

THE DEVELOPMENT OF AN ELECTRONIC DATA COLLECTION TOOL
AND COMPARISON OF THE ELECTRONIC AND MANUAL
METHODS OF LAND USE INVENTORY

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of City and Regional Planning & Master of Science in Engineering

(Transportation Planning Specialization)

by

Wesley Catanzaro

May, 2013

© 2013

Wesley Catanzaro

ALL RIGHTS RESERVED

Page ii

COMMITTEE MEMBERSHIP

TITLE: The Development of an Electronic Data Collection Tool and Comparison of the Electronic and Manual Methods of Land Use Inventory

AUTHOR: Wesley Catanzaro

DATE SUBMITTED: May, 2013

COMMITTEE CHAIR: Dr. Cornelius Nuworsoo, Associate Professor
City and Regional Planning Department

COMMITTEE MEMBER: Dr. William Riggs, Assistant Professor
City and Regional Planning Department

COMMITTEE MEMBER: Dr. Anurag Pande, Assistant Professor
Civil and Environmental Engineering Department

ABSTRACT

The Development of an Electronic Data Collection Tool and Comparison of the Electronic and Manual Methods of Land Use Inventory

Wesley Catanzaro

An important component of Planning Information Systems for municipal planning agencies is a comprehensive land use inventory that provides information on the location, distribution, and intensity of land uses throughout a community. This data is a necessary prerequisite for the informed creation of planning documents such as General Plans, Specific Plans, Housing Inventories, and Climate Action Plans. Beyond location, distribution and intensity of land uses, planners may also wish to incorporate additional information at the parcel level, such as the number of housing or commercial units, building condition, and/or access and connectivity to adjacent streets. Because some of this information is best observed in the field, agencies require methods of collecting this data that will ensure data precision, accuracy, and consistency, while minimizing data collection and processing time. Electronic data collection tools that are compatible with Geographic Information Systems provide a potential solution that can facilitate these desired data collection parameters. This research illustrates the development of an electronic data collection tool that planning agencies may utilize within various planning efforts, and compares the efficiency of the tool to traditional ‘pen-and-paper’ data collection methods in terms of time savings. It is recommended that planning agencies widely adopt and implement electronic tools for land use data collection, for the demonstrated benefits related to data consistency and reduced data collection time in the field.

Keywords: data collection, land-use survey, land-use planning, land-use inventory, survey instrument, geographic information systems

ACKNOWLEDGMENTS

First, I would like to thank professors Jason Henderson, Qian Guo, and XiaoHang Liu at the Geography Department at San Francisco State University their exemplary instruction regarding urban geography, city and transportation planning, and spatial analysis methods. My initial exposure to these topics which I received in your courses was a contributing factor in my decision to pursue a career in planning, and I am deeply indebted to you for the inspiration.

I would also like to thank the various faculty members in the City and Regional Planning and Civil Engineering departments at Cal Poly for providing excellent instruction and guidance in the topic of planning and transportation engineering, including Cornelius Nuworsoo, William Riggs, Anurag Pande, Kimberly Mastako, and Chris Clark. The lessons, experience and guidance that I have acquired while studying here have been invaluable.

Finally, I would like to thank Lana Russell-Hurd. Your companionship during this program has been of immeasurable worth and has helped me to persevere. Thank you for taking this journey with me.

TABLE OF CONTENTS

	Page
List of Tables	vii
List of Figures	viii
Chapter	
1. Introduction.....	1
2. Background and Literature Review	4
3. Key Criteria for Land Use Survey Instrument: Design and Existing Technology ..	9
Data Quality	10
Customization	11
Interoperability/Compatibility	12
Ease of Use	12
Hardware Considerations.....	13
Software Considerations	15
4. Description of Survey Instrument Specifications and Features	17
Data Need Assessment.....	17
Hardware and Software Selection and Specifications	20
Software Features.....	23
Customized Forms	23
Data Merge.....	27
Compatibility with Microsoft Access	27
5. Analysis Method and Findings	29
Testing Two Independent Sample Means.....	31
Results.....	32
6. Conclusion and Recommendations for Further Application	36
References.....	38
Appendices.....	39
Appendix A: Sample output tables for t-tests.....	40
Appendix B: Land Use Survey Instrument Configuration Guide.....	42

LIST OF TABLES

Table	Page
1	Characteristics of hardware devices for field data collection 14
2.	Data Needs for Newark Land-Use Inventory 19
3.	HanDBase Features and Specifications 23
4.	Data Entry Times and summary statistics for residential and commercial parcels in San Luis Obispo, CA 33
5.	Summary Statistics for combined land use data by the two methods 34
6.	Summary of Difference of Means Test 34
7.	Potential estimated time savings for a sample of commercial and residential land uses in San Luis Obispo 35
A-1.	Dataset Table 40
A-2.	Group Statistics for All Land Uses Test 40
A-3.	Independent Samples Test for All Land Uses 40
A-4.	Group Statistics for Residential Land Uses Test 41
A-5.	Independent Samples Test for Residential Land Uses 41
A-6.	Group Statistics for Commercial Land Uses Test 41
A-7.	Independent Samples Test for Commercial Land Uses 41

LIST OF FIGURES

Figure		Page
1.	The Motorola MC65 Rugged Mobile	14
2.	The Lenovo Thinkpad.....	14
3.	Apple’s iPhone and iPad.....	15
4.	Customized Land Use data collection form in HanDBase for iPad.....	25
5.	Customized Structure data collection form in HanDBase for iPad.	26

CHAPTER 1

Introduction

Effective planning practice hinges on the collection, analysis, and interpretation of a variety of information for the development and implementation of plans, policies, and programs that guide urban development. Data used in local planning efforts may come from various sources. For instance, federal agencies may be a key source of information for planning efforts that spearhead environmental protection, in the case of nationally administrated programs such as the National Environmental Protection Act. State agencies may also provide important information to local planners via similar environmental protection programs or other programs that are administered at the state level. While state and federal oversight and administrative support does affect some aspects of planning practice, local jurisdictions such as cities and counties still retain considerable control on the local development and implementation of plans, policies, and programs. Local planners and the local agencies and councils they represent and work with are the ultimate arbiters in how planning is practiced and development implemented at the local level. As such, they are also a key factor in the collection and development of local data that is not used by federal and state agencies.

Land-use data is one example of local data that is of primary importance for local and regional planning agencies. General Plans, Specific Plans, Climate Action Plans, and Housing inventories are locally produced documents that depend on assessments of existing and proposed locations, distributions and intensities of various land uses throughout a community. Thus, the successful design and implementation of these

planning initiatives depends on the accurate and comprehensive collection of land-use data, which may be difficult and costly to collect and maintain.

This research focuses on a method of collecting land-use data through direct observation that is expected to reduce data collection costs (in terms of time) for local agencies, and increase the consistency and precision of data within a land-use inventory. Specifically, this research examines how electronic data collection as facilitated through the use of portable electronic devices in the field can reduce data collection times when compared to the traditional method of data collection via pen-and-paper. The method and corresponding survey instrument described in this thesis was successfully deployed for the collection of land-use data as part of the City and Regional Planning graduate studio course at California Polytechnic State University, San Luis Obispo. In addition to reducing data collection time, additional benefits over traditional land-use data collection include increased quality assurance and consistency of attribute values through the implementation of a “decision-tree” survey design, as well as a geodatabase compatible format that is easily exported from the survey instruments and into a Geographic Information System.

This document is divided into six chapters. In addition to this introductory chapter, the remainder of the thesis is organized as follows:

Chapter 2 presents a review of the relevant literature relating to land-use data collection and compares secondary and automated methods of land-classification with direct observation methods.

Chapter 3 continues from the argument developed in Chapter 2 regarding the preference for direct observation, and describes key criteria that should be considered in

the design of an electron survey instrument. This chapter also presents a brief review of existing hardware and software that the planning practitioner or data manager may consider in survey instrument design.

Chapter 4 describes the survey instrument that was designed and deployed for land-use data collection during the City and Regional Planning studio courses that were previously mentioned. This chapter describes the data needs of the studio courses, and specifications and features of the selected hardware and software that was used to collect the land-use data for the courses.

Chapter 5 describes the methodology for comparing data collection times of the electronic method and traditional method, and presents the results of the difference of mean time analysis.

Chapter 6 concludes the thesis and provides recommendations for further application of the method described. The Appendix provides detailed information regarding the development and application of the survey instrument for professional practice.

CHAPTER 2

Background and Literature Review

Modern city and regional planning encompasses the development of policies, programs, and regulations which guide the orderly, informed growth of jurisdictions while also safeguarding the health, safety and welfare of their citizens. The interface between historic, existing, and projected or desired patterns of land-use, population and economic change, and the policies and corresponding implementation measures is what drives growth over time. Knowledge of existing conditions is especially important to the planning process, because it forms the basis of identifying current or projected deficiencies in the urban space as well as the resources that must be prioritized or rectified to meet the goals of the community. Dandekar (2003) acknowledges "A planner needs to get good information that provides insight into the problem, and enables changes in policy and action that can remedy or ameliorate the problem or turn it into an opportunity for change and development."

In terms of land-use planning and programming, the development of an existing conditions land-use inventory is an important first step. An ideal land-use inventory will be "in narrative and tabular form, and will indicate the amount, type, and intensity and/or density of existing land uses." (Berke, 2006). In addition to representing accurate baseline conditions, a jurisdiction's land-use inventory system should also facilitate regular updating via institutional methods (processing of building permits, certificates of occupancy, site plans, and subdivision approvals) or through field surveys, aerial photography or remote sensing images. Geospatial databases are optimal tools to

containing, utilizing and updating land-use data, as they allow planners to store a variety of pertinent information at different scales (parcel, census block or tract, neighborhood), and information within them can be easily queried, summarized, and represented cartographically. Knowing that a geospatial database is the desired method of storage of land-use data, it is important to consider the methods by which such a database will be meaningfully and reliably populated, either through a comprehensive, jurisdiction-wide inventory or updated through spot-checks for neighborhood or area planning purposes.

Whether for general plans, specific plans, housing elements/inventories, or climate action plans, the traditional method of conducting a land-use inventory involves the collection of data in the field. Typically, the agency interested in acquiring the data will deploy several people to survey existing conditions. Surveyors will take base maps, pens and markers, and survey forms to capture any number of variables that are to be utilized for planning (Kelly, 2010). Depending on the extent of land to be surveyed and the amount of data to be collected, this process may be prohibitively expensive in terms of money and time.

As mentioned above, alternate methods of collecting and updating land-use are available to planners, but while they may be less consuming in terms of time and money, they may not provide the level of detail and comprehensiveness desired. Institutional methods may be tailored to collect a wide variety of information, but they have little applicability to parcels for which the owner is not actively seeking development or subdivision approval. Property tax assessors may have computerized databases indicating general land-use categories, but these may not contain the level of detail required for comprehensive planning and/or may not be up-to-date (Kelly, 2010). Aerial satellite

imagery and "Streetview" programs such as those provided by companies such as Google and Microsoft (Google Earth/Maps, Bing Maps) have considerable appeal for their ease of use, but may also be subject to issues of current-ness and their limited ability to provide the planner with highly detailed information that may be desired at the scale of the individual parcel. For instance, a planner may find it difficult to determine the number of residential units within a large multifamily complex via satellite imagery. Street level pictures may be available, but they may have limited utility depending on the configuration of the building to the street, or access restrictions that prohibit street level photography altogether.

Remote sensing is another tool that may be used for land-use classification, but it presents challenges in terms of accurately classifying land-use at the level of detail that jurisdictions may desire for a parcel based land-use inventory. Zhan *et al*(2000) indicated that "The main reason is that an urban land use classification scheme is developed based on socioeconomic functionality instead of biophysical characteristics that are closely related to the spectral reflectance detected by remote sensing images (as cited in Wu *et al*, 2009). Wu *et al* (2009) developed an approach for detailed urban land-use classification based on land parcel attributes derived from GIS and remote sensing data. Their method classified urban land-uses with a 95.6 percent accuracy overall. However, office, industrial, civic, and transportation land-use classification was relatively inaccurate compared to single- and multi-family residential, commercial, open space, and undeveloped land-uses. Hu & Wang (2013) conducted additional research on increasing the accuracy of office, industrial, civic, and transportation parcels, but were only able to achieve an overall classification accuracy of 61.68%. These results should not preclude

the application of remote-sensing for urban land-use classification. Remote sensing techniques as employed in the research referenced here could certainly be valuable as a potential starting point for determining general land-use classes where an inventory has not been continually maintained or does not exist.

However, it is important to consider the limitations of remote sensing vis-à-vis the objectives of the land-use inventory that a jurisdiction aims to undertake. As with previously mentioned satellite imagery and ‘Streetview’ tools, remote sensing may not provide important information on the quality of the parcel, the number of units/offices, and vacancy, for instance. A very detailed land-use inventory may aim to capture information that is best surveyed in the field and which may provide a clearer indication of the overall character of a particular area. Allan B. Jacobs (1985) emphasizes the potential value of observation as a primary method of inquiry and analysis, noting:

Seeing people and their environments is quite different from learning about them second hand. Ideas of the poorness or wealth of a population suggested by income statistics may not be borne out by observations of those people in their daily lives. There is a great difference between reading age statistics and seeing people with gray or white hair, carrying packages, waiting for buses, using the laundromat... Seeing twelve boarded-up houses one after another in North Sacramento has a much more powerful impression than being told that there are quite a few vacant housing in an older area. Planners tend to be more careful in deciding on policies and actions when they associate real people's faces and images of real places with decisions.

While advances in technology have made it possible to gather land-use data with apparent efficiency and relatively little cost (notwithstanding initial, potentially expensive cash outlays for purchase of needed equipment), these advances may belie or exclude the observation of the variety of data that is best acquired *in situ*. Based on this premise, the objective of this research is to develop a method of land-use data collection that may be

used by planners in the field to most accurately and consistently collect a variety of desired data that is more efficient than traditional pen and paper methods. The following chapter sets the groundwork for the development of such a method through the consideration of key criteria in the design of a survey instrument, as well as hardware and software considerations.

CHAPTER 3

Key Criteria for a Land Use Survey Instrument: Design and Existing Technology

The method for land-use data collection developed in this research was successfully utilized for the General Plan studio course in the City and Regional Planning department at California Polytechnic State University, San Luis Obispo. It was successfully applied for two cities in the state of California; Newark, located in the San Francisco Bay Area, and Clear Lake, located in Lake County. While it was primarily developed in response to the immediate need to collect a large amount of land-use data within a short amount of time and with limited funding, it is assumed that these constraints are not significantly different for what may be present in professional planning practice. Similarly, it is assumed that desired specifications for hardware and software selection are transferable between academic and professional settings (i.e., customization, controls for data ensuring data quality).

However, potential alternatives to the choice and design of a viable survey instrument exist, some of which are presented here. Ultimately, the choice of hardware and software will depend on the data needs of a particular jurisdiction, the knowledge and resourcefulness of staff, and budgetary constraints. Depending on the setting, some of these criteria may be of little concern, or may need to be thoughtfully considered and balanced to determine the most feasible solution. In addition to cost, key criteria that should be evaluated in the beginning stages of a data collection and survey instrument

selection and design include data quality assurance, customization, interoperability, and ease of use. These criteria are described below.

Data Quality

Quality assurance and control of data is an important consideration, especially when dealing with large datasets. Data collection via pen-and-paper may be susceptible to illegible notation, potentially affecting efforts to achieve consistency and completeness of values when data is transferred into an electronic database. While a separate quality assurance phase can follow a collection period, an ideal survey instrument will somewhat negate this necessity by providing the surveyor with a set of pre-defined attribute values that are consistent across all survey instruments (in the case of multiple survey groups) and can be easily selected from a list or table rather than manually typed. This is especially pertinent for land-use data, where a particular land-use category will apply to thousands of parcels.

In the case of the General Plan studio course, our team sought to define land-uses on the parcel level by a primary (or general) category as well as more specific sub-categories, as is common to planning practice. For instance, a parcel with residential uses would have a general category of residential, but a specific land-use falling under the general residential category may be single-family, multi-family, duplex, triplex, or apartment. Sub-categories were conceived as to be selectable only when a particular corresponding general category was selected. This 'decision tree' approach was successfully implemented to provide a possible range of attribute values for the survey, while maintaining a high standard of data quality that required minimal subsequent clean-up to the collection phase.

Customization

The application of a decision-tree approach as well as consideration related to other attribute data types and data entry methods dictate that the selected hardware and software combination permit some degree of customization in survey design. Planners may wish to collect a range of nominal, ordinal, interval or ratio data for importation and symbolization in GIS. The ideal instrument will allow a full range of customization options for the collection and storage of these types of data, including the ability to define attribute names for easy identification, the ability to pre-define possible values as mentioned previously, or to establish safeguards for certain attributes that limit interaction but permit the display of already known values (in the case of a land-use survey, this may include street addresses, APN numbers or some other unique identifier for a particular parcel, or a field that displays the parcel acreage, for reference).

An additional consideration with respect to customization is the method and form of input. The design of the survey instrument should examine the limitations or features of the available hardware and their relationship to the software that will house the collected data. For instance, does the hardware feature expansion ports for connecting of devices that may be needed? If expansion is not necessary, are hardware input methods and specifics of the graphic user interface (GUI) changeable so as to permit efficient and intuitive data collection? A key aspect of this customizable data entry may include the creation of user-defined forms that appear when the surveyor is entering data. Depending on the software choice and time constraints, surveyors may need to use 'default' GUI/data

entry settings, but these may not provide for optimal functionality of data entry and display. In the case of a comprehensive, city-wide land-use survey, it may be worthwhile to invest additional time toward custom form design to facilitate efficient data collection that minimizes error.

Interoperability/compatibility

Another key criterion for selection, design, and use of survey equipment is its interoperability with a range of software that the planner will use for data manipulation, spatial analysis, and cartographic representation. The analyst should examine the format of the data files that are generated by the survey instrument, and confirm their compatibility with other hardware and software that will be used in analysis. If the survey instrument and related software does not provide output in a compatible format with a GIS, what options exist for converting the collected data?

Ease of Use

The ideal survey instrument should facilitate data collection. What constitutes "easy" may be subjective and vary from person to person. Thus, data managers should carefully evaluate the tradeoffs that exist between different survey instruments to determine which tools represent the most efficient use of technology given data needs, management capability, and the needs of surveyors when out in the field. Key aspects of ease-of-use may be addressed through customization options within the software, but may also need to consider survey instrument form factor. Instruments with expansive displays, for instance, may present a greater burden on the survey teams in terms of weight, increasing fatigue above what would be experienced with a smaller, more portable device.

Hardware Considerations

A number of hardware tools exist that can facilitate field collection of land use data. In 2003, The Association for GIS Professionals (URISA) published a white paper (NovaLIS Technologies, 2003) which highlighted the benefits of electronic field data collection for land assessors across a number of devices, including handheld computers, laptops, Personal Digital Assistants (PDAs), and Pen Computers. The hardware landscape has changed considerably since Waters' review. Laptops are now available in a variety of form factors, including 'ultra-portable' laptops which omit some features such as cd/dvd rom for decreased size, and tablet laptops which may be configured for touch-based interfacing but still include a keyboard for traditional two-hand data entry. PDAs, as popularized by the now defunct Palm Inc., have generally been supplanted by smartphones which provide more functionality within the same form factor. Finally, pen computers have evolved to include tablet devices, such as Apple's iPad, which are capable of processing user inputs without the use of an e-pen or stylus. Table 1, adapted from NovaLIS Technologies (2003), provides an up-to-date comparison of these hardware devices across typical specifications such as weight, screen size, and input methods. Figures 1, 2 and 3 illustrate examples of these device categories.

Table 1 - Characteristics of hardware devices for field data collection.

	Handheld Computers/Pocket PC	Laptops	Smartphones	Tablets
Weight	~ 2 lbs	2-4 lbs	<1 lb	1-3 lbs
Screen Size	Small screen	Full size screen	Small screen	Full screen
Input Method	Keypad	Keyboard (Tablet PCs may include touch screens)	Keypad or Virtual Keyboard	Virtual Keyboard, but may be compatible with wireless physical keyboards
One hand or two hand operation	One hand operation	Two hand operation	One hand operation	One hand operation
Examples	Motorola MC65	Lenovo Thinkpad	Apple iPhone, Samsung Galaxy	Apple iPad, Google Nexus

Source: Adapted from NovaLIS Technologies, Inc. (2003)



Figure 1- The Motorola MC65 Rugged Mobile.
Image from <http://www.motorola.com>.



Figure 2- The Lenovo Thinkpad.
Image from www.engadget.com



Figure 3- Apple's iPhone and iPad.

Software Considerations

The selection of software used in land use data collection is just as critical to the success of a land use inventory. Data managers should consider the needs of the agency in terms of the previously mentioned criteria to select the optimum software package. Because software features may be of primary importance over and above hardware selection, and because certain software packages may only operate on certain operating systems, the preference of software may determine the choice of compatible hardware, but this may not always be the case when certain hardware features such as durability, GPS, and wireless connectivity are needed.

There is a range of software packages that can be used to facilitate land use data collection. Examples include Pendragon Forms (Pendragon Software Corporation), Field Assets (LBS Wireless), and HanDBase (DDH Software). These software packages are supported across a variety of hardware platforms, and include several key features that are valuable for undertaking a land use inventory, including user-defined attributes, customized forms for data entry, and transferability of data to desktop database programs such as Microsoft Access. ESRI also provides a number of mobile applications under its ArcGIS line of products that work across a variety of Windows, iOS and Android devices, and which may represent the best choice in terms of seamlessly integrating field data into an ArcGIS desktop environment.

This chapter has presented various criteria and reviewed a sample of commercially available hardware and software that may be considered in the planning and implementation of an electronic land-use data collection effort. The ultimate choice of tools to be used for such an effort will ultimately depend on the data needs and goals of the survey and available resources in terms of staff knowledge, existing information systems, and money. The following chapter describes the survey instrument that was designed and successfully deployed for electronic data collection in the cities of Newark and Clearlake, California.

CHAPTER 4

Description of Survey Instrument Specifications and Features

Data Needs Assessment

This chapter describes the survey instrument that was successfully utilized for the collection of land use data in the California cities of Newark and Clearlake. The land use data collection effort for these cities was a portion of the overall scope of work that the Cal Poly City and Regional Planning department conducted to update the general plans of these cities. The land use inventory and corresponding General Plan update was conducted by a different team for each city, thus, some of the data that the Newark team was interested in collecting was not collected for the city of Clearlake. Table 2 presents the data that the Newark team collected for its land use inventory. Data types marked with an asterisk indicate data that was not collected by the Clearlake team.

Table 2 also illustrates the schema that was developed for collecting data on land uses at the parcel level. This schema was first designed by the Newark team, which represented the initial design and deployment of the survey instrument. Land Use types were divided into seven general land use categories: Residential, Commercial, Mixed-Use, Public Facilities, Open Space, Vacant, and Other. All of these general land use categories, with the exception of Vacant and Other, were further subdivided into specific land use categories. For instance, specific residential land uses include single-family detached, single-family attached, multi-family duplex, multi-family triplex, multi-family quad (for four residential units in a building), and apartment (five or more residential units among one or more structures). Additionally, the secondary and tertiary land uses

were to be applicable only when a mix of uses existed at the parcel. In this case, the surveyor would indicate the primary land use as a type of mixed-use, and then enter the secondary land-use, such as residential, which was evaluated to comprise the majority of the parcel. The tertiary land use would indicate the other land use type that existed but did not constitute the majority or dominant use of the parcel.

This specification represents the decision-tree characteristic referenced in the previous chapter, and was a primary consideration in the design of the survey instrument to ensure data quality. The author concluded that the survey instrument should allow the survey teams to select general and specific land uses from a list of predetermined values, with specific land-use designations only selectable when a corresponding primary land-use designation is selected. This method of data collection would ensure that data values would be consistent across different survey teams, because it would not be necessary to manually type in attribute values. This specification would eliminate the need to clean the data insofar as locating and correcting misspelled values for individual parcels within the land use inventory. However, it would not ensure that parcels were accurately surveyed. For instance, it would still be possible for a surveyor to incorrectly classify a parcel as ‘single family attached’ where a more correct classification would be ‘multi-family duplex’. Thus, it remains necessary for members of the data collection effort to mutually specify and reach consensus on the land use types so that accuracy of the data is maximized. Even with this shortcoming, the decision tree approach still results in increased consistency of attributes between parcel records within the database.

Table 2 - Data Needs for Newark Land-Use Inventory

	Data	Parent Values	Child Values
Parcel Data	Address	Any alphanumeric address at the parcel	N/A
	Street	Any alphanumeric street name	N/A
	Street Suffix	Ave, Road, Blvd, St, Way, Cir, Expy, Dr, Ln, Pl, Ct, Plz	N/A
	Sidewalk Condition	Good Fair Bad	N/A
	Primary Land Use Secondary Land Use Tertiary Land Use	Residential	Single Family Attached Single Family Detached Duplex Triplex Quadplex Apartment
		Commercial	Retail Service Office Industry
		Mixed-Use	Residential Commercial Commercial Mixed-Use Public Commercial
		Public Facilities	School Church Police Fire Waste Community Center Civic/Government ROW
		Open Space	Park Agriculture Conservation Streams/Drainage/Channels
		Vacant	N/A
		Other	N/A
Parcel Notes	Any alphanumeric notes	N/A	

Table 2 continued - Data Needs for Newark Land-Use Inventory

	Data	Parent Values	Child Values
Structure Data	Number of Stories	Numeric Value	N/A
	Number of Units	Numeric Value	N/A
	Number of Vacant Units	Numeric Value	N/A
	Business Name	Any alphanumeric entry	N/A
	Building Condition	Good Fair Bad Poor	N/A
	* Presence of Unit Conversion	Yes No Maybe	N/A
	* Paved Access to Street	Yes No	N/A
	* Sidewalk Access	Yes No	N/A
	Structure Notes	Any alphanumeric entry	N/A

The creation of pre-defined values was additionally specified for other parcel attributes. These attributes include Street Suffix, Sidewalk Condition, Structure Condition, Presence of Unit Conversion, Paved Access to Street, and Sidewalk Access. While these attributes were populated from a list of pre-determined values, they did not utilize a second-tier of classification as in the case of the specific land-use designations.

Hardware and Software Selection and Specifications

Having determined the primary criteria for the survey instrument, the author began to explore hardware and software options for possible use. In addition to facilitating pre-determined attribute values and the desired decision tree approach

(customization and data quality), other aspects relating to the criteria specified in the previous chapter were considered.

Because the survey instrument was conceptualized to eliminate the need to transcribe data collected via pen-and-paper into an electronic format, it was important that the instrument's database allow for interoperability with other desktop programs and file formats that were compatible with ArcGIS desktop, which was the ultimate destination of the field-collected data. The land-use data would be stored in a geodatabase that could be updated instantaneously via Microsoft Access. Thus, the survey instrument software needed to permit the export of collected data in a Microsoft Access compatible format.

Additionally, ease of use was another key criterion for both software and hardware. This is a somewhat subjective measure that is dependent upon the experience and comfort of the users of the software, including the data manager and the data collectors. In evaluating hardware and software options against technological trends, the author concluded that a touch-screen interface would represent the least cumbersome and most intuitive option for the survey teams. Such interfaces, which serve as the primary method of interfacing among tablet devices, benefit from the inclusion of context sensitive virtual keyboards which delimit character types based on the attribute type for which data is being collected. This has the added corollary benefit of a larger screen display when compared to handheld computer/pocket pc device, as the surface area of the screen is not compromised by the surface area that houses a tactile keyboard.

These criteria, as well as resource limitations pertaining to time and money informed the selection of Apple Inc.'s iPad device as the hardware component for the

survey instrument. The General Plan team procured four second-generation iPads which would each be shared by two field surveyors. One surveyor would observe and describe the existing conditions for the parcel, while the other surveyor would enter the corresponding data described by the observer.

Given the decision to use Apple's iPad, the choice of database software would be limited to software that was available in Apple's App Store. After sampling various database applications that would meet the specified criteria, HanDBase, published by DDH Software, was chosen as the preferred application to facilitate the land use inventory. HanDBase is a mobile relational database application that is available on various platforms, including iOS, Android, and Windows. Table 3 lists the specifications of the program. For additional information regarding software specifications, see the Bibliography for information regarding the HanDBase user's manual.

Table 3 - HandBase Features and Specifications

Feature	Specification
Maximum Number of Databases	Depends on hardware memory limitations
Maximum Number of Fields per Database	100
Maximum Number of Records per Database	65,000
Field Types	Text, Integer, Float, Pop-Up, Check-Box, Unique, Signature, Date, Time, Link, Linked, Note, Heading, DB Popup, Calculated, Relationship, Conditional, External
Maximum size of a Text or DB Popup Field	254
Minimum number in an Integer Field	-2,147,483,647
Maximum number in an Integer Field	2,147,483,647
Minimum number in a Float (depending on decimal place settings)	-2147483647.9999
Maximum number in a Float (depending on decimal place settings)	2147483647.9999
Maximum size of a note field	2000 characters
Maximum number of Popup Items per field	60
Maximum number of characters in a Popup Item description	80

Software Features

In addition to the specifications listed in Table 3, HandBase also includes a number of features that facilitate the field collection of land-use data. These features include the customizable data entry forms, the ability to merge data collected on different devices, and compatibility with .MDB and .CSV files (for importing and exporting, respectively).

Customized Forms

While HandBase does allow for instant data entry upon installation, users can create customized data entry forms to fit the task. Carefully designed forms may further assist with the speed and accuracy of the data collection process, and maybe present a

more intuitive method of entry to surveyors in the field. Figures 4 and 5 illustrate the customized forms that were created to collect land use and structure data for the City of Newark. Both forms consist of several elements used to capture data for individual parcels, including tabs to cycle between the land use form and the structure forms, text fields which can receive manual input or display information selected from a drop down menu, and decision tree boxes in which secondary (child) values are filtered based on the selection of a specific primary (parent) value. Elements can be freely arranged such that intuitive data entry is maximized while data entry time is minimized. After entering/selecting data values for a parcel and corresponding structures, the surveyor simply touches the onscreen prompts to either proceed to the next record or take the application back to the table view. Doing so will save the data collected for the parcel in the database.

iPad 8:40 PM 19%

Cancel OK

Address and land use Structure 1 details Structure 2 details Structure 3 details

MAP ID: 1413 APN: 92-119-35 Acreage: 0.15208195089...

Street address (number): 8278

Street name: Juniper

Street suffix: AVE Street_suffix

Sidewalk condition: G G F B

Survey Date: Date

No Date

Parcel notes:

Primary land use: Single family detach... Residential > Commercial > Mixed-Use > Public Facilities > Open Space > Misc > Edit Popups

Secondary land use: Residential > Commercial > Mixed-Use > Public Facilities > Open Space > Misc > Edit Popups

Tertiary land use: Residential > Commercial > Mixed-Use > Public Facilit... > Open Space > Misc > Edit Popups

Previous Next Tools

Figure 4 – Customized Land Use data collection form in HandBase for iPad.

iPad 9:19 PM Not Charging

Cancel OK

Address and land use Structure 1 details Structure 2 details Structure 3 details

Structure 1 notes:

Number of stories: **Number of units:**

Number of vacant units:

Business name(s):

Condition of building:

- G
- F
- P
- B

Presence of unit conversion:

- Y
- N
- Maybe

Paved access to street:

- Y
- N
- Edit Popups

Sidewalk access:

- Y
- N
- Edit Popups

Previous Next Tools

Figure 5- Customized Structure data collection form in HandDBase for iPad.

Data Merge

HanDBase also allows data collected across different devices to be merged into one database, which is useful when data collection is spread across more than one survey team and across multiple shifts. To do this, the application uses the hardware's wi-fi capability to connect devices that are connected to the same wireless network. The user may determine which device will receive the data from the other devices (this is the 'target' device). The user can activate 'Connect' mode within the application on the other devices one at a time, which will display a unique ip address that is then entered into the web browser of the target device. The browser on the target device will then display all of the databases currently saved on the device to which it is connected. The user simply needs to select the corresponding parcel database for the data to transfer and merge. If the wi-fi network to which the devices are connected does not facilitate 'Connect' mode via the browser window, or if a wi-fi network is otherwise unavailable, the devices may also be connected via 'ad hoc' mode to facilitate data transmission and merging.

Compatibility with Microsoft Access

While the mobile HanDBase application is capable of functioning independently for the design and distribution of a land use database, it is also capable of importing predesigned Microsoft Access databases through supplemental desktop software. This option is recommended as it may be more efficient for the user to design the database with a traditional desktop interface rather than via the iPad's touch based interface. After performing initial configurations in Microsoft Access, the database can be imported into the HanDBase desktop application for further refinements and conversion into the .PHB format that is specific to the mobile HanDBase application, at which point the database

can be downloaded to all of the survey devices. After all data is collected, the database can be exported as a .CSV file for importation into Microsoft Access and updating of the geodatabase via user-defined queries.

This chapter has described design consideration, specifications, and features of the survey instrument that was successfully deployed for land use data collection for the cities of Newark and Clearlake. Appendix A provides detailed instructions that the practitioner can use to replicate the electronic data collection method that is the focus of this report. The practitioner is additionally advised to consult the HandBase User's Guide for a more detailed reference concerning the features and use of the software. The following chapter outlines the analysis methodology that was applied to estimate the relative advantage of the electronic method of land use data collection compared to traditional pen-and-paper collection, and presents the findings of the analysis in terms of time savings.

CHAPTER 5

Analysis Method and Findings

The analysis methodology that was used to demonstrate the hypothesized benefits of electronic data collection over traditional pen-and-paper data collection consisted of a statistical difference of means test (t-test) between two sample data collection efforts (one electronic and one traditional) to illustrate the statistical significance of data collection time savings. Thus the null and experimental hypotheses evaluated at the 95% confidence level are:

H₀: There is no significant difference in data collection time between the two methods.

H₁: Data collection time via the electronic method is significantly different (smaller) than the traditional method.

To test the statistical significance of time savings, a total of 120 parcels were surveyed in San Luis Obispo, California so that 60 parcels (N=60), were surveyed by each method. Each sample group consisted of thirty residential parcels, and thirty commercial parcels. The residential and commercial parcels were selected so as to achieve similarity between the sample groups; All residential parcels surveyed were single family residential parcels, and all commercial parcels were located outside of San Luis Obispo's central business district and included on-site parking. For each parcel, various land use and structure data were obtained. Land use information included address, sidewalk condition, and Primary Specific Land Use. Structure data was collected for up to three structures for each parcel, and included the number of stories, the number

of units, the number of vacant units, the building condition, and for commercial parcels, business names.

The traditional pen-and-paper method utilized a tabular paper form with rows representing individual parcels and columns representing parcel attributes, which is a format similar to Microsoft Access database. It is assumed that for a citywide land use inventory, this database inspired form design would be preferable to a form design that would potentially dedicate a single page to each surveyed parcel and which could be more similar in design to the electronic forms (Figures 4 and 5 from the previous chapter).

It is expected that a highly designed paper form similar in design to the electronic form could potentially be used, and that it would provide data collection time savings similar to those of the electronic form. For instance, pre-defined values could be listed on a designed paper form, and the surveyor would simply need to circle the applicable data value for the parcel being surveyed. This method of data entry could be reasonably expected to take a similar amount of time to enter as with an electronic, touch-based interface. However, the data collectors/managers would still need to transfer the written records to the electronic database for digital representation and spatial analysis, which would take longer than if the information was collected electronically from the start. Thus, a designed paper form, while representing a potential advantage in data collection time over a tabular form, would still be inferior to electronic methods when the sums of data collection time and data transfer time are considered.

Based on the previously stated utility assumption of a tabular form, this research only considered the time savings of the electronic form in comparison to the pen-and-

paper tabular form. Further research is recommended to verify time savings between electronic form and designed paper form data collection techniques.

The conditions under which the survey was administered should also be noted. First, all 120 parcels were surveyed by a single person over four days, with each day consisting of a single approximate two-hour survey period. While it may have been possible to condense the surveying into two or possibly one single day, this method allowed the research to control for fatigue, assuming that a longer survey period would result in decreased efficiency. Additionally, the survey was conducted by one single individual who was familiar with the electronic data collection method, which may lead to concerns regarding the reproduction of the results when other individuals apply the method. While it is possible that the indicated results may not be representative of a “typical” survey effort, it is assumed that a sufficient level of training would achieve comparable results irrespective of age, education, and/or experience and familiarity with the survey instrument.

Testing Two Independent Sample Means

In order to test the difference in means between the two samples, the pooled estimate of the standard deviation of the sampling distribution of the difference in sample means was calculated (Healey, 2012). Equation 1 presents the pooled estimate formula used for small samples.

Equation 1: Pooled Estimate of the Standard Deviation of the Sampling Distribution of the Difference in Sample Means

$$\sigma_{\bar{x} - \bar{y}} = \sqrt{\frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2 - 2}} \sqrt{\frac{N_1 + N_2}{N_1 N_2}} \quad (1)$$

Where:

$\sigma_{\bar{x} - \bar{x}}$ is the standard deviation of the sampling distribution of the difference in sample means;

N is the number of cases in each sample;

s is the standard deviation for each sample;

1, 2 are subscripts for sample one and sample two respectively; and

The value for the pooled estimate was then taken as the denominator in Equation 2, in order to calculate the test statistic.

Equation 2: Test Statistic for Two, Small Sample Means

$$t \text{ (obtained)} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{\bar{x} - \bar{x}}} \quad (2)$$

Where:

\bar{x} is the sample mean, and;

$\sigma_{\bar{x} - \bar{x}}$ is the standard deviation of the sampling distribution of the difference in sample means;

Results

Table 4 displays the time in seconds taken to record data for the residential and commercial parcels via electronic and traditional methods, and corresponding descriptive statistics. The data indicates lower mean data collection times, variances, and standard deviations via electronic collection methods compared to pen-and-paper data collection. The differences in summary statistics between land use types (residential vs. commercial), is primarily due to the additional business name data component that was collected for each commercial parcel.

Table 4- Data Entry Times and summary statistics for residential and commercial parcels in San Luis Obispo, CA

Data point	Data Entry Time (Seconds)			
	Residential		Commercial	
	Electronic	Traditional	Electronic	Traditional
1	16	20	20	23
2	16	22	20	28
3	16	23	22	28
4	16	23	24	29
5	16	23	25	30
6	17	23	25	32
7	17	23	25	32
8	17	24	25	33
9	17	24	25	34
10	17	24	25	35
11	17	24	27	44
12	17	24	28	53
13	17	25	31	54
14	17	25	33	54
15	18	25	33	54
16	18	25	33	63
17	18	25	33	64
18	18	26	36	67
19	18	26	37	69
20	18	27	40	72
21	18	27	40	79
22	18	27	43	80
23	18	27	45	89
24	19	27	45	94
25	19	28	50	100
26	19	28	51	100
27	20	28	55	103
28	20	28	55	119
29	22	29	118	122
30	23	30	180	163
Mean	18	25	42	65
Median	18	25	33	58.5
Mode	17	23	25	54
Variance	2.8	5.3	1023.4	1193.4
Standard Deviation	1.7	2.3	32.0	34.5

Table 5 presents similar summary statistics when residential and commercial land use data collection times are combined for both data collection methods. Significance testing for the difference of means was similarly performed between the two methods separately for residential, then separately for commercial, and finally for combined land uses. The results of significance testing indicate that the difference in mean data collection time between the two methods is statistically significant for all three sample groups at the 95% confidence level (rejection of H_0). Table 6 summarizes the results of the difference of means tests. Appendix A includes detailed SPSS group statistics and independent sample tables for the three t-tests mentioned previously.

Table 5 - Summary Statistics for combined land use data by the two methods

Summary Statistic	Full Dataset Data Entry Time (seconds)	
	Electronic	Traditional
Mean	30	45
Median	21	28
Mode	17	23
Variance	647.6	987.2
Standard Deviation	25.4	31.4

Table 6 – Summary of Difference of Means Tests

	t (obtained)	t (critical) (2-tailed)	Degrees of Freedom	Sig. (2-tailed)
All Land Uses (Unequal Variances)	-2.941	1.98	113.118	.004
Residential (Unequal Variances)	-14.291	2.0	52.787	.000
Commercial (Equal Variances)	-2.707	2.0	58	.009

Finally, ArcGIS was used to estimate potential time savings of the electronic data collection method compared to the traditional method assuming the calculated differences in mean data collection times. Parcel and land-use spatial data was acquired from the City of San Luis Obispo. This data was spatially joined within ArcGIS to assign a specific land use to every parcel in the city. The joined data was examined to determine the city specified land uses that were surveyed. All surveyed residential parcels were determined to be ‘low density residential’, while the surveyed commercial parcels were either ‘neighborhood commercial’, ‘general retail’, ‘service and manufacturing,’ or ‘tourist commercial’. Table 7 presents the total number of residential and commercial parcels that fall under these land use categories and provides an estimate of potential time savings between the two survey methods by multiplying the mean data collection times by the total number of parcels with residential and commercial land-uses. Corresponding values for the 95% confidence interval are parenthetically displayed. The calculated estimates indicate that if every residential and commercial parcel specified above were to be surveyed, the electronic data collection method would produce time savings of 12.1 hours and 8.68 hours for residential and commercial parcels, respectively. It is expected that commensurate time savings would be realized across a variety of different land uses and in different urban places.

Table 7 – Potential estimated time savings for a sample of commercial and residential land uses in San Luis Obispo

Parcel Type	Number of parcels	Mean Time Electronic (Sec)	Mean Time Traditional (Sec)	Total Time Electronic (Hrs)	Total Time Traditional (Hrs)	Time Savings-Hours
Residential (single family only)	6,221	17.9 (17.3-18.5)	25.33 (24.5-26.2)	30.9 (29.9-32)	43.7 (42.3-45.3)	12.8 (12.4-13.3)
Commercial (neighborhood commercial, general retail, service and manufacturing, tourist commercial)	1,359	42 (30.2 – 53.1)	65 (52.5-77.3)	15.86 (11.4-20)	24.54 (19.8-29.2)	8.68 (8.4-9.2)

CHAPTER 6

Conclusion and Recommendations for Further Application

The collection and maintenance of land use data is an important component of planning information systems and the planning actions which they inform.

Comprehensive and up-to-date information on the location, intensity, and distribution of land uses throughout a community is of primary importance in the updating and development of General Plans, Climate Actions Plans, Specific Plans, and Housing inventories. While automated methods of data collection do exist and may be a useful starting point in undertaking a comprehensive inventory, such methods, such as the classification of parcels via remote sensing, may not accurately provide detailed information that is best observed in the field.

Given the comparative benefits of and potential necessity to collect land use information through direct observation, planners are faced with the task of evaluating and comparing data collection tools. The purpose of this research was to demonstrate the advantage of electronic based data collection compared to tradition pen-and-paper methods. In addition to time savings that were shown to be statistically significant compared to tradition data collection, electronic data collection provides additional benefits depending on the selection of hardware and software, including built-in data quality assurance through the use of decision tree configurations, and interoperability with geodatabases.

This research has illustrated the application of just one hardware and software combination, although numerous alternatives with different features and interfaces exist.

Data management professionals should carefully consider the tradeoffs between cost, data needs, survey instrument limitations, and ease of use for themselves as well as the data collection team. Readers interested in further application of the HandDBase software for land use data collection should see Appendix B of this document for a detailed procedure regarding survey instrument preparation with respect to the iPad version of the software. For additional information regarding HandDBase on other platforms and detailed user's guides, refer to DDHSoftware's website (www.ddhsoftware.com).

References

- Berke, P. R. (2006). *Urban Land Use Planning*.
- Dandekar, H. C. (2003). *The Planner's Use of Information*.
- Healey, J. F. (2012). *Statistics: A Tool for Social Research*, Ninth Editions, Wadsworth.
- Hu, S., & Wang, L. (2013, February). Automated urban land-use classification with remote sensing. *International Journal of Remote Sensing*, 34(3), 790-803.
- Jacobs, A. B. (1985). *Looking at Cities*. Cambridge: Harvard University Press.
- Kelly, E. (2010). *Community Planning* (2nd ed.). Washington D.C.: Island Press.
- NovaLIS Technologies, Inc. (2003). *Handheld Gadgets and the Appraiser: The Transformation of Field Data Collection*. (p. 9). Charlotte, NC: URISA.
- Wu, S.-S., Qiu, X., Uery, E. L., & Le, W. (2009). Using Geometrical, Textural, and Contextual. *Annals of the Association of American Geographers*, 76-98.
- Zhan, Q., Molenaar, M., & Gorte, B. (2000). URBAN LAND USE CLASSES WITH FUZZY MEMBERSHIP AND CLASSIFICATION. *International Archives of Photogrammetry and Remote Sensing*, 33, 1751-1760.

Appendices

A. SPSS output tables for t-tests.

B. Land Use Survey Instrument Configuration Guide

Appendix A

SPSS output tables for t-tests.

Table A-1: Dataset Table

		Method		Total
		Electronic	Manual	
Land Use	Residential	30	30	60
	Commercial	30	30	60
Total		60	60	120

Table A-2: Group Statistics for All Land Uses Test

Method		N	Mean	Std. Deviation	Std. Error Mean
Time	Electronic	60	29.77	25.448	3.285
	Manual	60	45.12	31.420	4.056

Table A-3: Independent Samples Test for All Land Uses

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Time	Equal variances assumed	8.389	.005	-2.941	118	.004	-15.350	5.220	-25.687	-5.013
	Equal variances not assumed			-2.941	113.118	.004	-15.350	5.220	-25.691	-5.009

Table A-4: Group Statistics for Residential Land Uses Test

Method		N	Mean	Std. Deviation	Std. Error Mean
Time	Electronic	30	17.90	1.668	.305
	Manual	30	25.33	2.309	.422

Table A-5: Independent Samples Test for Residential Land Uses

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Time	Equal variances assumed	5.605	.021	-14.291	58	.000	-7.433	.520	-8.474	-6.392
	Equal variances not assumed			-14.291	52.787	.000	-7.433	.520	-8.477	-6.390

Table A-6: Group Statistics for Commercial Land Uses Test

Method		N	Mean	Std. Deviation	Std. Error Mean
Time	Electronic	30	41.63	31.991	5.841
	Manual	30	64.90	34.546	6.307

Table A-7: Independent Samples Test for Commercial Land Uses

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Time	Equal variances assumed	2.577	.114	-2.707	58	.009	-23.267	8.596	-40.474	-6.060
	Equal variances not assumed			-2.707	57.661	.009	-23.267	8.596	-40.476	-6.057

Appendix B

Land Use Survey Instrument Configuration Guide