# ADDRESS POINTS AND A MASTER ADDRESS FILE:

# IMPROVING EFFICIENCY IN THE CITY OF CHINO

by

Michael Taylor Kellison

---

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

December 2012

# Table of Contents

# List of Tables



# List of Figures

## Abstract

One of the major responsibilities of a city government is management of real property, both public and private, within its jurisdiction. Classically, land is described by parcel (an areal geospatial feature) while structures are referenced by address (a pseudo-spatial text string). Handwritten, typed, and computerized address lists in spreadsheets and non-geospatial databases have been and continue to be used by the various departments within city governments. Inevitably, these lists are unevenly updated and inconsistent in various ways. Modern data management systems, specifically Microsoft Excel, contain tools for standardizing tabular data, including addresses. Geographic information systems (GIS), which can be used to manage parcel and address data directly, have traditionally relied upon street centerline or parcel geocoding to spatialize an address and determine its location. Utilizing Excel and geocoders together, to create a complete and reliable master address file (MAF), can help a city government operate more efficiently. Explicitly spatializing the relationship between addresses and parcels by converting textual addresses to address points (APs) in a GIS database, is critical for many aspects of city business operations, because doing so allows the points to be mapped. This thesis demonstrates that an accurate and complete set of APs is a superior solution to street centerline or parcel geocoding. APs can be created from a city government's multiple, internal spreadsheets and databases, utilizing Microsoft Excel and GIS in combination with street centerline and parcel geocoding, resulting in an MAF and APs that can be used citywide.

# Introduction

One of the major responsibilities of a city government is management of real property, both public and private, within its jurisdiction. Classically, land is described by parcel (an areal geospatial feature) while structures are referenced by address (a pseudo-spatial text string).

Commonly, as city governments grow and evolve, technical solutions are implemented at different times and with different goals, resulting in handwritten, typed, and computerized address lists in spreadsheets and non-geospatial databases throughout the various departments of the organization. These address lists are inevitably inconsistent, contain duplicative or conflicting information, store addresses in different formats, etc. As an example, an entire address string might be entered into a single field in one list and parsed into multiple fields in a different list.

Combining inconsistent data stored in different formats can be a complicated work effort. Modern data management systems, specifically Microsoft Excel, contain tools for standardizing tabular data, including addresses.

Geographic information systems (GIS), which can be used to manage parcel and address data directly, have traditionally relied upon street centerline or parcel geocoding to spatialize an address and determine its location. Utilizing Excel and geocoders together, to create an accurate and complete master address file (MAF), can help a city government operate more efficiently. Additionally, spatializing the relationship between addresses and parcels by converting textual addresses to address

points (APs) in a GIS database, is critical for many aspects of city business operations, because doing so allows the points to be mapped and the relationships to be viewed.

A case study of the City of Chino, California, where the author works as a Planner and de facto GIS Manager, demonstrates that an accurate and complete set of APs is a superior solution to street centerline or parcel geocoding. APs can be created from a city government's multiple, internal spreadsheets and databases, utilizing Microsoft Excel and GIS in combination with street centerline and parcel geocoding, resulting in an MAF and APs that can be used citywide.

## Motivation

Working the public counter for a city government can be a challenging job, as residents, brokers, and developers are constantly trying to determine what is and is not permitted per a city's municipal code. Between the phone calls and the continuous stream of people coming to the counter, there is often pressure to find information quickly. A reliable GIS can be an extremely valuable tool in these situations, as most questions deal with specific real property. Locating a subject property quickly helps to move the process along in an efficient manner. However, a reliable GIS is only as good as the data behind it and nothing is more frustrating, when working the public counter, than entering an address into a GIS and receiving the result "no matches found."

It is difficult to understand how a city government, whose job it is to oversee the properties within its city boundaries, does not have a reliable address list associated with its GIS. Unfortunately, this exact scenario has played out many times for me and others working the public counter for the City of Chino. The stop-gap solution to this problem is to rely on street centerline geocoding, but results from this method are only an approximate location. It takes additional steps of asking questions or looking up other records to determine exactly where the address is located. If this same scenario were to occur when a first responder is trying to locate an address, the consequences could be very serious.

A two-step process was developed to deal with this problem. The first step was to create an MAF from the multiple independent address lists within the City, which

involved reconciling the format differences and other substantial inconsistencies between the lists. The second step was to produce an explicit GIS layer, the APs. This step was more labor intensive, as field work was required to verify the locations of some of the address points. The effort was determined to be worth the work, however, because of the positive impact an MAF and APs would have on the accuracy and efficiency of future work efforts citywide. For counter staff and others, the benefit of entering an address, and not only receiving a result every time, but receiving a result that shows the exact location of the address, is compelling to any department using the City's GIS.

# Background

## Parcels (Lots)

A parcel, which is sometimes called a lot, is an area of land that is owned by single entity (individual, group of individuals, corporation, etc). Parcels are often created during the subdivision of land for the purposes of development, such as a residential subdivision for the development of single-family houses. Addresses are created to identify the location of a building constructed on a parcel. In the case of a single-family house, the relationship between the house and parcel is one-to-one, meaning there is only one address per parcel. In many other types of development, such as multi-family housing, commercial, and industrial development, there are sometimes multiple buildings or units constructed on a single parcel. In these situations there might be multiple addresses associated with a parcel. Therefore, there is a one-to-many relationship between parcels and addresses, with the possibility of multiple addresses per parcel. Figure 1 demonstrates this case, where multiple industrial buildings, each identified by a unique address, were developed on a single parcel.


**Figure 1 - Parcel/Address - One-to-Many Relationship**

**In this business park, five different buildings, each with their own address, sit on one parcel (outlined). The building at the southwest extends into a second parcel, but is identified by a different address within each parcel.**

## Street Addresses

A street address (hereafter simply an address) describes a physical location by reference to position along a street, road, or other transportation route. Addresses are generally associated with a building, such as a house, business office, or government office, but are sometimes created for park facilities, utility equipment, or even vacant lots needing to be described by a physical location.

As indicated on the website for the United States Postal Service, addresses have been used in the United States for at least 300 years, since the first postal service was established in the late 17th Century. The Post Office Department, predecessor of the United States Postal Service (USPS), was created in 1775. The USPS, as it is called today, was officially created in 1971. Addresses are often associated with the USPS, as people regularly use addresses when mailing packages and letters. Some people might think that the USPS establishes new addresses, but this is not the case. New addresses are established by the city or county having land use jurisdiction over development projects, and are generally assigned concurrently with development of a structure following a numbering system that has been sanctioned by the USPS. This numbering system includes patterns that make it easier to locate an address, such as having even numbered buildings on one side of a street and odd numbered buildings on the opposite side. Once an address is established by the city or county, the USPS is informed of the new address.

In addition to creating addresses, city and county governments use addresses in many aspects of their operations. Planning and building departments use addresses when reviewing and approving development related projects that require the issuance of building permits. Community services departments use addresses to send out information about community programs and events. Utility billing departments use addresses when tracking utility usage and for sending out bills. Police and fire departments use addresses when responding to service calls. Each department's need

for address-related information is different, and often a department will manage its own address list, which leads to inefficiencies in the form of duplicative efforts and inconsistencies between departments.

Address lists are critical to the operations of city government, because it is necessary to know where homes and businesses are located throughout the city. Creation of an MAF, which is an accurate and complete list of every address within the city, could resolve the problems of inconsistency and inefficiency. However, textual address lists on their own are limited, because they are only pseudo-spatial text strings that require additional spatializing to determine their location.

## Geocoding Methods

Geocoding offers one approach to spatializing street addresses. Traditional geocoding methods use street centerlines as the basis for determining the location of an address. This concept was first introduced with the TIGER (Topologically Integrated Geographic Encoding and Referencing) files developed by the U.S. Census Bureau (Klosterman and Lew 1992). Street centerline geocoding relies on a range of address numbers associated with a segment of street centerline data. For example, if a street centerline segment contains an address number range of 200-400, a geocoder will likely suggest that 300 is located at the middle point of that segment, which may not be correct (Figure 2).


**Figure 2 – Urban Street Centerline Geocoding**

**An urban example of a how a street centerline geocoder works at its basic level. Address 300 is positioned halfway between 200 and 400 along the street centerline.**

In most cases, street centerline geocoders are accurate enough for the casual user looking for the location of a building, especially in urban areas. They can usually even determine which side of the street the address is located, by considering whether the subject address number is odd or even. Imagine a different scenario, where the address is in a rural area and is off a dirt road. A geocoder that approximates the location of that address based on interpolation might mistake the position of the dirt road substantially and certainly would not be close to the actual structure (Figure 3).

Viramontes Express, a composting facility located in Chino, provides a good example of this situation.

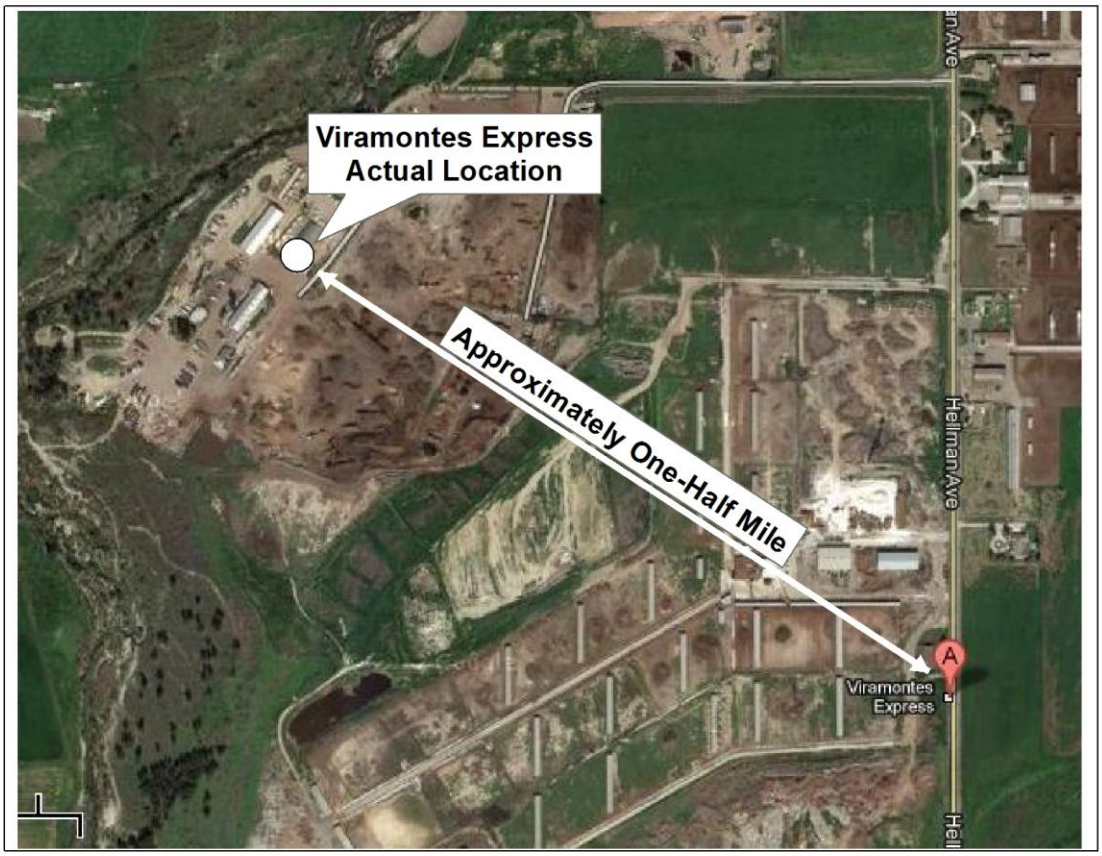


**Figure 3 - Rural Geocoding**

**Viramontes Express (17130 Hellman Avenue, Corona, CA) is about one-half mile away, as the crow flies, and almost one mile away in driving distance, from the geocoded location.**

This level of error could be quite problematic for some services, especially emergency services that need to locate addresses quickly.

Additionally, street centerline geocoding has limited ability to identify a specific suite/unit in a shopping center or apartment complex, a problem Goldberg (2010) calls sub-parcel geocoding (Figure 4).

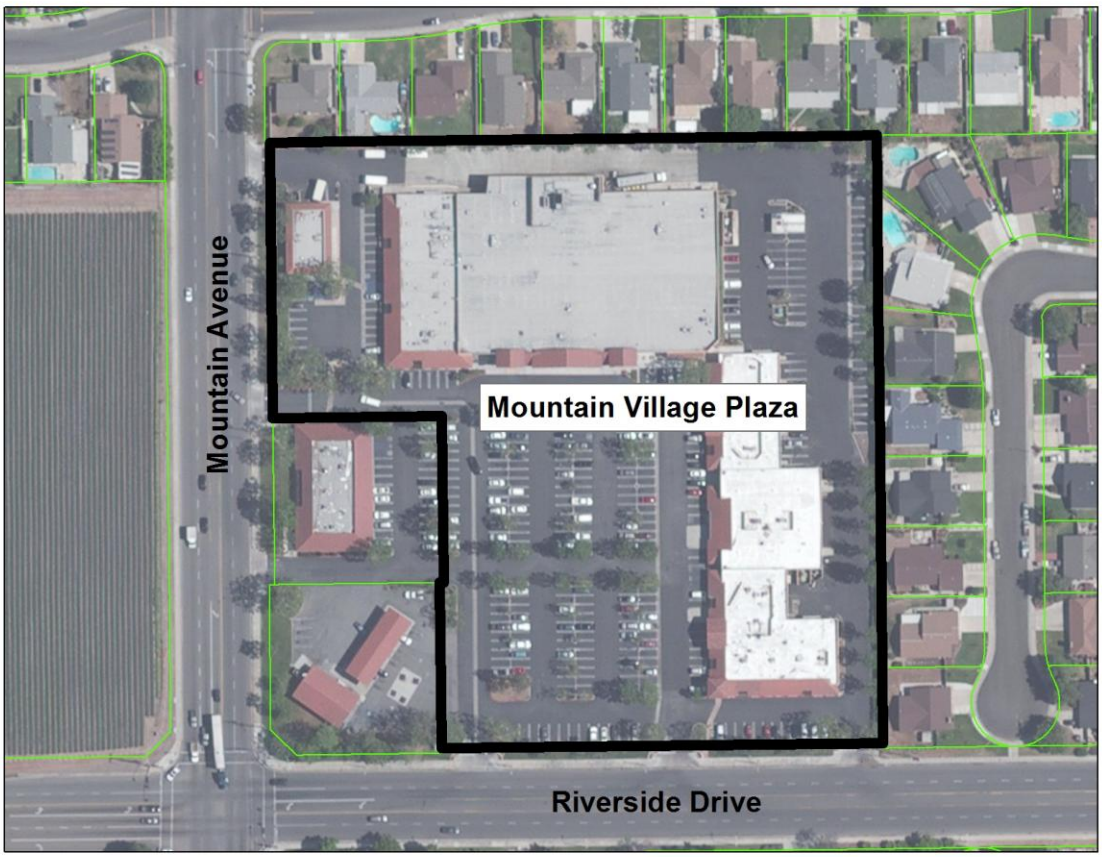


**Figure 4 - Sub-Parcel Geocoding**

**Mountain Village Plaza sits on a single parcel. However, each business within the plaza has a unique address.**

"This case occurs when multiple structures are residing on the same land parcel such as in apartment/condominium-type properties and large campuses such as universities and business parks or in the case of large farms where a single small structure may be located somewhere within a much larger parcel" (Goldberg 2010, 40). Goldberg goes on to suggest that future technological advances, such as aerial imagery recognition software, or secondary data sets with additional details might help to address these specific problems.

Many city governments have dealt with similar geocoding problems and have found solutions in the form of APs. An AP is a point placed on a map, generally at the center of a parcel or on top of a specific building, with its purpose being to show the location of an address. Many of the publications available regarding the creation of APs focus on county level efforts (Castaneda an Knippel, 2011; DeMeritt, 2009; Pima County, 2012; Zhou, 2008), perhaps because of problems encountered using traditional geocoding methods in rural areas, which are more common within county jurisdictions.

Hinton et al. (2009, p. 42) note "Strip malls, apartments, townhomes, rural structures and poorly addressed areas presented challenges and delayed the delivery of mission-critical services. This problem has been resolved through using GIS data points [APs] located at the precise location of the address." By creating APs and locating them at the precise location of the structure in which the address is assigned (Figure 5), emergency services know exactly where they are heading and do not have to rely on the approximation of a street centerline geocoder.

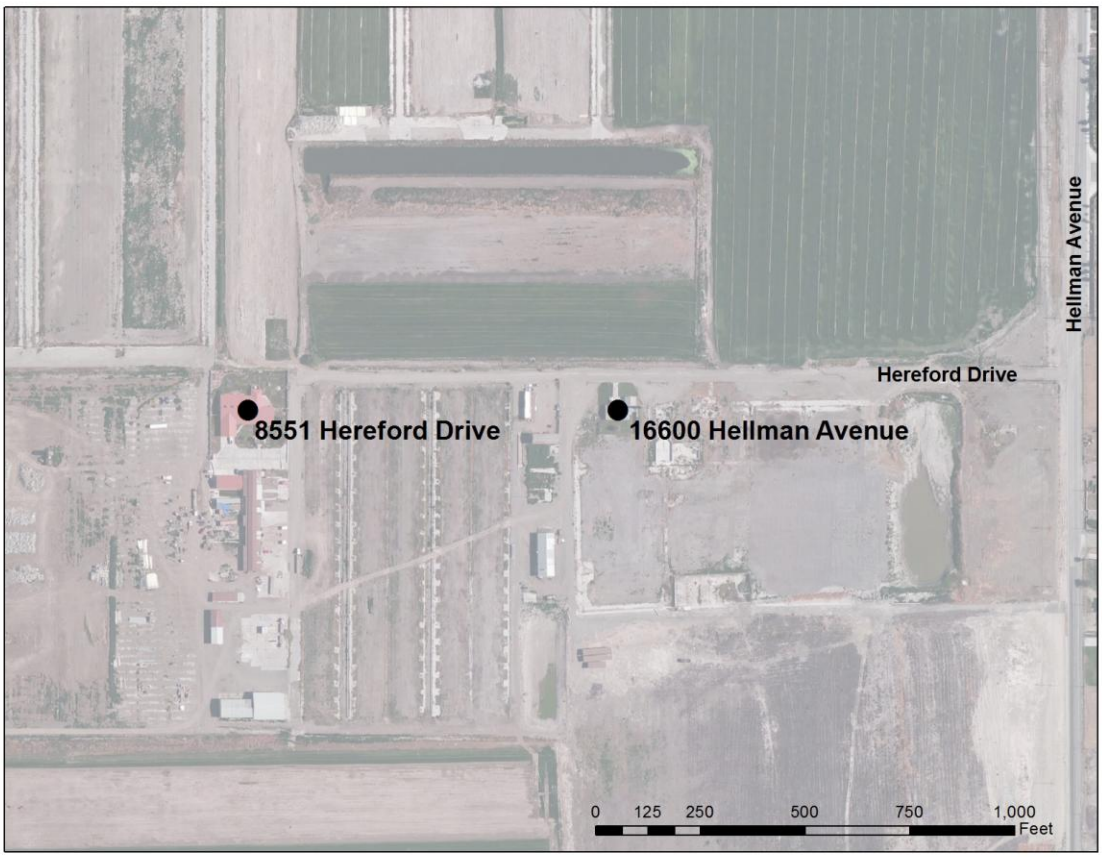


**Figure 5 - APs (Rural)**

**16600 Hellman Avenue is accessed from Hereford Drive. A street-centerline geocoder would locate the address 1000 feet to the east along Hellman Avenue.**

Zandbergen (2008) compared parcel, street centerline, and address point geocoding techniques. He found that parcel geocoding, which searches the address field of a parcel dataset, resulted in the lowest match rate, most likely due to parcel datasets containing only one address field, even though a parcel might have multiple addresses within its boundaries. Street centerline geocoding resulted in the highest match rate, but also produced false positives (geocoded non-existent addresses). Zandenberg concluded “Address points appear very promising as an address data model

for geocoding…they provide an extra validation of the address input data, since it is less likely a false positive will be introduced" (p. 231) (Figure 6).

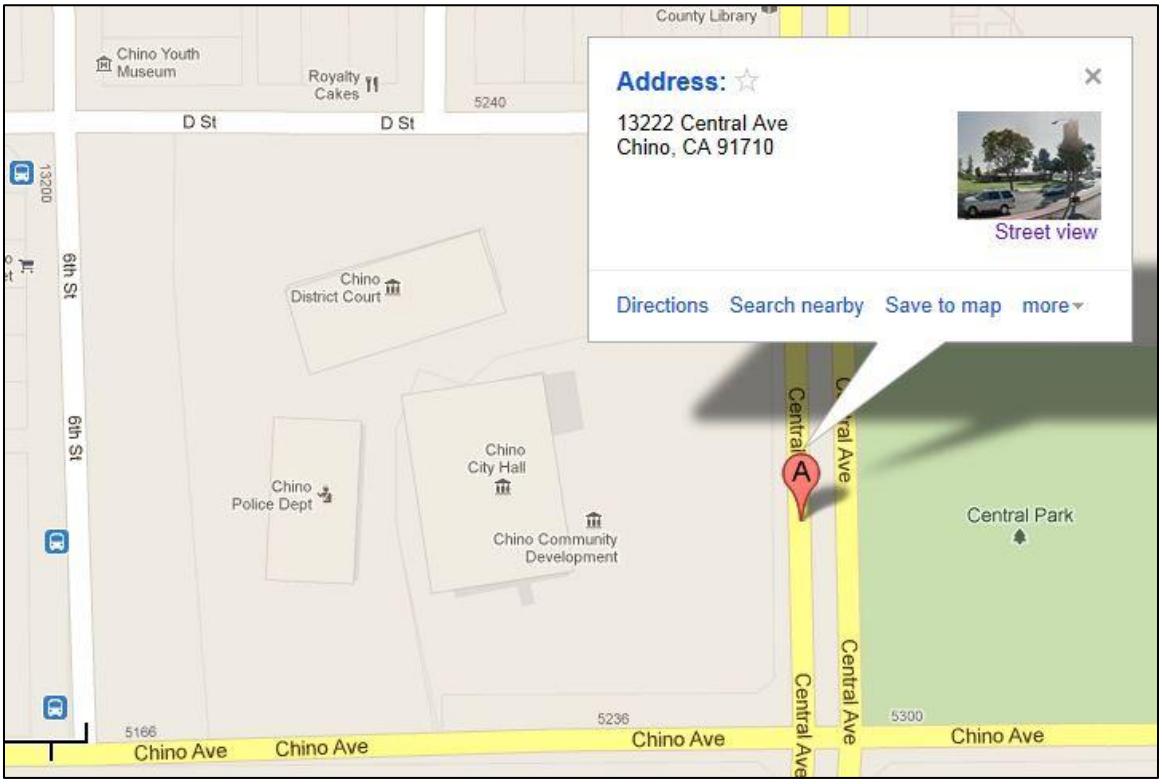


**Figure 6 - False Positive Result**

**13222 Central Avenue, Chino, CA, in Google Maps. However, that address does not exist, so the result shown is a false-positive.**

## Esri – Address Data Management

Environmental Systems Research Institute (Esri), a private company based in Redlands, California, is one of the global leaders in the GIS industry. They have developed the ArcGIS suite of software products that are used by public and private organizations worldwide. Esri's ArcGIS for Local Government team develops tools that assist in responsibilities commonly associated with local government. According to their

website, their Address Data Management tool, "can be used to maintain road centerlines with address ranges, facilities, site addresses [APs], and related mailing addresses…to streamline the collection, maintenance, and use of authoritative address information" (Esri).[1] The existence of this tool demonstrates the importance of maintaining accurate address data and utilizes both street centerline geocoding and APs. In practice, when a user of the tool clicks on a street centerline near the new (or existing) building, an AP is created by "reverse geocoding" from the street centerline address range. The user is free to edit the address and/or reposition the point to improve its accuracy.

## Other Agencies

In an article published in the Spring of 2011, Dakota County, Minnesota discussed their need for a countywide database of APs. Their immediate need for APs was to support a new computer aided dispatch (CAD) system to be used by their emergency services personnel, the Dakota Communications Center. They relied on both commercial data and address lists from cities within their jurisdiction to piece together a comprehensive list of addresses within the County and used street centerline geocoding to determine approximate locations of addresses not represented in their parcel layer. For those addresses that could not be geocoded, they worked with the appropriate cities to determine if the address contained errors or was a false record (Castaneda and Knippel, 2011).

[1] http://resources.arcgis.com/en/communities/local-government/01n40000002z000000.htm.

An extreme example is described by DeMeritt (2009) for Boone County, Kentucky, which consolidated address lists from 30 different agencies as part of a consortium aimed to standardize APs countywide. According to DeMeritt, they "…consolidated information from many different agencies into a single usable dataset for its clients without custom coding or manual processes by using the ArcGIS Data Interoperability Extension" (p. 17). A takeaway from the Boone County project was to provide database table fields for each component of a complete text address to maximize flexibility for end users. For example, they created a single field for the standardized text address (address number, street name, street type, etc.) we well as separate fields for each one of the address components, so when used within GIS applications, labels for addresses could be more detailed (complete text address), or less detailed (address component). Separating the address components into individual fields facilitates use of the list on future projects, as it is easier to recompose individual components as needed than it is to extract components from a complete text address.

Other notable agencies that have created APs include the District of Columbia and Great Britain. According to The District of Columbia's website, they created an APs dataset for their Master Address Repository (MAR), which is a database of streets, buildings, and points of interest deserving of an address. Great Britain's national mapping agency, Ordnance Survey, first created their APs dataset (ADDRESS-POINT) in the early 1990's. As indicated on their website, ADDRESS-POINT utilizes the more than 27 million addresses contained in Britain's Royal Mail Postcode Address File (PAF). Both

the District of Columbia and Great Britain make this data available for purchase by the general public.

## Case Study

Described below is a case study involving the creation of an MAF and APs for the City of Chino, located in San Bernardino County, California. In the following text, the City of Chino may be referred to as either City or Chino and the County of San Bernardino may be referred to as either County or San Bernardino.

Several years ago, the City purchased Esri ArcGIS Desktop and Server software, and a GIS manager was hired to develop a citywide GIS. The system was launched internally, giving City staff the ability to view property information about specific parcels. Unfortunately, with the turn in the economy, the GIS manager position was eliminated, data stopped being updated, and the system became outdated and unreliable.

The foundation of the City's GIS was, and is, parcel and address data (Parcel Layer) maintained and provided by the County. The County provides an updated Parcel Layer, in the form of a shapefile, to cities within its jurisdiction every two weeks. While the parcel polygons and parcel numbers are generally reliable, the address data contained in the Parcel Layer is often inaccurate or incomplete. As already noted, while the Parcel Layer only allows for a simple address to be associated with each parcel, it completely ignores the existence of multiple addresses located within a parcel.

The City also utilizes a number of packaged software systems for various department functions, and each system contains its own address list. One of these systems is used by the Planning and Building divisions for the tracking of development

applications and permits (the Project Database). At the time of this case study, the address list in the Project Database had not been updated in several years. Chino's Finance Department, which includes the business licensing and utility billing divisions, utilizes another system (hereafter Finance Database). The address list contained in this system is much more current and reliable than that of the Project Database, but is still known to contain errors.

The need to consolidate the City's multiple address lists was obvious. While there were other address lists being used in the City, this case study focused on the consolidation of what are arguably the three most important address lists – the Parcel Layer, the Project Database, and the Finance Database – into a shared MAF.

## Address Components

An address format had to be established that worked well with ArcGIS for the creation of APs, and could also be leveraged by the Project Database, Finance Database, and other address lists to be consolidated in the future. In Southern California, the minimal components of an address are address number (house number), street name, city, state, and zip code. A variety of additional components, such as prefix direction (north, south, east, west), street type (road, avenue, street, etc.), suffix direction (NW, SW, NE, SE), building designators, and suite/unit designators, may also appear. There is no single correct format, and every jurisdiction needs to decide what is appropriate for them. Ultimately, however, all addresses must be acceptable to USPS.

For Chino, the format of the MAF was driven by the format of the address lists within the Parcel Layer, Project Database, and Finance Database. The Parcel Layer parsed addresses into six components, the Project Database parsed addresses into four components, and the Finance Database parsed addresses into six components. Because it is easier to recompose individual components as needed than it is to extract components from a complete text address, it was decided to parse the addresses into individual components in the new MAF. However, following the example of Boone County, additional fields of component combinations, including a field for a complete text address, were also included in the MAF. These combinations were created using formulas within Excel to avoid potential errors associated with retyping the addresses. The following tables are excerpts showing the component fields of the three subject address lists:

**Table 1 - Parcel Layer**

**The Parcel Layer parsed addresses into six components; as of July 2012 it had 22,863 records containing a parcel number, with 20,813 of those containing a valid address.**

| NUMBER | PREDIR | STREETNAME | STREETREETTYPE | CITY | STATE |
|---|---|---|---|---|---|
| 8919 | S | MERRILL | AVE | CHINO | CA |
| 9467 | | MERRILL | AVE | CHINO | CA |
| 14741 | S | CARPENTER | AVE | ONTARIO | CA |
| 9032 | | MERRILL | AVE | ONTARIO | CA |
| 9031 | | EUCALYPTUS | AVE | ONTARIO | CA |
| 8911 | | EUCALYPTUS | AVE | ONTARIO | CA |
| 4365 | | WILSON | ST | CHINO | CA |
| 4355 | | WILSON | | CHINO | CA |
| 4345 | | WILSON | | CHINO | CA |
| 4335 | | WILSON | ST | CHINO | CA |
| 4323 | | WILSON | ST | CHINO | CA |
| 4301 | | WILSON | ST | CHINO | CA |
| 12760 | | WRIGHT | AVE | CHINO | CA |
| 12759 | | RAMONA | AVE | CHINO | CA |
| 12746 | | WITHERSPOON | RD | CHINO | CA |

**Table 2 - Project Database**

**The Project Database parsed addresses into four components; as of July 2012 it had 21,299 address records.**

| CITY_ID | STREET_NAME | STREET_DIRECTION | STREET_NO |
|---|---|---|---|
| CHINO | PHILADELPHIA ST | | 931 |
| CHINO | PARCEL MAP | | 1056 |
| CHINO | SIGN ST | | 1234 |
| CHINO | RIGHT OF WAY ST | | 1234 |
| CHINO | SONOMA CT | | 1235 |
| CHINO | RAMONA AVE | | 1275 |
| CHINO | MILLS AVE | | 2220 |
| CHINO | RIVERSIDE DR | | 3220 |
| CHINO | RIVERSIDE DR | | 3242 |
| CHINO | RIVERSIDE DR | | 3258 |
| CHINO | RIVERSIDE DR | | 3340 |
| CHINO | CHINO AVE | | 3413 |

**Table 3 - Finance Database**

**The Finance Database parsed addresses into six components; as of July 2012 it had 20,641 address records.**

| CITY | NUMBER | STATE | STREET_TYPE | STREET | ZIP_CODE |
|---|---|---|---|---|---|
| CHIN | 5808 | CA | CT | HORSESHOE | 91710 |
| CHIN | 13425 | CA | AVE | MOUNTAIN | 91710 |
| CHIN | 13459 | CA | AVE | MOUNTAIN | 91710 |
| CHIN | 12951 | CA | AVE | BENSON | 91710 |
| CHIN | 13302 | CA | PL | BARCELONA | 91710 |
| CHIN | 12951 | CA | AVE | BENSON | 91710 |
| CHIN | 12951 | CA | AVE | BENSON | 91710 |
| CHIN | 12951 | CA | AVE | BENSON | 91710 |
| CHIN | 13374 | CA | PL | BARCELONA | 91710 |
| CHIN | 12951 | CA | AVE | BENSON | 91710 |
| CHIN | 12951 | CA | AVE | BENSON | 91710 |
| CHIN | 12951 | CA | AVE | BENSON | 91710 |

As an example of the inconsistencies discovered between and within the three address lists, Table 4 shows the different terms used for street type within each address list.

**Table 4 - Street Type Terms**

**The column on the left shows the full name of the term, while the other columns show how the term, or abbreviation for the term, was used in the different address lists. The Project Database and Finance Database used both fully spelled-out term names and abbreviations.**

| TERM | PARCEL LAYER | PROJECT DATABASE | | FINANCE DATABASE | |
|---|---|---|---|---|---|
| AVENUE | AVE | AVE | AVENUE | AVE | AVENUE |
| BOULEVARD | BLVD | BLVD | | BLVD | |
| CIRCLE | CIR | CIR | | CIR | |
| COURT | CT | CT | COURT | CT | |
| DRIVE | DR | DR | | DR | |
| LANE | LN | LN | LANE | LN | |
| LOOP | LP | | LOOP | | |
| PARKWAY | PKWY | PKWY | | PKWY | |
| PLACE | PL | PL | | PL | |
| ROAD | RD | RD | | RD | |
| STREET | ST | ST | STREET | ST | |
| TERRACE | TER | | | | |
| TRAIL | TR | | TRAIL | | |
| WAY | WY | | WAY | WY | WAY |

For the new MAF, the full term name was used to eliminate confusion about what the term means. Abbreviations being used in the three address lists had to be converted to the full term name. This was an easy exercise using the *find and replace* tool in Microsoft Excel. The address components adopted for the Chino MAF are shown in Table 5.

**Table 5 - MAF Components**

**Address components adopted the Chino MAF**

| Component | Description |
|---|---|
| NUMBER | Address number |
| PREDIR | Prefix direction for the address |
| STREETNAME | Street name for the address |
| STREETTYPE | Street type for the address |
| BUILDING | Building designator |
| SUITE_UNIT | Suite or unit designator |
| CITY | City in which the address is located |
| STATE | State in which the address is located |
| ZIP | Zip Code in which the address is located |
| NAMETYPE | Street Name and Street Type combined |
| PROPADDR | Number, Predir, Street Name, Street Type, Building, and Suite/Unit combined |
| FULLADDR | All individual address components combined (Number, Predir, Street Name, Street Type, Building, Suite/Unit, City, State, and Zip) |

Additional components were included beyond what existed in the three address lists, including building designator and suite/unit designator, thus avoiding the problem noted by Zandbergen (2008) that APs are often only created one per address number, leaving suites/units to be determined by other means. For Chino, each building and suite/unit was treated as a separate address in the MAF so that an AP could be created for each. This required additional work in the way of site visits and collection of directory maps for apartment complexes, campuses, and business parks (Figure 7).


**Figure 7 - Directory Map**

**An example of a directory map obtained to determine different addresses within a parcel, including suite/unit designators.**

## Creating the MAF and APs

With address components decided, the next step in developing the MAF was creating a draft set of APs. Point features were obtained from the Parcel Layer using the Feature to Point (Data Management) tool in ArcGIS. This process resulted in a single point feature at the centroid of each polygon (parcel) feature. The existing attributes from the parcels were automatically added by ArcGIS to the point features, including

address information that was associated with the parcel. Many undeveloped parcels did not have addresses associated with them, so those point features were deleted.

The draft set of APs contained quite a few records that were incomplete. 447 records were missing a street type, 1,589 records contained no zip code, and another 365 contained zip codes that were incorrect. The zip codes were easily corrected in ArcGIS using a spatial selection within a zip code polygon, and then assigning the proper zip codes to the selected addresses (Figure 8).

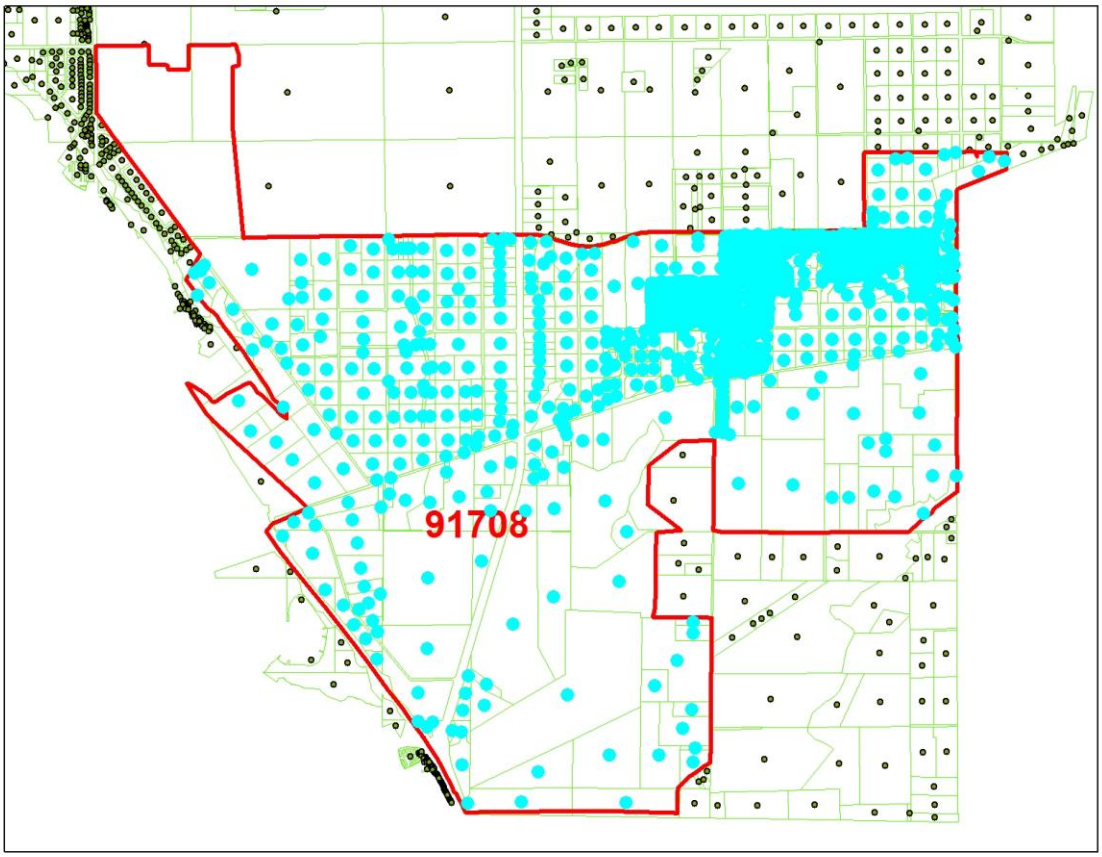


**Figure 8 - Zip Code Assignment**

**Zip code polygons were used to select the APs within them (here for the 91708 zip code) and the zip field reassigned accordingly.**

The street types required a bit of research, consisting mostly of comparing the incomplete records to complete records to determine the appropriate street type. The Parcel Layer data, including the parcel number field, was extracted from ArcGIS into a Microsoft Excel spreadsheet. The Project Database and Finance Database were also added to the spreadsheet and parsed into individual components consistent with the Parcel Layer. For each row, the parsed data was recombined into an additional column to create a complete property address (see Table 5) that functioned as a unique identifier. Excel's Remove Duplicates tool was then used to eliminate duplicates, with priority on removing duplicates without a parcel number. This step was critical, because the majority of records were duplicated amongst the address lists being combined. Once this step was completed, 23,288 addresses remained, making up the final MAF.

## Geocoding

After combining the address lists, there remained 2,475 addresses not associated with a parcel. Geocoding was used to determine where the APs for these addresses should be located. A new table was created, containing only the addresses without a parcel association, and this table was geocoded in ArcMap using the standard options available in version 10.0 US Streets Geocode Service (ArcGIS Online). Approximately 2,000 of the 2,475 records were matched and added to the map as APs (Figure 9). The geocoded APs were generally located along the correct side of the street, near the parcel in which they were to be placed. Using Chino's aerial imagery from 2011, the geocoded points were manually moved from the street to a location on

top of the building. The APs created from the Parcel Layer were also moved at this time to their ultimate location on top of the building.


**Figure 9 - Combined APs from Parcel Layer and Geocoding**

**The APs symbolized in green came directly from the Parcel Layer; those in red, representing additional property addresses within a parcel, came from geocoding, with manual adjustments as needed.**

Most of the remaining unmatched addresses were erroneous records that needed to be eliminated from the dataset. A few geocoded addresses required special attention and/or modifications to the data, as they included address numbers with too many/few numbers, street names misspelled, and new streets not yet included in Esri's geocoding service.

Individual suites/units were still largely unrepresented at this point, so APs were created using directory maps, such as the one shown in Figure 7, for suites/units. For multi-unit complexes for which maps were not obtained, or for addresses that were difficult to locate using aerial imagery, it was necessary to perform site visits for final confirmation. Figure 10 shows the workflow for the creation of Chino's APs.


**Figure 10 – Workflow**

**The workflow for creating Chino's APs, including step numbers, descriptions, and estimated man hours spent to complete each step.**

# Results

## Address Points

The "problem examples" displayed in Figures 2-6 were resolved effortlessly by using APs. Another specific example of where APs proved to be an improvement over parcel geocoding and street centerline geocoding was at the Chino civic center. Figure 11 shows the parcel boundaries and address information contained within the Parcel Layer. Three of the four parcel polygons in this area did not contain address information. This void of information was a common occurrence within the Parcel Layer in situations where multiple adjacent parcels were owned by the same entity. One parcel was generally populated with relevant address information, and the remaining parcels did not contain any address information. Additionally, two of the buildings in the civic center fell across parcel lines, confusing matters even more. Relying on parcel geocoding in this situation simply did not work, as the Parcel Layer did not accurately align with the buildings onsite and did not properly reflect addresses assigned to those buildings.

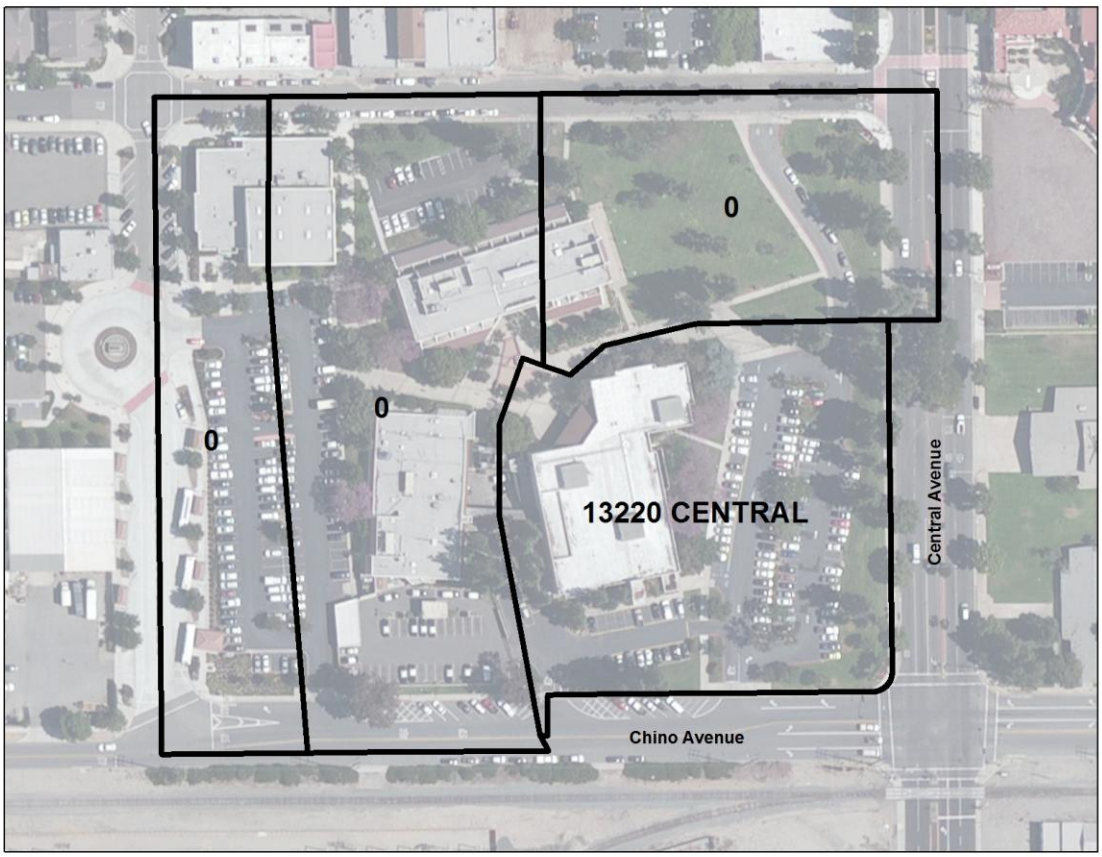


**Figure 11 - Civic Center - Parcel Geocoding**

**Many parcels within the Parcel Layer did not contain address information. This commonly occurred when a property was vacant or the parcel was part of a collection of parcels owned by one entity.**

Due to the campus setting of the Civic Center, street centerline geocoding also fell short of being able to identify which building on the campus was actually the building of interest. In Figure 12, street centerline geocoding identified a point along Central Avenue in an attempt to locate 13250 Central Avenue. However, the geocoded point was actually placed closer to 13260 Central Avenue, thus providing very confusing results to someone unfamiliar with the Civic Center.

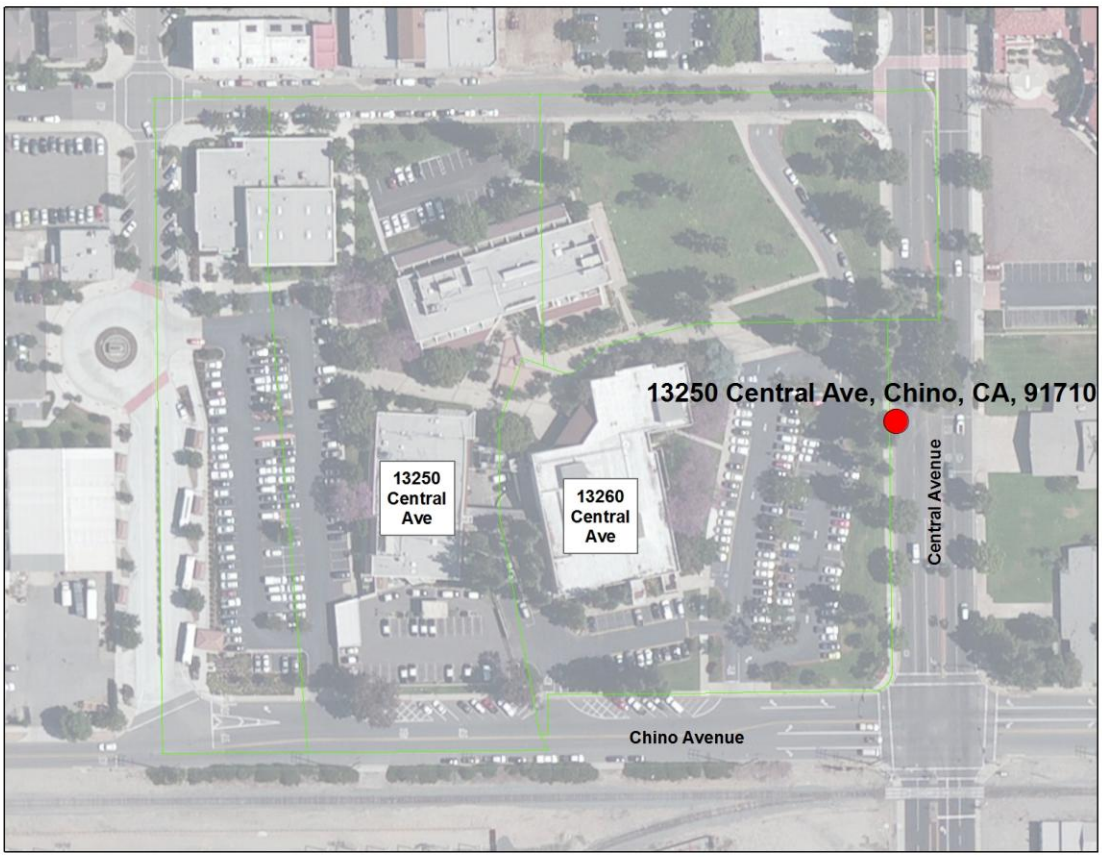


**Figure 12 - Civic Center - Street Centerline Geocoding**

**13250 Central Avenue was geocoded along Central Avenue, closer to 13260 Central Avenue, in this campus setting.**

APs resulted in the most accurate representation of addresses at the Civic Center. Figure 13 shows the APs, labeled with the address number, street name, and street type, thus clearly identifying which address belonged to which building. The address information was correct and the location of each address was clear.

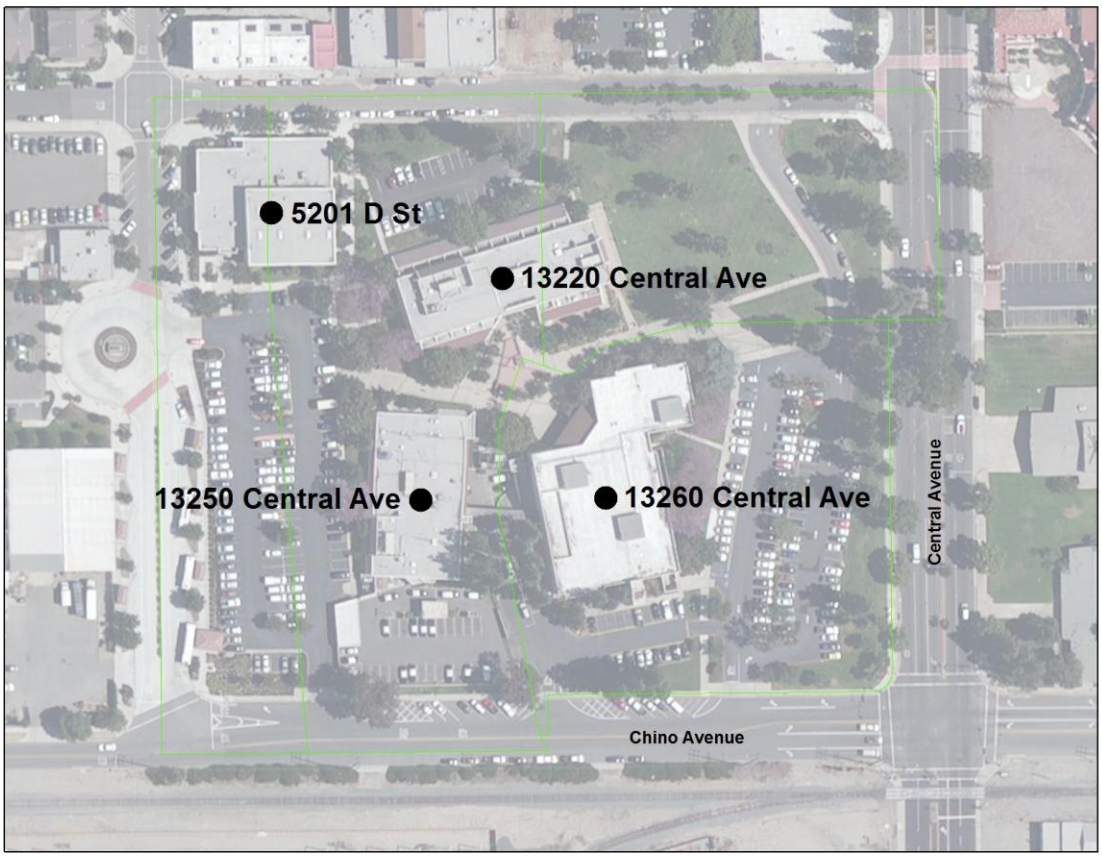


**Figure 13 - Chino Civic Center - APs**

**Major buildings of the Chino Civic Center, correctly identified by APs.**

## Master Address File

Toward the end of this case study, my department received an inquiry from the Chino Finance Department for an address list that might aid them in verifying the accuracy and completeness of the Finance Database. They explained that the Finance staff has attempted to regularly update the Finance Database, and they have struggled to find a reliable source of addresses. They also mentioned that the Police Department regularly contacts them to verify addresses within the City, because Police staff members are also trying to keep their systems updated. This meeting confirmed the

timeliness and significance of this thesis project, but also made it obvious that the MAF needed to be made available to staff citywide.

## Chino Intranet Site

After consultation with the Information Technology Manager, it was decided that the best way to make the MAF, and APs, available to staff citywide was through the City's Intranet site. This site is used to extend a variety of information and services to staff, including a link to the City's Municipal Code, help desk services, a phone directory, and more. Figure 14 shows the addition of links for both the MAF and APs under the heading of GIS Links in a test environment for the City's Intranet site. Figure 15 shows the GIS viewer configured with APs. Administrative approvals from City management were still needed at the time of this study before the links could be made active on the City's live Intranet site.

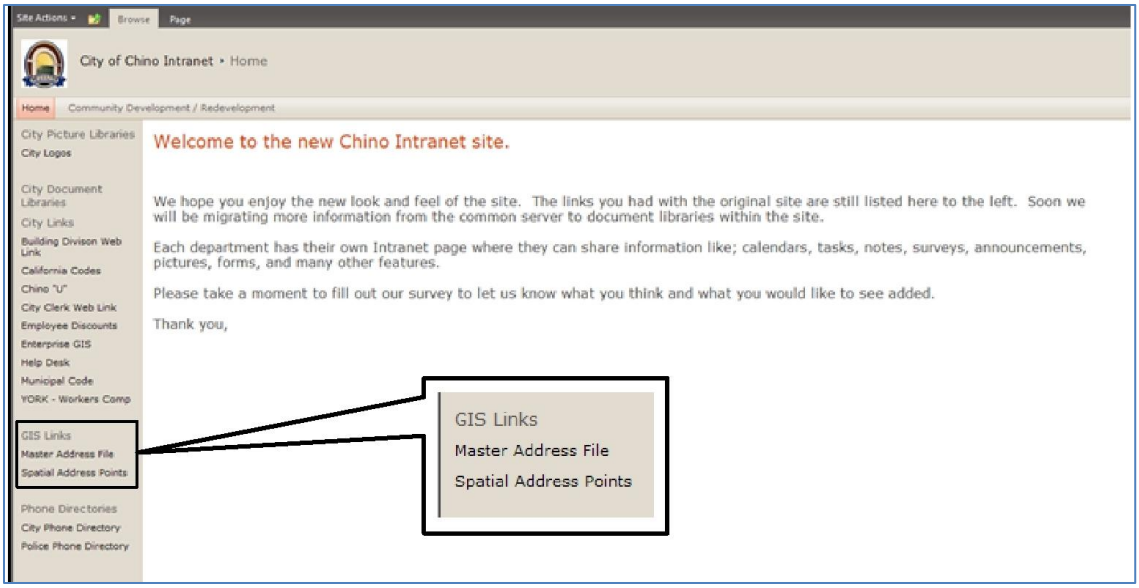


**Figure 14 - Chino Intranet Site**

**The Chino Intranet Site is available to all City staff. The MAF is a downloadable Excel file while the APs open in a simple GIS viewer.**

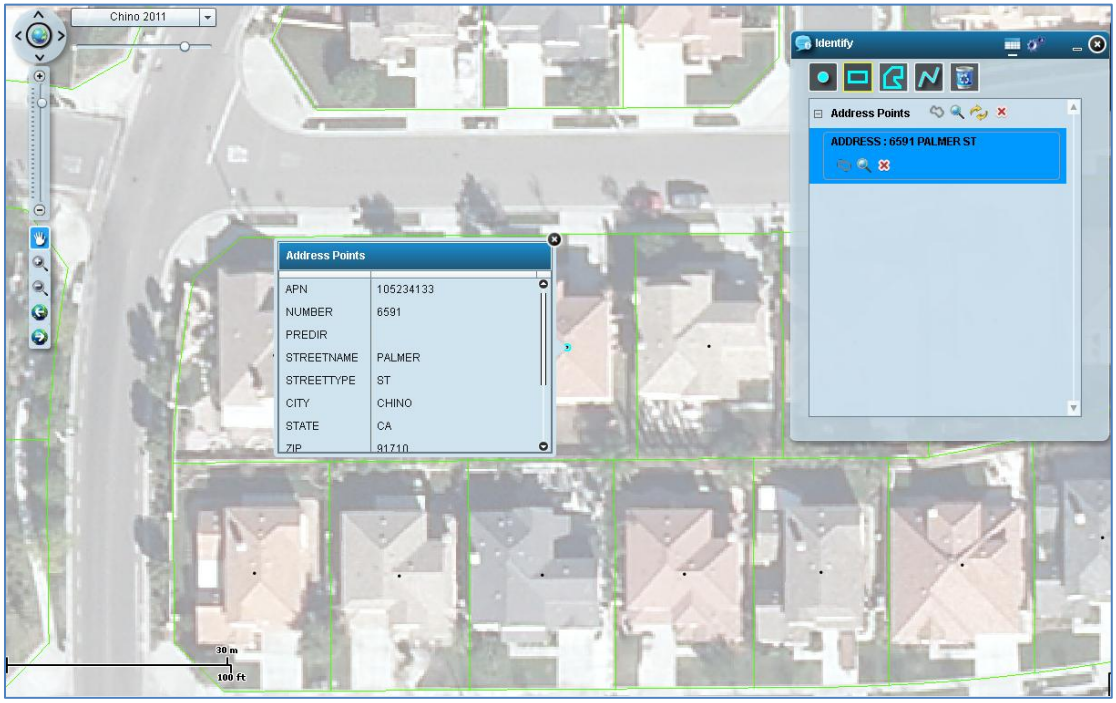


**Figure 15 – APs in a GIS Viewer**

**The GIS viewer allows City staff to search and view APs overlaid on aerial imagery.**

# Discussion

## Maintenance – MAF and APs

It is critical that the MAF and APs be maintained as addresses are created, amended, or retired. In Chino, the GIS Manager is the administrator of the MAF and APs, but addresses are managed by the Building Division. When address changes occur, the Building Division forwards the information to the GIS Manager. APs are coded with a status of either current or retired, thus maintaining spatial records of existing addresses and those that no longer exist. APs are easily re-associated with parcels through a spatial join, thereby automatically updating ownership information for each AP with every Parcel Layer update. A new MAF is created with every address change by exporting the APs table to a Microsoft Excel spreadsheet. Following this procedure, as opposed to manually updating the MAF, ensures consistency between the MAF and APs.

## Importance

The creation of an MAF and APs increases the accuracy and efficiency of work produced within the City; it also reduces the number of errors resulting from incomplete address lists and uncertain address locations. Identifying addresses, specific buildings, and even suites within buildings can be accomplished with a reliable MAF and APs, eliminating guesswork that might otherwise lead to costly errors. Staff time will no longer be wasted on updating multiple address lists throughout the City. Employees working the public counter will not encounter false-negatives; instead, they will know

with confidence when an address truly does not exist, and more importantly, where addresses do exist. Emergency services can be confident that they have the most current and precise location of an address when responding to calls, whether it is a single-family house or a suite within a business park.

This thesis demonstrates that an accurate and complete set of APs is a superior solution to street centerline or parcel geocoding. APs can be created from a city government's multiple, internal spreadsheets and databases, utilizing Microsoft Excel and GIS in combination with street centerline and parcel geocoding, resulting in an MAF and APs that can be accessed citywide through its Intranet site.

## Next Steps

This thesis focuses on construction and maintenance of an MAF for the City of Chino. To facilitate its use, as mentioned earlier, the MAF is accessible through the City's Intranet site as a web-hosted GIS service and also as a downloadable Excel file. However, to integrate the MAF with other production systems in the City, selected contents from it need to be re-updated into those systems.

For example, an immediate next step is to re-update the Project Database from the MAF, in parallel with the Parcel Layer update. The Parcel Layer, which the County updates every two weeks, can be used to update two tables in the Project Database (Figure 16): the Parcels table, which contains a list of all the parcel numbers in the City, and the People table, which contains name and address information about the owners of those parcels. Parcels are commonly subdivided and sold, so this update would enable the City to continually keep the parcel numbers and ownership information in the Project Database, now four years out of date, current. One possible mechanism for doing this update is SQL Server Integration Services, as discussed in the Appendix.

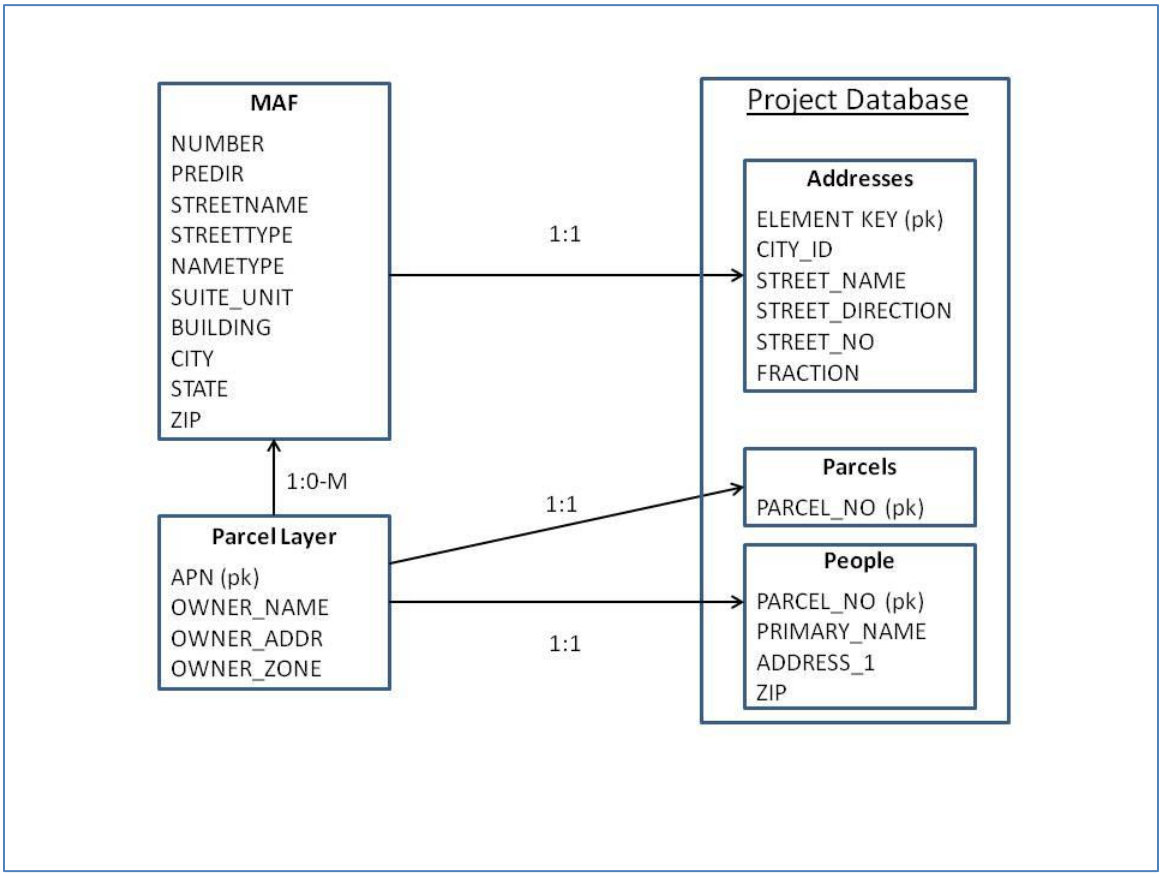


**Figure 16 – Data Flow Diagram**

**The MAF and Parcel Layer can be used to update several tables within the Project Database.**

The MAF, after being spatially joined to the Parcel Layer, will update the Addresses table in the Project Database. With these three tables regularly updated, the Planning and Building divisions will again be able to make use of the GIS functions associated with the Project Database, specifically referencing permits and project numbers to addresses.

# Bibliography


Babinski, Greg. 2009. GIS as an Enterprise Municipal System. *Government Finance Review 25(1):* 22-30.

Castaneda, Dan, and Randy Knippel. 2011. Desktop GIS: Building a County-Wide Address Points Database. Office of Geographic Informatin Systems. Dakota County. http://www.co.dakota.mn.us/Departments/GIS/Newsletter/Spring2011_Desktop_AddressPoints.htm. (Accessed July 12 - August 22, 2012).

City of Torrington. 2008. Building an Address Point Database Using an Out-Of-The-Box ArcGIS Server Website. http://www.ct.gov/gis/lib/gis/Torrington_Address_Points.pdf. (Accessed July 7, 2012).

DeMeritt, Matthew. 2009. A New Model for Address Point Management: Data interoperability improves customer service in Kentucky county. *ArcUser* (Winter): 16-19.

Devogele, Thomas, Christine Parent, and Stefano Spaccapietra. 1998. On spatial database integration. *International Journal of Geographical Information Science* 12(4): 335-352.

Dueker, Kenneth J. 1974. Urban Geocoding. *Annals of the Association of Amercian Geographers* 64(2): 318-325.

Environmental Systems Research Institute. 2003. An Esri Technical Paper. *An ArcGIS Address Data Model for the City of Calgary.* Redlands, CA.

Environmental Systems Research Institute. 2011. An Esri White Paper. *Integrating Esri Production Mapping with ArcGIS for Land Records.* Redlands, CA.

Environmental Systems Research Institute. Maintain Accurate Address Data. http://resources.arcgis.com/en/communities/local-government/01n40000002z000000.htm. (Accessed July 29, 2012).

Goldberg, Daniel. 2011. Advances in Geocoding Research and Practice. *Transactions in GIS 15(6)*: 727-733.

Goldberg, Daniel. 2010. Spatial Approaches to Reducing Error in Geocoded Data. PhD Dissertation, University of Southern California. ProQuest LLC.

Hinton, Curt, David Holdstock, and Crystal Phillips-Mustain. 2009. GIS for Public Safety: Are You Providing a Full Suite of Tools to Protect Your Community? *Emergency Number Professional Magazine* (August): 41-44.

Huebner, Francis C. 1906. Our Postal System. *Records of the Columbia Historical Society* 9: 126–74.

Jung, Christian, Daniel Karch, Sebastian Knopp, Dennis Luxen, and Peter Sanders. 2011. Efficient Error-Correcting Geocoding. *Karlsruhe Institute of Technology*: 1-15.

Klosterman, Richard, and Alan Lew. 1992. TIGER products for planning. *Journal of the American Planning Association* 58 (3) (Summer): 379-385.

Landrum, Rodney. 2007. SQL Server Integration Services: Gather information automatically. *SQL Server Magazine* (May): 29-34.

Ordnance Survey. Address-Point. http://www.ordnancesurvey.co.uk/oswebsite/products/address-point/index.html. (Accessed August 2, 2012).

Pima County Geographic Information Systems. Situs (Property) Address. http://gis.pima.gov/whatis/aboutsitus.htm. (Accessed July 6, 2012).

Sweet, Bob. 2003. Address Standards Guide For City of Oklahoma City Addresses. http://gis.okc.gov/website/documents/Training/Procedures/OklahomaCityGISStreetAddressStandards.pdf. (Accessed June 27, 2012).

The District of Columbia. Office of the Chief Technology Officer. Master Address Repository. http://octo.dc.gov/DC/OCTO/Maps+and+Apps/Online+Mapping/All+Online+Maps/Master+Address+Repository. (Accessed August 15, 2012).

Trainor, Timothy. 2003. U.S. Census Bureau Geographic Support: a response to changing technology and improved data. *Cartography and Geographic Information Science* 30 (2): 217-223.

United States Postal Service. The history of the United States Postal Service. An American History. 1775-2006. http://about.usps.com/publications/pub100/pub100_001.htm. (Accessed August 12, 2012).

Walker, Bruce. n.d. Saving Money: Creating an Address Point Layer "In House." http://proceedings.esri.com/library/userconf/serug10/papers/pap/saving_money-creating_an_address_point_layer_in_house.pdf. (Accessed July 12, 2012).

Wikipedia. Topologically Integrated Geographic Encoding and Referencing. http://en.wikipedia.org/w/index.php?title=Topologically_Integrated_Geographic_Encoding_and_Referencing. (Accessed August 4, 2012).

Wikipedia. United States Postal Service. http://en.wikipedia.org/w/index.php?title=United_States_Postal_Service&oldid=505187702. (Accessed August 5, 2012).

Yang, Duck-Hye, Lucy Mackey Bilaver, Oscar Hayes, and Robert Goerge. 2004. Improving Geocoding Practices: Evaluation of Geocoding Tools. *Journal of Medical Systems* 28(4) (August): 361-370.

Zandbergen, Paul A. 2008. A comparison of address point, parcel and street geocoding techniques. *Computers, Environment and Urban Systems* 32: 214-232.

Zhou, Tong. 2008. *Creation of a Master Address Point Layer - A Collaborative Approach to a Complex Project.* Paper presented at the ESRI International User Conference.

# Appendix

## SQL Server Integration Services

SQL Server Integration Services (SSIS) is a system that exists within Microsoft SQL Server software, which provides a framework to construct extract, transform, and load (ETL) processes. This type of data management software is exactly what is needed to maintain addresses within the Project Database by moving updated data from the MAF into the Project Database.

The process of creating an ETL process begins in SQL Server Business Intelligence Development Studio (BIDS), which is a specially designed version of Visual Studio that allows a user to create a process by dragging the different steps from a toolbox into the data flow window. Each part of the process is then configured, so it knows where the data is coming from (source), what it should do with the data (transformation), and where to put the data (destination). Figure 17 shows how data can be merged/joined from the MAF and Project Database, then inserted back into the Project Database, thereby updating address information within the Project Database.

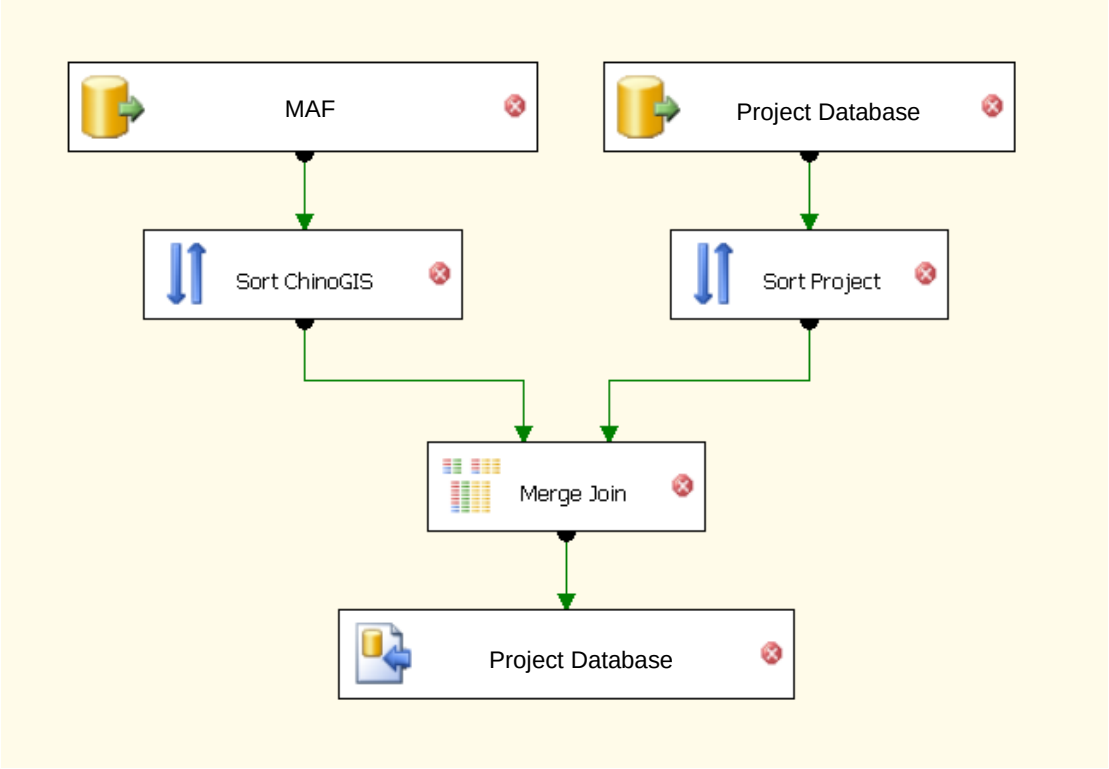


**Figure 17 – ETL Process**

**Data from multiple source files can be joined and loaded into a destination file.**