

The Development of a Master Address Repository for Washington, DC

Authors: David Jackson & Vivek Srivastav

Abstract

A comprehensive address database has a broad range of uses in the efficient delivery of government services. Constructing a Master Address Repository (MAR) for the city of Washington, D.C., is a time-consuming, complex, and challenging task. The repository now has more than 136,000 unique, valid addresses assembled and researched from more than 20 unique sources. Along with building the repository, comprehensive methods and testing procedures are utilized to check and maintain the MAR's integrity and accuracy. The MAR has more than just a full address. It contains various critical geographies such as Ward and Police District, the X,Y, coordinates, as well as attribution that links back to the District's Street Spatial Database (SSD)

I. Introduction

An address point contains important information about a particular location. For local government, addresses represent the most significant locator for databases, services and record records. An address record contains information including the street number, street name, street type, quadrant, x coordinates, y coordinates, full address, ward, ANC, zipcode, and property information.

The District of Columbia's, Master Address Repository (MAR), contains over 135,000 active address records, over 4,500 alias records that are linked to the address records and over 1,000 retired address records in the database. A retired address record is one where an address existed at a particular location but no longer exists due to demolition or reconfiguration of the building.

“The MAR allows the District government to more easily compare information across databases and agencies. The format and quality of address information varies across government systems, making it nearly impossible for government to identify all the activity associated with a given address. Now, addressing standards are built into the MAR. These standards define the valid components of DC addresses (AID, Number, Street Name, Street Type, and Quadrant), including their correct formatting and spelling. (DC GIS Website)”

A. History

Prior to the MAR, DC GIS had a less comprehensive address database used in the DC Guide (<http://about.dc.gov/index.asp>). It was a file derived primarily from the property points published by the [Office of Tax and Revenue](#). Moreover, each agency had their own address list which rarely matched others. There was no single address inventory for the District.

In July 2005, the MAR was published containing more than 128,500 address records. The MAR was compiled from 16 different government databases which contained addresses.

- DC Office of Tax and Revenue (OTR)
- Water and Sewer Authority (WASA)
- Metropolitan Police Department (MPD) –ADDRESS
- Department of Motor Vehicle (DMV) - Drivers License
- DMV - Registered Vehicles
- Dunn & Bradstreet Business Directory
- Mayor’s Call Center (MCC)
- MPD – STREET
- OTR - Property Owners
- OTR - Property Inventory
- Board of Elections (BOE) - Registered Voters
- Department of Health (DOH)
- Emergency –911 (E-911)
- DC Department of Transportation (DDOT) - Street Centerline
- Department of Consumer and Regulatory Affairs (DCRA)

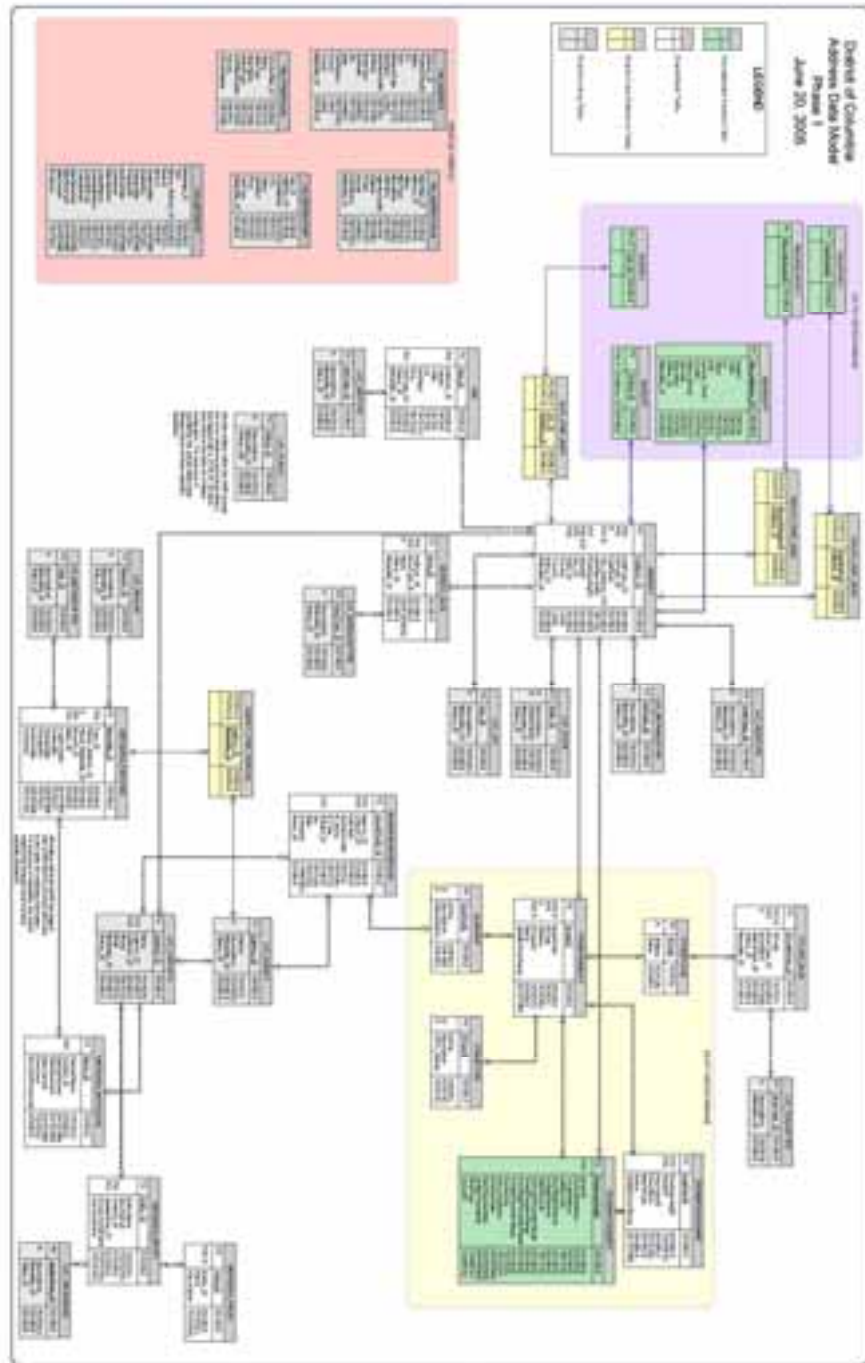
Since July 2006, an alias and address team consisting primarily of two GIS Analysts has added over 7000 address records, and over 4000 alias names have been added. Furthermore, over 2000 addresses have been updated (X,Y position moved, wrong address number, incorrect street name or street type).

II. MAR Database Design

The MAR contains not only address records but many other tables that contain valuable information. The MAR is stored in Oracle 9.2.07. The database is linked via a unique street identifier to DDOT’s Street Spatial Database. In fact addresses points cannot be added to the database unless they have a streetname value from the DDOT database.

Alias Records

Not all address records contain full address such as the Washington Monument, which does not have a street number and thus no full address. The street number and full address fields will thus have a null value. However, the ‘Washington Monument’ is an alias name for a record representing that location. The Washington Monument will be linked to an address record via a unique address identification number (AID). An address alias table which stores place names for locations in DC is linked to the address records. There are more than 4000 alias records in the MAR database. The alias record table contains important government offices, memorials, statues, hospitals, schools, police stations and fire stations.



The Master Address Repository Database Model

Each address in the database has been assigned a unique Address Identification Number (AID), map coordinates, and other useful information, including:

- Property square
- Lot number
- Ward
- Advisory Neighborhood Commission (ANC)
- Police Service Area (PSA)
- Zipcode
- Neighborhood Cluster

III. Address Verification Process

Most of the time spent on the MAR involves address verification. During Address verification, a decision is made about individual address records. In the address verification process an address record can be added, updated, retired, deleted or linked to an alias names. Address verification consists of:

- a. Gathering Addresses
- b. Checking Individual Addresses
- c. Digitizing Addresses

A. Gathering Addresses

Addresses are gathered from a wide range of sources. However, these sources can be categorized into three categories.

- i. Addresses Submitted by Users
- ii. Comparing Imagery, Building Polygons to Addresses
- iii. Database Matching

i. Addresses Submitted by Users

Increasingly, the MAR is being integrated into DC Government processes. Inevitably, some users are finding addresses that they consider valid but are not in the MAR. DC government has created a web site where users can verify and report missing addresses to the MAR team. If the submitted address is along a valid theoretical street range and is not found in the MAR, the user can then click on a web link and be forwarded to a site where they can submit the address. The user enters their contact information, along with the address in question is recorded in a MAR table which contains addresses to be verified. The information is then emailed to the MAR team as a ticket. The MAR team investigates the questionable address and responds to the user in a timely manner.

ii. Comparing Imagery and Building Polygons to Addresses

Addresses also are compared against the imagery and building polygons. Many building polygons do not have an address because they represent a garage, or sheds. Using the building polygons and more current raster imagery, the whole city was reviewed to look for buildings that did not contain an address. These locations of these buildings were

marked and later addresses were found..

iii. Database Matching

The MAR team acquires databases from a wide range of DC Government agencies. The database matching technique yields the most missing or faulty addresses. Here are some examples of databases that have been used to improve the MAR:

- Building Permits (Department of Consumer & Regulatory Affairs)
- Voter Database (Board of Elections and Ethics)
- Basic Business License (Department of Consumer & Regulatory Affairs)
- Student Database (District of Columbia Public Schools)
- DC Government Employees (Office of Personnel)
- Answers Please (Department of Human Services)

The addresses in the databases are automatically scrubbed, a process that ensures that the addresses are changed to meet the MAR format. The scrubbing process uses an ArcObjects program.

Many of the addresses that are contained in DC Government files do not meet the MAR format due to the following reasons:

- **Misspelled Street Names:** ‘CONNECTICUT AVENUE’ instead of ‘CONNECTICUT AVENUE’
- **Abbreviations:** ‘CT AVENUE’ the state postal abbreviation instead of ‘CONNECTICUT AVENUE’ or ‘AVE’ instead of ‘AVENUE’
- **Capitalization:** ‘Connecticut’ instead of ‘CONNECTICUT’
- **Incorrect Quadrant:** Quadrants are the unique address scenario for DC. There are four quads in Washington DC..... NE, NW, SE and SW. Even a complete address may be wrong due to Quad.

Once the addresses have been scrubbed, a simple attribute join is performed to see which addresses match. All addresses that do not match existing MAR addresses are then exported out and are ready for manual verification. These are potential missing addresses. These missing candidate addresses are then processed through an interpolation program that places the addresses between or on the side of existing MAR addresses.

B. Checking Individual Addresses

Many sources are checked during the address verification process. The more sources that recognize the candidate address, the more likely it will be added to the database. Below are some sources that are being used for address verification:

- United States Postal Service (USPS) Data
- Orthophotography Imagery
- Pictometry Imagery

- Street Centerlines
- Existing Addresses
- Internet Search
- Field Research
- Other DC Government databases (ex. DC Employees, Voter Rolls, Student database)

Based on these sources we determine if an address is valid and can thus be digitized. Sometimes, it is determined that the address cannot be solved in the office environment and thus field research is required.

However, it is preferable to solve address issues in the office environment rather than physically visiting the site. Field research consumes much time, but it is sometimes necessary as part of the address verification process.

ii. Field Research

On the plus side, field research, offers a pleasant break from the working in the office. Field verification occurs on a weekly basis. The city has been divided into 15 MAR Zones which allows field research sites to be grouped together in a more efficient manner. Each week field research sites are printed out and a route is planned. For downtown sites walking is used as a method of transportation for the field researches, for the non downtown sites driving is more effective.

In field research, a two person technician team, travels to the physical site of the address in question. At the site, the technicians verify the address number, street name and type and adjacent addresses. Based on the field research address points are added or modified. Also, while in the field the technicians look for new construction and unusual addresses. The technicians also inform the DC Department of Transportation (DDOT) if there are discrepancies between physical reality and DDOT's Street Spatial Database.

C. Digitizing Addresses

After individual addresses or record changes are verified, these are digitized. ArcGIS 9.2 software is used for digitization. The changes are digitized in a shapefile. During the digitization the type of change is recorded.

D. Address Anomalies

Address anomalies are basically unusual addresses. These address anomalies are significant because these cases are reality and had to be accounted for in the repository. The following anomalies were found while working on the MAR:

i. Out of Sequence

The out of sequence addresses were real challenge. The only way to confirm these cases are by field research.

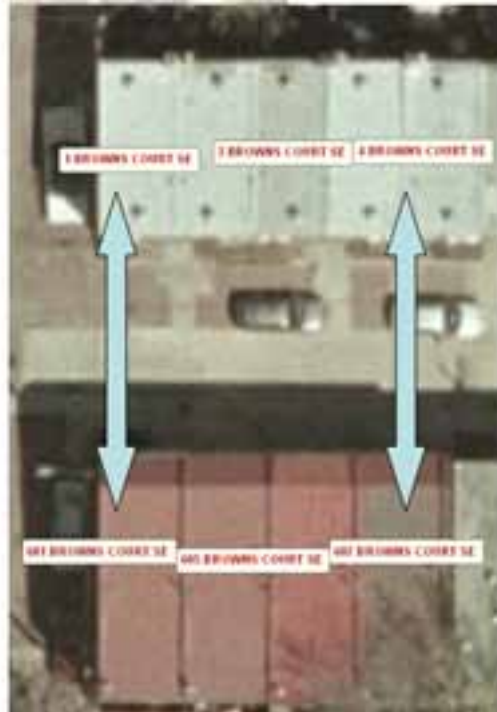


ii. Address Far from Its Associated Street Name



iii. No Address Pattern Along The Same Street Segment

For example Browns Court, assigned addresses for this segment has no pattern. On one side of the street, the address numbers ranged from 1 to 11 and on the other side it is 601 to 607.



iv. 1/2 Addresses

1/2 addresses are found when there was not any possibility of assigning a logical number between two previously used addresses. For example between 2222 V STREET NW and 2224 V ST NW, there is found 2222 1/2 V STREET NW.



v. REAR Addresses

There are very few cases of rear addresses that exist. They are assigned to buildings that fall behind a primary address. The rear portion of the address is considered an address number suffix.



vi. Addresses on the Wrong Side of the Street

As a normal practice, even and odd addresses fall on different sides of the street. However, there are exceptions.



IV. Data Loading

The digitized data from the weekly address verification process are now reviewed and then loaded into the MAR database. The four steps in the data loading process are:

- Quality control for digitized data
- Separate data into categories
- Loading into Oracle using SQL Scripts
- Quality control post data loading

A. Quality Control for Digitized Data

The digitized records from the address verification process undergo quality control procedures. The quality control process has been standardized and has the following steps:

- The digitized data contains all necessary fields
- Values in each field are within the field's domain
- There are no duplicate addresses in the
- The unique roadway segment identifier matches an existing DDOT's Street Spatial Database along with the associated street name and street type
- A spatial join between the digitized points and the DC Quadrant layer

B. Separate Data into Categories

Once the quality control process has been completed, the data is then divided into categories. The division into categories allows for customized SQL scripts to be run on each category that load the data into Oracle.

The records are then separated into the following categories:

- New Addresses, Full Address, No Alias Name, Not In MAR
- New Addresses, Full Address, Alias Name, Not In MAR
- New Addresses, No Full Address, Alias Name, Not In MAR
- Updated Addresses, In MAR
- Retired Addresses
- Deleted Addresses
- Alias Change
- Alias Add
- Alias Delete

The data is then stored as dbf files on a shared drive for the Oracle database administrator to load into the MAR database. An email is sent which informs the Oracle DBA and other interested parties of the week's progress. The email details the number of records in each category, what sources were used, and a map showing locations of updated and new records. The data is now ready for to be loaded into the Oracle database.

C. Loading it into Oracle using SQL Scripts

The dbf files are first loaded into MS Access and then imported into Oracle using customized SQL scripts. These scripts were written by the database administrator. For each category there is a unique script that is run.

D. Quality control post data loading

After the data has been loaded into the ArcSDE using Oracle, it is tested to ensure that the data loaded is correct. Just as each category is loaded individually, so to each category is tested individually. For one, this ensures that each and every category was actually loaded. There are two reasons why data loading would produce incorrect results.

- SQL Script that was used has a problem
- The data given does not meet specified requirements

Additionally, if any problems are found with the loaded data, tests will be administered to determine if the problem was data or the SQL scripts. If it was found that there were problems with a SQL script, that individual SQL script can be modified.

Data that did not load properly during a weekly data load will be loaded in the subsequent week.

V. MAR's Role in DC Government & Beyond

Nearly, 80% of databases in local government contain some spatial data. Similarly, a large percentage of DC Government databases contain spatial data. The most common is an address. Addresses are used for crime incidents, student, building permits, health inspections, voter rolls, business licenses.

MAR “acts as a local gazetteer. The MAR is available to all Internet users. It employs spell checking and lists of street name and street type “aliases” (eg, MLK Ave) to make locating addresses as easy as possible. In addition to locating addresses, the MAR also helps users find intersections (4th Street NW and D Street NW), blocks (4th Street NW between D Street and E Street, or the 400 block of 4th Street NW), and place names (One Judiciary Square, also known as 441 4th Street NW). In most cases, the MAR returns the specific building desired. (DC GIS Website)”

As soon as there is a valid address in the one of these government databases it can be geocoded against the Master Address Repository (MAR).

A. Ways to Access the Master Address Repository

The MAR can be accessed in multiple ways.

- Feature class in an enterprise wide ArcSDE
- Address Points as shapefile
- MAR Web Services
- Online Mapping Applications (ArcIMS)

i. A Feature Class in an Enterprise Wide ArcSDE for DC Government

An address point feature class derived from the MAR database is available in the enterprise ArcSDE server. The DC GIS enterprise geodatabase is accessed by many agencies during their regular daily GIS activities.

ii. Address Points as a Shapefile from the Website

An extract of the MAR database is published as a shapefile. The shapefile is available for download off the DC GIS website (<http://dcgis.dc.gov>). The Address Point shapefile is available to the public just like many of DC GIS's other datasets.

iii. Online Mapping Applications

The MAR is now being incorporated into web based mapping applications that use ArcIMS. These applications include the DC Atlas. Addresses can be searched, and mapped by the general public.

iv. MAR Web Services

The MAR is one of the first citywide applications to take advantage of DC's new service-oriented architecture. The capabilities of the MAR are made available to DC agencies as a series of Web services that send and receive XML. Web services such as the MAR allow agencies to take advantage of them while minimizing costly changes to existing systems (DC GIS Website).

B. How DC Agencies are Incorporating the MAR

i. DDOT

The DC Department of Transportation is using the MAR database to update the address number ranges in their Street Spatial Database. Each address is associated with a unique roadway segment. By finding all the address points that are associated with a roadway segment, DDOT can then re populate the actual ranges in their database

ii. Surveyor's Office

At the Surveyor's Office, the online address verifier is used to quickly identify which property square rests on. The property square information is then used to reference plat maps.

iii. DCRA

DC's Department of Consumer & Regulatory Affairs is using the MAR to validate addresses before issuing building permits for existing structures. The DC Government needs to know where construction is occurring in the city.

For new buildings that are to be constructed, a new address must be issued by DCRA. Before issuing a new address DCRA uses the MAR to make sure an existing address does not

Often building permit applicants apply for a permit without having the correct address. The MAR allows DCRA to find and validate addresses for the purpose of issuing building permits.

iv. Metropolitan Police Department

The police department's Tactical Intelligence and Crime Analysis Unit is using the MAR for geocoding crimes. Every weekday, in the early morning, crimes are geocoded and the daily crime maps are studied by the police briefings.

The public can request crime maps and reports for areas of District. These requests are fulfilled by the Research & Resource Development Division. The MAR is used to locate the addresses that are submitted by the public. Once these addresses are located a map can be produced showing crime in the surrounding area.

The Synchronized Operation Command Complex (SOCC) is using the MAR to locate and map the locations of specialized police units, protestors and parade marchers.

C. Beyond the DC Government

The MAR is not only proving useful within the DC Government, but is being applied by other companies and organizations.

Individuals from the real estate industry are using the address database to locate properties. The addresses are a quick way to access underlying property information.

VI. Lessons Learned & Limitations

While developing a master address repository, many lessons were learned and limitations recognized.

- The Importance of Working with Other DC Agencies
- Time Consuming and Expensive Project
- Need for Maintenance
- Automating the Process
- Limits of the Data

A. The Importance of Working with Other DC Agencies

It has been essential to collaborate with other DC Agencies while developing the MAR. Other agencies have provided valuable tables that contain thousands of address records. These tables have enabled the MAR team to find missing and faulty addresses.

Also, the MAR database has certain dependencies that rely on other DC Agency databases. For example the street centerline database used by DCGIS is maintained by DDOT. In order to add a new address record the street name and type has to be in a table contained in DDOT's Street Spatial Database.

B. Time Consuming and Expensive Project

Developing an address database for a complicated city like DC is not a short term or inexpensive process. The process of verifying and digitizing thousands of addresses is time consuming. Given, the address peculiarities and all the mistypes in DC agency databases and sheer number of addresses in Washington, DC, building a Master Address Repository is a major undertaking.

C. Need for Maintenance

Maintenance of the Address repository is as important as building it. New and updated addresses are uploaded every week and even a small mistake in loading may spoil entire thing. For example, if an address location (XY) is messed up then a user cannot even see a single address on the correct location. The District of Columbia is an old city, by American standards, and the older parts of the city tend to have more address peculiarities. . The address pattern for a lot of places is not same. Furthermore, a lot of demolitions of old structures and new construction happen in this city. Keeping up to date is a major task. To handle these situations, addresses are retired whose structure is demolished and address is not in use. The source for addresses of building under construction is the building permits database which is compiled by the Department of Consumer & Regulatory Affairs (DCRA). DCRA is responsible for assigning new addresses in the District.

D. Automating the Process

Automating additional parts of the process would speed up the project. Portions of the pre and post weekly QC process should be automated. An automated QC procedure would check for duplicate addresses, mistyped street names and types, null values, addresses located in the wrong quadrant. Automating portions would save time and money.

E. Limits of the Data

In such a complicated city as Washington, DC the Master Address Repository will never be perfect. Nevertheless, a completeness rate of greater than 99.8% is achievable and an accuracy rate greater than 99.9% is achievable. At these confidence levels, users will rarely come across missing or faulty addresses.

VII. Conclusion

The Master Address Repository is a valuable resource that allows DC Government to work more effectively. The MAR allows the District Government to easily compare address information across databases and agencies. Previously, the format and quality of address information varied across DC Government systems, making it nearly impossible for the government to identify all the activity associated with a given address. The MAR has created a standard for addresses across DC Government.

The response to the MAR has been exceptional, as many DC agencies are incorporating this valuable resource into their daily processes. The MAR is more efficiently allowing the DC Government to deliver government services, to residents, businesses and visitors.

Certainly, building an address database for a complicated city such as Washington is an expensive and time consuming process. The cost and time can be reasonably well quantified. However, the benefits of having a solid address repository are very difficult to quantify. Despite, the difficulty of quantifying the benefits, building an address repository that is widely used has tremendous long term benefits.

Despite the progress that has been made, there is still a significant amount of work left. As the completeness and accuracy of the MAR increases, the time spent finding a faulty or missing address also increases. The MAR team thoroughly believes that 'Every Address is Important'.

About the Authors

David Jackson is a GIS Analyst working for DC Government. David graduated with a BS in GIS and Geography from the University of Maryland, College Park. He is the team lead on the Address and Alias Data Group at DC GIS.

Vivek Srivastav is a GIS Analyst working for ArcBridge Consulting for the past five years. He has extensive experience with addresses.