Avoiding Common Pitfalls of Data Collection

by Donald R. Narup, President, CRA Information Services

Compliance officers are faced with highly complex data collection and reporting challenges that are essential not only to the supervision, but also the well being of the institution. Since the establishment of new compliance regulations in 1996, the requirements placed on banks to collect, track, maintain, report and analyze thousands of pieces of CRA and HMDA data has increased. Recent changes in the HMDA regulation will intensify this situation.

Compliance officers are often tasked to create their own processes for data collection and analysis to meet these growing requirements. The most common resources used are mainframe systems and financial spreadsheets—neither of which provide an efficient methodology to manipulate the data and prepare the required analysis. Without a customized program designed to handle CRA and HMDA data, these homegrown processes are even more vulnerable to common data collection pitfalls. These pitfalls have the potential to adversely affect the bank's ability to make informed planning and compliance decisions vital to proper management of the institution.

This article will offer tips to help compliance staff recognize and avoid some of the more common pitfalls. These tips are based on years of practical experience in receiving and analyzing compliance data from institutions of all asset sizes and geographic areas of the U.S. This is by no means a complete listing, nor does it address all of the issues surrounding any one item mentioned. It will hopefully help you to begin to assess the effectiveness of your bank's data collection system.

Start Now

Establishing procedures and training staff is the most time consuming part of the effort. Give yourself sufficient time to smooth out the wrinkles before you are faced with the first annual submission. Don't wait until the bank is at the magic $250 million asset mark and becomes a reporting institution to begin collecting information. Starting from scratch, it could take at least a year to get data collection processes in place, especially if you don't have customized software to set up the data fields and proper codes.

Making the Mainframe Work

If you decide to collect data on a mainframe system, determine whether there are any CRA data collection fields already available. If not, you will need to find out how to add the necessary fields to the system. Most systems allow customers a certain number of fields they can customize. They are most commonly called flex data fields. The system you use may have a different name.

Don't be surprised if only a few, or no flex data fields are available. The system has likely been in place for many years, during which time people have set up fields for any number of purposes (many long forgotten). Deleting fields and reports that are no longer needed may require management approval, which may not happen according to your timetable.

Once you determine which data collection fields are available, examine the loan boarding document. Rework it to include the new compliance data fields, or make an addendum page that includes them. Following are some of the more problematic areas you will encounter when adapting a mainframe system for data collection purposes.

CRA Address

- Use a specific CRA address field, as the customer's mailing address may not be the CRA address. Many systems have several address fields that can be used to capture address information. Having separate CRA and mailing address fields also ensures that data is collected in a consistent format.

- Keep in mind that P.O. box addresses are not acceptable. The actual street address to where the loan proceeds are going is the correct CRA address.

- Avoid the common pitfall of using the "collateral description" field to enter a CRA address. Usually, this field is only 35 characters in length, which is often not adequate to accommodate the full address.

- Require data to be entered in the CRA address field even if it is the same as the mailing address. This avoids confusion over whether the CRA address is the same as the mailing address or was simply not entered.

Geocoding

- Addresses need to be geocoded. Geocoding is a skill, requiring particular knowledge of how it works and the database engines used to do it. Not getting a match on the FFIEC geocoder may be due to the way an address was entered in the search fields, and not because it's not in the database.

- Based on my experience, twelve to fifteen percent of all addresses entered on mainframe systems have some type of error that prevent them from being geocoded. Besides spelling, the most common error is placing an apartment or unit number between the street name and directional prefix, as with "123 A E Main St." Address formats have a street number followed by a directional prefix such as N, S, E, W, a street name and then a suffix such as Ave, Blvd, Ln and St. Geocoders do not recognize the "A" as a prefix and will not make a match. The address must be reformatted to "123 E Main St #A" before it will geocode. By entering it correctly the first time, you save yourself some duplication of effort.

- Always use standard U.S. postal abbreviations. The abbreviation for Trail is TRL, not TR. In some geographies, MLK will not generate a match but Martin Luther King will. 4532 S Highway 76 may not match but 4532 US Hwy 76 S or 4532 County Road 76 or 4532 Kentucky Hwy might. If 2016 First St doesn't get a hit, try 2016 1st St.

- Do not enter an address range, as in "1234-1242 Main St." Only one street number can be used to geocode an address. For multiple properties at different locations, pick the one most advantageous to the bank, such as the address in a low- or moderate-income census tract. For multiple addresses at a single location, one street number will suffice.

Putting geocode information on the mainframe is your choice. If you don't need to, then don't. The FFIEC and the Census Bureau's America Fact Finder make geocoding information available at no charge on their web sites at **www.ffiec.gov** and **www.factfinder.census.gov**. However, entering one address at a time gets a little tedious and probably is not an efficient use of time or personnel.

> *"The key to good data collection is to collect it the way you will use it."*

There are very inexpensive software products available that will scrub addresses to check for spelling errors, correct non-standard formats and append current zip and zip+4 codes. These products are designed for bulk mailers; but the address scrubbing features used for bulk mailing purposes are also excellent to use for geocoding purposes. Additionally, they have the capacity to geocode as many as 1,000 records in a few seconds for as little as $50. The more records geocoded, the more the cost per thousand decreases. Yet another alternative for much less than the cost of personnel, an institution can eliminate a lot of work and save far more than the costs they now incur, by using an outside professional geocoding service. These are a few ways to ease your compliance burdens.

Borrower Income

Depending on the CRA loan type classification, there are two different data collection requirements for reporting income. Small business/small farm classified loans require a code number; all other loan types require a dollar amount. Do not mix codes and dollar amounts in the same field. Putting two different kinds of data in the same field causes problems in analysis and reporting.

- Don't customize the codes. CRA and HMDA have well-defined specifications for reporting data using a standardized code. If a small business/farm standardized income code is 1 = $1,000,000 or less and 2= greater than $1,000,000, why use A or B as a code? Using the field titled "income" where the codes A=low, B=moderate, C=middle, D=upper, E=small business and F=small farm is an example of putting two different types of information in the same field as well as using customized coding. To make this information useful, it would need to be converted to another form. The key to good data collection is to collect it the way you will use it.

- Don't allow data entry personnel to round the income dollar amount to the nearest $1,000—the way regulations ask for it. Because rounding errors are common, it is not unusual to see an amount like $36,600 rounded to 40 instead of 37. It is better to have data entry personnel simply copy the full dollar amount of the borrower's gross annual revenue or income. Later, compliance personnel can decide which loans need a revenue code or a rounded dollar amount when the information is needed for reporting purposes.

Proper Coding

- *Small Business/Small Farm:* It is very important to be consistent when designating which loans are classified as small business or small farm. Start by designating the person who will be responsible for making that classification decision and make sure they are knowledgeable of the regulations. It may also help to print some basic guidelines on the mainframe boarding document. For instance, "small business must be $1 million or less, used for a business-related purpose, not collateralized by residential property."

- *Federal Call Codes:* At the time the loan is input a federal call code must also be determined and entered. A common problem occurs when the CRA loan type is changed from small business/farm after the loan has been input. The loan type code is changed, but seldom is the federal call code assigned at boarding corrected. The result is that incorrect amounts are reported on the institutions quarterly call report.

- *CRA Action/Affiliate Codes*: Two other essential data collection fields are CRA action code (1=originated, 6=purchased) and affiliate lending flag (1=originated/purchased by your institution and 2=originated / purchased by an affiliate of your institution). If your bank purchases loans or has an affiliate, these CRA data collection fields need to be added the mainframe.

Understanding Spreadsheets

Everyone has a different way of formatting the collected data into a spreadsheet. Following are some basic rules that I've found to be effective.

- Don't create a separate spreadsheet for every month. At the end of a year you will probably have to combine them all anyway. You can always sort the monthly data using tools in the spreadsheet program.

- Put your collected data in columns, which allows you to sort, and not in rows. Each column should be a separate data collection field.

- If your data processing system uses the same loan number at each renewal, put a -1 or -2 or some designation *after* the loan number. This makes it a unique reporting number per regulations and eliminates duplicate account numbers. Don't put the designator before the account number, as it will throw the account number out of numerical order when sorted.

- Don't enter an entirely different number in the CRA data collection program provided by the government. While the number is unique, examiners cannot cross-reference the loan.

- BE CONSISTENT. All personnel should be entering data the same way. Merging spreadsheets into a usable format becomes difficult if one spreadsheet has a date format of yy/mm/dd and another enters it as dd/mm/yy, or if one has separate address fields for city, state and zip and another has a combined city/state/zip field. Format individual fields the same across spreadsheets on such aspects as number of decimal places and type of data field (numerals, text, etc.).

About the Author

After 30 years in the banking industry, Mr. Narup retired in 1991 as president and CEO of Laguna Bank in Laguna Beach California. He is the president of CRA Information Services located in San Diego, California. CRA Information Services provides compliance data collection and analysis software, and is an outsourcing facility for CRA and fair lending analysis and other compliance-related services for financial institutions nationwide. He can be contacted at (858) 573-2995 or address e-mail to compliance@crahelp.com.

Conclusion

CRA data are one of the best sources of information an institution has. However, without a standardized data collection program, thousands of dollars and thousands of work hours can be wasted. CRA and HMDA data are tangible company assets and not something collected only for the regulators to be reported once a year. This is information that can be used to determine usage of products and services within a marketing area, to show where there is a need for new products and services to meet the needs of specific communities, and to develop marketing strategies that will increase sales and profitability. Don't be satisfied with just collecting data; for an effective compliance effort you must use it. The decisions of your staff are only as good as the information they have can access. ***CI***

edited by Anita Todd