# *Special Report:*
Geocoding - Achieving the highest accuracy

Ensure that your institution will meet the 2014 Consumer Financial Protection Bureau (CFPB) Home Mortgage Disclosure Act (HMDA) accuracy requirements

Executive Summary: Although geocoding is only one part of the examination accuracy testing, it is substantially different from the other submittable data that is tested. The geocode is probably the most important data element because of its demographic impact. There is no such thing as 100% accuracy for geocodes on a complete LAR due to the complexities and subtleties of the subject, but there is at least one proven solution that can achieve greater than 98% accuracy. This document contains statistical and anecdotal evidence describing how to achieve the highest geocoding precision level possible to help avoid a time-consuming resubmission process or a costly penalty.
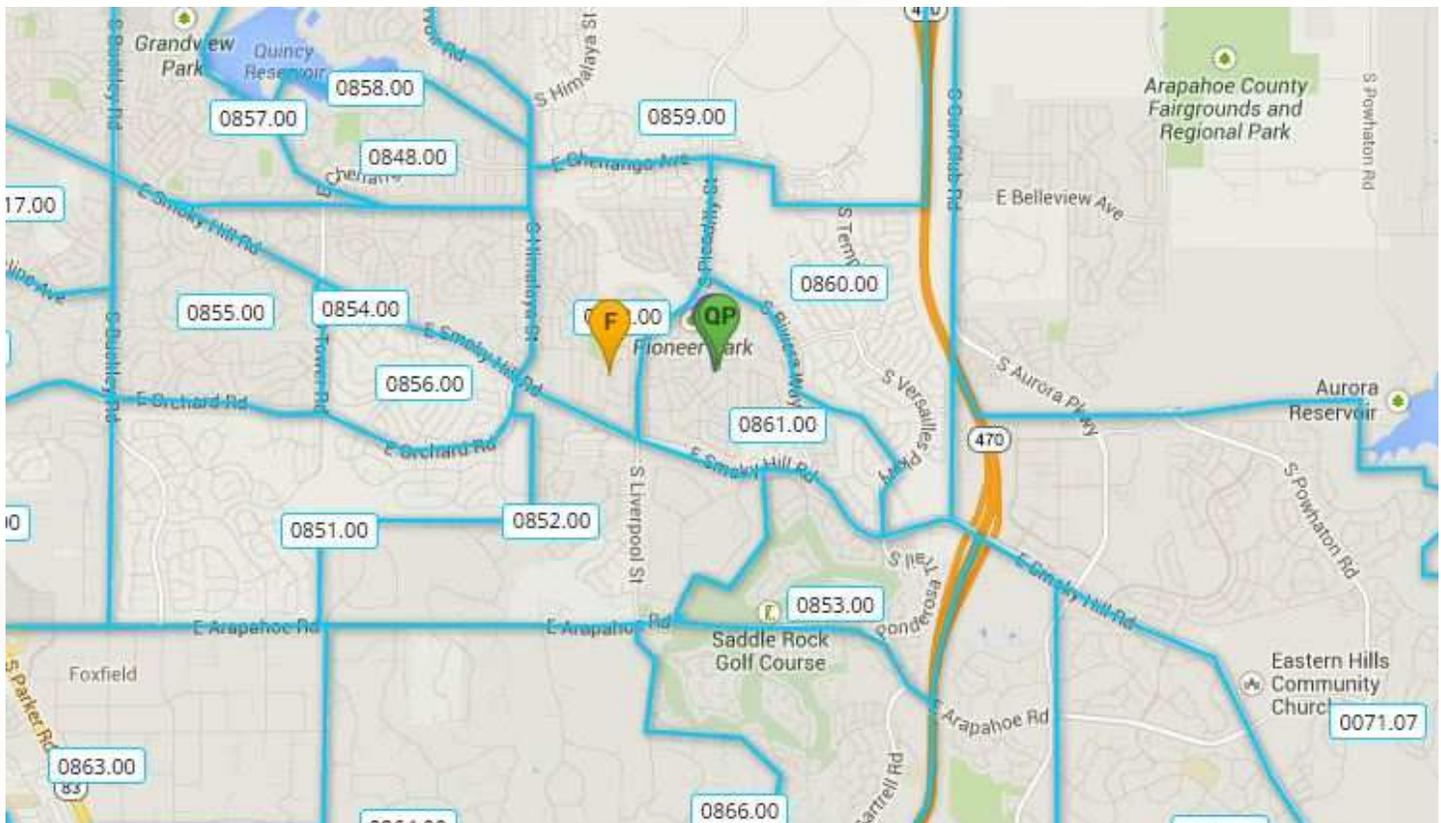
Prepared by: John R. Woloshen, CEO, RATA Associates, LLC – November 2013

The New CFPB Examination Guidelines

According to two documents recently released by the Consumer Financial Protection Bureau (CFPB) (**CFPB Bulletin 2013-11** and **CFPB Examination Procedures and HMDA Resubmission Guidelines**, **Oct 2013**), on January 18, 2014 examiners will begin using modified rules and guidelines for assessing accuracy on critical Loan Application Register (LAR) fields. The agency has already proven their intentions to strictly enforce these finalized requirements by enforcement actions during recent examinations. Two institutions were required to resubmit 2011 and 2012 HMDA data because they failed to meet the agency's accuracy requirements. Fines of over $425,000 have been levied and this seems to be just the tip of the iceberg. Getting HMDA data as accurate as possible on all LAR submission fields is more important than ever, and the geocode fields (MSA/MD, State, County, and Census Tract) are the most challenging fields to get to compliance-grade accuracy levels.

Subtleties of Geocoding

The geocoding fields, unlike the other LAR fields, are not just data that come from the loan application. Geocodes are the numerical translation of the address into the corresponding MSA, State Code, County Code and most importantly Census Tract. For 90% of the addresses, this is a relatively simple task. When you consider ongoing changes to geographic data (new developments, new streets, new houses, changes to street names, etc.) and the fact that census tract boundaries don't always line up with streets, this task quickly becomes very complex.

In order to properly geocode an address, data from two independent sources – the United States Postal Service (USPS) and the United Stated Census Bureau (USCB) – must be merged. Each data source updates its geographic data without taking the other into account. In order to geocode with an automated geocoder, institutions must run a post-deliverable address through the geocoder, which will try to determine what census tract the range, street, and address fall into - if it has record of the address at all. As with any type of conversion, the end result is only as good as the data being processed. There are many issues that affect the geocoding result, such as mistakes within the address fields, missing data elements, or addresses that are too new to be recognized. Geocoder address databases are constantly being updated and the one with the best and most up-to-date data will typically yield the best results.

There are only a few commercially-available geocoders on the market, and they range in price and quality of results. Some geocoders can "hit" more addresses but the quality of the result may be lower, and others may "hit" fewer addresses but the ones they do hit will normally be correct. **Fact: No geocoder can find all addresses or be correct on all geocodes returned. There are errors associated with every method and dataset.**

FFIEC as the Baseline

Regulators use the FFIEC Geocoding/Mapping System as a baseline for comparing geocode results during examinations. The source data for this geocoder is updated twice annually. Users of this web-based tool, found on the www.ffiec.gov website, can submit individual addresses to be geocoded. If the address is found, the geocoder returns the State, County, MSA/MD, and Census Tract.

The FFIEC geocoder will typically find 90% of the valid addresses that are submitted. This means if you process 100,000 addresses, you will still have to geocode at least 10,000 addresses using other methods.

RATA has performed rigorous testing on the results of the FFIEC geocoder and determined that it is approximately 99% accurate on the addresses which return solid matches. From that perspective it is a very good geocoder on the addresses that it can find, but it does have errors and address availability limitations. The other issue with the FFIEC geocoder is that you must type addresses into a web form, so it is not an efficient or cost-effective method for large-volume institutions.

Institutions need a commercially viable solution that can match or exceed the FFIEC geocoder's accuracy, geocode the additional 10% that the FFIEC can't find, and do it in an automated and cost-effective way.

And then there was… APOLLO

Now that we have learned that the FFIEC can only hit 90% of the addresses, that at least 1% of their results will be incorrect, and that addresses have to be manually entered into the FFIEC's website, it looks like we need to find a better geocoder. Since there is no single stand-alone geocoder on the market can do better than the FFIEC by itself, many years ago RATA Associates developed APOLLO, a compliance-grade geocoding server that utilizes multiple geocoders, multiple address databases, artificial intelligence, spatial services, and advanced address standardization.

For over 20 years RATA has been refining its robust geocoding methodology and recent test results have confirmed that APOLLO returns the same result as the FFIEC on 99.997% of the addresses that the FFIEC can actually find, and then does the remaining addresses to the best precision levels that can be done at processing time. See Accuracy and Precision Test Results below.

Over the past 5 years APOLLO has processed more than 35 million addresses and over 100 million since it was first introduced in 1990. Typically it has processed about 20 percent of the total combined HMDA and CRA data submitted to the government by reporting institutions each year since 1990.

APOLLO Accuracy Testing vs. the FFIEC

For our in-depth accuracy tests we processed random files from two customers that do nationwide lending through APOLLO and then processed the same addresses using the FFIEC's geocoder.

Test File 1: 8000 addresses:

7298   - APOLLO matched the FFIEC results exactly.

30*   - Mismatch between the FFIEC results and APOLLO.

672   - Could not be done by FFIEC – "Address not found"

*Of the 30 geocodes that differed from the FFIEC (.00375% of the file) we ZOOM geocoded the addresses and conclusively verified that **APOLLO was correct on 20**, the FFIEC was correct on 5, and 5 were undeterminable.*

Test File 2: 10,000 addresses:

9406 **- APOLLO matched the FFIEC results exactly.**

18* **- Mismatch between the FFIEC results and APOLLO.**

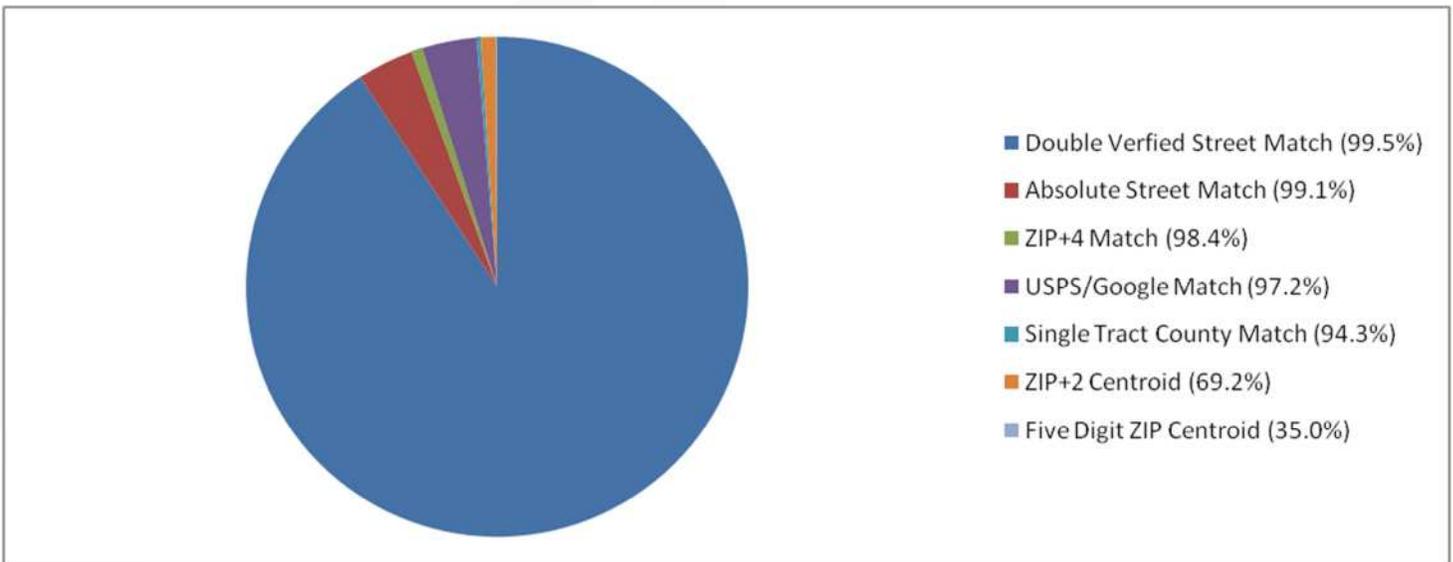576 - Could not be done by FFIEC – "Address not found"

Of the 18 geocodes that differed from the FFIEC (.0018% of the file) we conclusively verified **APOLLO was correct on 14, the FFIEC was correct on 2, and 2 were undeterminable.*

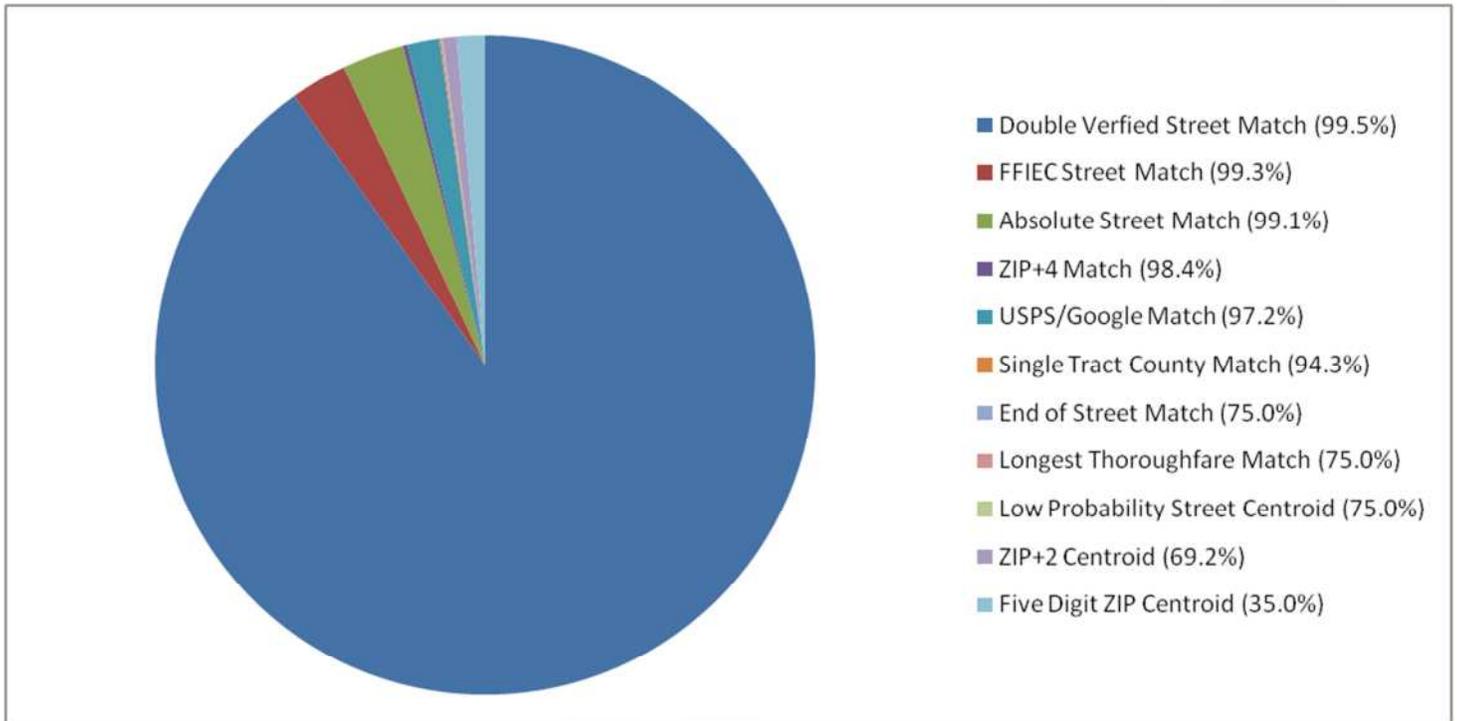## APOLLO Overall Precision on the Test Files

As stated above, there are many subtleties to geocoding that can affect how an address is hit by the geocoders. When addresses are too new or have missing or invalid data elements (ranges, directional, incorrect thoroughfares, ZIP codes, etc.), fallback geocoding methods are needed to get the closest result possible. Each of these fallback methods should be minimized and in our test cases were used on 2% of the overall file.

Full file precision levels are calculated by taking the number of addresses that were hit by different geocoding methods throughout the process and then multiplying the number of records by the average percentage that are normally correct for that method.

Test File 1 – Overall Geocoding Precision from APOLLO: 99.06 percent accurate results

Test File 2 – Overall Geocoding Precision from APOLLO: 98.26 percent accurate results

Legend:
- ■ Double Verfied Street Match (99.5%)
- ■ FFIEC Street Match (99.3%)
- ■ Absolute Street Match (99.1%)
- ■ ZIP+4 Match (98.4%)
- ■ USPS/Google Match (97.2%)
- ■ Single Tract County Match (94.3%)
- ■ End of Street Match (75.0%)
- ■ Longest Thoroughfare Match (75.0%)
- ■ Low Probability Street Centroid (75.0%)
- ■ ZIP+2 Centroid (69.2%)
- ■ Five Digit ZIP Centroid (35.0%)

The RATA ZOOM Geocoder for Auditing and Improving Results

In addition to the APOLLO service, RATA also offers an internet-based geocoding tool which we call the ZOOM geocoder. ZOOM works similar to the FFIEC's geocoder in that you key one address at a time into it, but that is where the similarity ends. ZOOM will run the address through 7 online geocoders including the FFIEC's geocoder and plot the results on a Google map with census tract overlays. This enables the user to ZOOM down to the actual property in question in either map or satellite view. In most cases a user can visually find and verify the census tract. In fact, ZOOM is how we verified the differences between APOLLO and the FFIEC. The ZOOM geocoder is available to all institutions for a minimal per-record charge and is offered to any examiner at no charge. www.zoomgeocoder.com

General APOLLO Test Conclusions

Based on the test results shown above and historical results, we are highly confident that APOLLO will meet or exceed all regulatory HMDA/CRA geocoding standards, including the new CFPB examination tolerances. In summary, APOLLO can geocode 100% of your file, can meet or exceed the FFIEC's accuracy, and is a cost-effective and efficient method for any institution.

RATA Geocoding Services

RATA offers five types of Geocoding services:

Premium (GeoPlus – compliance grade)
Standard (Lower cost "marketing grade")
Passport (Direct integration with loan origination systems)
Custom (Special processing unique to your institution)
Residual Address Matching (RAM) (Used for processing the 'address not found' records from other geocoders)

About the Company - RATA Associates, LLC

RATA Associates, LLC was founded in 1987 with the primary goal of providing cost-effective quality compliance software and service solutions to the financial services industry. Our RATA Comply Suite of software products help institutions deal with all aspects of HMDA, CRA, Fair Lending, Regression Analysis, Mapping, Peer Analysis and Compliance–grade geocoding, which was the focus of this document.

Over the past 5 years the RATA Comply software has processed more than 35 million applications and helped our clients submit approximately 20 percent of the total combined HMDA and CRA data submitted to the government by all reporting institutions each year since 1990.

The RATA Comply Suite is scalable and flexible enough to fit any financial institution. Pricing for all products and services are lending volume based, which means they are cost-effective regardless of the asset size or loan application volume of the reporting institution. Sound interesting?

For more information please feel free to visit our website at www.rataassociates.com

Regarding this special report: **Thank you for reading. I hope you found this informative. If you have questions or comments, please send an e-mail to me at jrw@rataassociates.com.**

Finally, please feel free to pass a copy of this report to others you feel may find it of interest.

John R. Woloshen, CEO, RATA Associates, LLC