair and Maintenance	811211	Consumer Electronics Repair and Maintenance
		811212	Computer and Office Machine Repair and Maintenance
		811213	Communication Equipment Repair and Maintenance
		811219	Other Electronic and Precision Equipment Repair and Maintenance
	323000 - Copying and Printing Services	323114	Quick Printing
		561439	Other Business Service Centers
	811310 - Commercial and Industrial Repair and Maintenance	811310	Commercial and Industrial Machinery and Equipment (except Automotive and Electronic) Repair and Maintenance
	811400 - Personal and Household Goods Repair and Maintenance	811411	Home and Garden Equipment Repair and Maintenance
		811412	Appliance Repair and Maintenance
		811420	Reupholstery and Furniture Repair
		811430	Footwear and Leather Goods Repair
		811490	Other Personal and Household Goods Repair and Maintenance

Destination Category	Reclassified Specific Destination	Original NAICS Code & Classification
Personal Services *(continued)*	812110 - Barber Shops and Beauty Parlors	812111 Barber Shops 812112 Beauty Salons
	812200 - Death Care Services	812210 Funeral Homes and Services 812220 Cemeteries and Crematories
	812300 - Laundry and Drycleaners	812310 Coin-Operated Laundries and Drycleaners 812320 Drycleaning and Laundry Services (except Coin-Operated)
	812900 - Other Personal Services	812910 Pet Care (except Veterinary) Services 812921 Photofinishing Laboratories (except One-Hour) 812922 One-Hour Photofinishing 812990 All Other Personal ServicesProfessional, and Similar Organizations
	813110 - Religious Organizations	813110 Religious Organizations
	813300 - Civic, Social, Political, Business and Other Organizations	813311 Human Rights Organizations 813312 Environment, Conservation and Wildlife Organizations 813319 Other Social Advocacy Organizations 813410 Civic and Social Organizations 813910 Business Associations 813920 Professional Organizations 813930 Labor Unions and Similar Labor Organizations 813940 Political Organizations 813990 Other Similar Organizations (except Business, Professional, Labor, and Political Organizations)
	492000 - Couriers and Messengers	492110 Couriers 492210 Local Messengers and Local Delivery
Airport	use land use data	
Transportation	480000 - Transportation	*All Transportation (480000) classifications not included in previous destination categories*
Wholesale	420000 - Wholesale Trade	*All Wholesale Trade (420000) classifications not included in previous destination categories*
Information	510000 – Information and Communications	*All Information and Communications (510000) classifications not included in previous destination categories*
Real Estate / Rental	531000 - Real Estate, Rental, and Leasing	*All Real Estate, Rental, and Leasing (530000) classifications not included in previous destination categories*
Professional Services	541110 - Professional, Scientific and Technical Services	*All Professional, Scientific and Technical Services (540000) classifications not included in previous destination categories*
Administration / Support	560000 - Administration and Support	All *Administrative and Support and Waste Management and Remediation Services (560000) classifications not included in previous destination categories*

Destination Category	Reclassified Specific Destination	Original NAICS Code & Classification
Other	100000 - Other	*Agriculture, Forestry, Fishing, and Hunting* *Mining* *Utilities* *Construction* *Manufacturing* *Management of Companies and Enterprises* *Public Administration*
Unclassified	999999 - Unclassified	999999 Unclassified

APPENDIX B:
NAICS Reclassification SPSS Syntax

‘ To use this syntax, open the Dun & Bradstreet business table in SPSS
‘ Click File…New…Syntax
‘ Copy and paste all of the text into the syntax window that opens
‘ Click the Run button
‘ The syntax will run and create three new columns in the table
‘ “Dest_code” is the specific destination recode.
‘ See Appendix A for a list of the titles of each specific destination recode
‘ (i.e. Dest_code 441100 = Motor Vehicle Dealers and Maintenance)
‘ “Dest_cat is the destination category code
‘ “Cat_Descrip” is the text label for each destination category (i.e. “Shopping”)

```
‘ Recode Dun & Bradstreet primary NAICS classification to new specific destination code
‘ Place new code in new column titled ‘Dest_code’
RECODE
  primarynaics
  (522210 thru 525990=521000)  (481111 thru 488999=480000)  (420000 thru 429999=420000)
  (510000 thru 512129=510000)  (512133 thru 519999=510000)  (531000 thru 532209=531000)
(532300 thru 539999=531000)  (541100 thru 541999=541110)
  (561100 thru 562999=560000)  (921110 thru 928120 = 921000)
  (551000 thru 551114=551000) (999990=Copy) (441110=441100)  (441120=441100)
(441210=441100)  (441221=441100)
  (441222=441100)  (441229=441100)  (441310=441100)  (441320=441100)
  (532111=441100)  (532112=441100)  (532120=441100)  (811111=441100)
  (811112=441100)  (811113=441100)  (811118=441100)  (811121=441100)
  (811122=441100)  (811191=441100)  (811192=441100)  (811198=441100)
  (442110=442000)  (442210=442000)  (442291=442000)  (442299=442000)
  (443111=443000)  (443112=443000)  (443120=443000)  (443130=443000)
  (444110=444100)  (444130=444100)  (444120=444101)  (444190=444101)
  (444210=444101)  (453930=444101)  (444220=444220)  (446120=446120)
  (447190=447190)  (447110=447190)  (448110=448100)  (448120=448100)
  (448130=448100)  (448140=448100)  (448150=448100)  (448190=448100)
  (448210=448100)  (448310=448100)  (448320=448100)  (451110=451100)
  (451120=451100)  (451130=451100)  (451140=451100)  (451220=451100)
  (451211=451210)  (451212=451210)  (452111=452100)  (452112=452100)
  (452990=453000)  (453110=453000)  (453210=453000)  (453220=453000)
  (453910=453000)  (453920=453000)  (453991=453000)  (812113=453000)
  (812191=453000)  (812199=453000)  (453998=453000)  (532210=532000)
  (532220=532000)  (532230=532000)  (532291=532000)  (532292=532000)
  (532299=532000)  (532310=532000)  (453310=453310)  (454111=454000)
  (454112=454000)  (454113=454000)  (454210=454000)  (454311=454000)
  (454312=454000)  (454319=454000)  (454390=454000)  (445110=Copy)
  (445120=Copy)  (452910=Copy)  (445210=445200)  (445220=445200)
  (445230=445200)  (445291=445200)  (311811=445200)  (445292=445200)
  (445299=445200)  (445310=445200)  (722110=722100)  (722211=722100)
  (722212=722100)  (722213=Copy)  (722310=722300)  (722320=722300)
  (722330=722300)  (722410=Copy)  (713940=713940)  (611620=713940)
  (512131=512130)  (512132=512130)  (711110=711000)  (711120=711000)
  (711130=711000)  (711190=711000)  (711211=711000)  (711212=711000)
  (711219=711000)  (712110=712000)  (712120=712000)  (712130=712000)
  (713110=712000)  (713120=712000)  (713210=712000)  (713290=712000)
  (713950=712000)  (713990=712000)  (721120=712000)  (713910=713900)
```

```
  (713920=713900)  (712190=713900)  (713930=713900)  (721211=713900)  (721214=713900)
  (446110=Copy)  (446130=446190)  (446191=446190)  (446199=446190)
  (621111=Copy)  (621210=Copy)  (621112=621100)  (621310=621100)
  (621320=621100)  (621330=621100)  (621340=621100)  (621391=621100)
  (621399=621100)  (621410=621100)  (621420=621100)  (621491=621100)
  (621492=621100)  (621493=621100)  (621498=621100)  (622210=621100)
  (622310=621100)  (621511=621500)  (621512=621500)  (621610=621500)
  (621910=621500)  (621991=621500)  (621999=621500)  (622110=Copy)
  (623110=623000)  (623210=623000)  (623220=623000)  (623311=623000)
  (623312=623000)  (623990=623000)  (519120=Copy)  (491110=Copy)
  (522110=522000)  (522120=522000)  (522130=522000)  (522190=522000)
  (624410=Copy)  (611110=Copy)  (611210=611200)  (611310=611200)
  (611410=611200)  (611420=611200)  (611430=611200)  (611511=611200)
  (611512=611200)  (611513=611200)  (611519=611200)  (611610=611600)
  (611630=611600)  (611691=611600)  (611692=611600)  (611699=611600)
  (624110=624000)  (624120=624000)  (624190=624000)  (624210=624000)
  (624221=624000)  (624229=624000)  (624230=624000)  (624310=624000)
  (721110=721110)  (721191=721110)  (721199=721110)  (721310=721110)
  (811211=811210)  (811212=811210)  (811213=811210)  (811219=811210)
  (323114=323000)  (561439=323000)  (811310=Copy)  (811411=811400)
  (811412=811400)  (811420=811400)  (811430=811400)  (811490=811400)
  (812111=812110)  (812112=812110)  (812210=812200)  (812220=812200)
  (812310=812300)  (812320=812300)  (812910=812900)  (812921=812900)
  (812922=812900)  (812990=812900)  (813110=Copy)  (813311=813300)
  (813312=813300)  (813319=813300)  (813410=813300)  (813910=813300)
  (813920=813300)  (813930=813300)  (813940=813300)  (813990=813300)
  (492110=492000)  (492210=492000)  (812930=480000)  (Else=100000) INTO  dest_code .
VARIABLE LABELS dest_code 'Destination Type Code'.
EXECUTE .



‘ Recode specific destination codes into larger destination categories
‘ Place new code in new column titled ‘Dest_cat’
RECODE
  dest_code
  (441100=1)  (442000=1)  (443000=1)  (444100=1)  (444101=1)  (444220=1)
  (446120=1)  (447190=1)  (448100=1)  (451100=1)  (451210=1)  (452100=1)
  (453000=1)  (532000=1)  (453310=1)  (454000=1)  (445110=2)  (445120=2)
  (452910=2)  (445200=2)  (722100=2)  (722213=2)  (722300=2)  (722410=2)
  (713940=3)  (512130=5)  (711000=5)  (712000=5)  (713900=5)  (446110=6)
  (446190=6)  (621111=6)  (621210=6)  (621100=6)  (621500=6)  (622110=6)
  (623000=6)  (519120=7)  (491110=8)  (522000=9)  (521000=9)  (624410=10)
  (611110=10)  (611200=10)  (611600=10)  (624000=11)  (721110=11)  (811210=11)
   (323000=11)  (811310=11)  (811400=11)  (812110=11)  (812200=11)
  (812300=11)  (812900=11)  (813110=11)  (813300=11)  (492000=11)  (480000=13)
   (420000=14)  (510000=15)  (531000=16)  (541110=17)  (560000=18)
  (921000=18)  (551000=18)  (100000=19)  (999990=20)  INTO  dest_cat .
VARIABLE LABELS dest_cat 'Destination Type Category'.
EXECUTE .
```

```
‘ Create label field for larger destination categories
‘ Place label in new column titled ‘Cat_descrip’
STRING cat_descrip (A30) .
RECODE
  dest_cat
  (1='Shopping')  (2='Food')  (3='Fitness')  (4='Parks')  (5='Entertainment')
   (6='Health Care')  (7='Library')  (8='Post Office')  (9='Bank or'+
 ' Insurance')  (10='Education')  (11='Services')  (12='Airport')
  (13='Transportation')  (14='Wholesale Trade')  (15='Information')  (16='Real'+
 ' Estate and Rental')  (17='Professional Services')  (18='Administration and'+
 ' Support')  (19='Other')  (20='Unclassified')  INTO  cat_descrip .
VARIABLE LABELS cat_descrip 'Destination Category Description'.
EXECUTE .
```

APPENDIX C:
Instructions for Matching Dun & Bradstreet and Parcel Point Data Using Oracle Database (provided by CS GRA)

1- Text manipulation using gawk (open source software) and Textpad to modify input files to be in ready format for the next stage.
2- Load data (business data and parcel points) into database (I have used Oracle Express edition), throwing away business addresses with not matched zip code.
3- Parse business addresses and parcel points using program with c# and load the parsed data to data base to join, the parsing program also standardizes street names using Alternate name table. The program replaces abbreviation (e.g. COUNTY ROAD ïƒ CO RD)
4- As the database is read only, we create an index over all attributes and perform a join query to get the matched addresses.
5- When the grid has been introduced; other program has been developed to perform the search within the grid only.

Recommendation for Next data

1- Addresses for parcel point and business address should be geo-coded to make the search more narrowed to improve the speed to find matched pair, no grid is necessary.
2- Alternate name should be standardized; I think the address get from geo-coding can be used to standardize the names.
3- Street names table that holds the street name along with direction and type.

APPENDIX D:
Working with D&B – Parcel Matches Made by CS Student

Files returned by the CS student are text files with three columns of information, the Duns number, the parcel pin that the business was matched to, and a "score" showing the accuracy of the match (similar to the score given to geocoding matches in ArcGIS).

Import this table into Access along with the original table of all D&B businesses (located at D:\Dunn&Bradstreet\originaldatadownload) and the table of parcel points for the 7 county metro area (the original parcel point file had duplicate points and did not contain accurate X/Y coordinates, so a cleaned version of this file with X/Y coordinates for each point in UTM zone 15 projection is available at D:\Dunn&Bradstreet\originaldatadownload\D&Bbusinesses_Metroparcels.mdb\Parcel_points_CLEANED).

Create a query that adds the rest of the D&B and parcel attributes to these three columns.

Select about 20 random businesses and double check that the D&B and parcel addresses are good matches. Even building numbers should be matched to even building numbers and odds to odds. Make sure that highways are not being matched to county road parcels and vice versa.

Save the query as a new table by right clicking on the name of the query and exporting the query as a dbf table.

Open ArcGIS, add the table, right click on its name, and click Add XY Data…

In the dialogue box, select the table fields with the X and Y data for the parcel points and select the UTM 1983 Zone 15 projection. Click OK.

Right click on the resulting XY event layer, click Data…Export, and save the points as a shapefile.

APPENDIX E:
Matching D&B Businesses to Parcel Points Using the Adapted Geocoding Method

1. Add the Lawrence Group street centerline shapefile (D:\Networks\TLG_metro.shp) and the table of unmatched Dun & Bradstreet business addresses to ArcGIS.

2. Right click on the name of the table and select Geocode Addresses…

3. In the address locator dialogue box select the locator named "TLG 2006". If this locator is not listed in the box, click the Add button. Navigate to the folder G:\KKRIZEK\Data\GISData\AddressLocators or navigate to your local Address Locators folder (listed below the lettered drives in the pull-down "Look in:" menu). If you cannot find the address locator, click Cancel and you can create a new one easily using the directions on ESRI's website: http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?id=4533&pid=4531&topicname=Creating_an_address_locator

4. Click OK

5. In the Geocode Addresses dialogue box, use the pull-down menus to indicate the fields that different address data elements are located in.

6. Type in the location where you want the output shapefile to be stored and the name of the file.

7. Click the "Geocoding Options…" button and uncheck "Match if candidates tie". You can also adjust the spelling sensitivity and the minimum score for something to be considered a match in this window if you want. Click OK.

8. Click OK and ArcGIS will geocode the table and add a shapefile containing the geocoded points to the map. The Review/Rematch Addresses window will appear and show the geocoding success rate.

9. If there are addresses that are tied, click the "Addresses with candidates tied" radio button and click "Match Interactively". For each of the addresses with tied candidates, highlight the address that is the closest match in the candidates window, then click Match. After matching all of the tied addresses with tied candidates, click Done.

10. Next, click the "Unmatched Addresses" radio button and click "Match Interactively". Go through each of these addresses and correct any spelling errors or errors in ArcGIS's address standardization as necessary to find the closest match candidate for each business, highlight it in the candidates window, then click Match. After you have found a match for as many unmatched businesses as possible, click Done.

11. After geocoding the Dun & Bradstreet businesses, add the parcel points shapefile to ArcGIS (D:\2007parcels\Metro\all_7_parcel_pts.shp)
12. Since the geocoded Dun & Bradstreet points are usually drawn reasonably close to the correct parcel and the parcel points shapefile has over 1 million features, you will want to create a new shapefile that just contains the parcel points that are located within a

reasonable distance of the geocoded Dun & Bradstreet businesses. To do this, click on the Selection menu, then click on Select by Location

13. In the Select by Location window, select features from the parcel points shapefile that are within a distance of 300 meters of features from the geocoded Dun & Bradstreet shapefile. Click OK. (This may take a couple minutes to process.)

14. Right click on the parcel points shapefile and click Open Attribute Table

15. Click on the Option button at the bottom of the table window and select Export…

16. Make sure the pull-down menu at the top of the Export window says "Selected Records" and enter the location where you want to save the selected features and the name of the new table. Click OK.

17. Click Yes in the pop-up window to add the new table to the current map.

18. Since all of the parcel address data elements are contained in separate fields, you will have to create one field that contains all of the street address data before you can geocode it. Right click on the new parcel points table and click Open

19. Click the Option button at the bottom of the table and click Add Field…

20. Name the field "Address2" and designate it a text field, length 50. Click OK.

21. Select the Address2 column in the table, right click on the column title, and select Field Calculator

22. In the text box, write the following: [BLDG_NUM] & " " & [PREFIX_DIR] & " " & [PREFIXTYPE] & " " & [STREETNAME] & " " & [STREETTYPE] & " " & [SUFFIX_DIR]

23. Click Calculate. Close the table.

24. Repeat steps 2-10 to geocode the new table of parcel points. (Geocoding will probably take a while for the parcel dataset.)

25. Right click on the geocoded Dun & Bradstreet shapefile and click Joins and Relates, Join…

26. In the pull-down menu at the top of the window select "Join based on spatial location"

27. Select the geocoded parcel point shapefile as the layer to be joined.

28. Click the second bullet point so that each business will be given all of the attributes of the parcel point that is closest to it and a field indicating the distance between it and the closest parcel point.

29. Indicate the location where you want the output shapefile to be saved and its name. Click OK. (this might take a while to process)

30. Query the resulting shapefile for records with a value of zero in the distance field. These records are exact matches and contain the unique ID and address data for both the business and the parcel it's located within. They can be linked back to the original parcel point location (as opposed to the geocoded location) using the parcel pin number and the directions found in Appendix D.

APPENDIX F:
Grid Cell Matching Methodology

To create a grid using these instructions, you must first download and install the Hawth's Tools extension to ArcGIS on your computer. The extension and accompanying documentation can be downloaded for free from:
http://www.spatialecology.com/htools/download.php

1. Open ArcGIS and add a shapefile of the boundaries of the 7-County Metro (G:\KKRIZEK\Data\GISData\Boundaries\counties.shp)

2. In order to make sure that points located on the borders of the 7-County Metro are included in a grid cell, you must first make create a buffer. Open ArcToolbox and select Analysis Tools, Proximity, Buffer. Create a single, 300 meter buffer of the 7-County Metro and save it as "Metro_buffer".

3. Click the Tools menu and select customize. Select Hawths Tools in the Toolbar list and drag and drop it next to an existing toolbar.

4. Click Hawth's Tools, Sampling Tools, Create Vector Grid.

5. Select the Metro Buffer as the extent, 300 meters as the distance between lines, make sure Save as Polygon is checked, and set the projection as 1983 UTM Zone 15. Click OK.

6. Add the shapefile of the geocoded, unmatched Dun & Bradstreet businesses to the map. Right click on the shapefile, click Joins and Relates, Join…

7. Select Join based on spatial location from the pull-down menu at the top of the window, select the grid as the file to join to, and click the correct radio button so that each business point is given all of the attributes of the polygon it falls within. The resulting shapefile will contain all of the Dun & Bradstreet attributes as will as the unique ID of the grid cell the business falls within.

8. Repeat steps 6 and 7 with the parcel points file.

9. Send both tables to the Computer Science Research Assistant for matching.