

# Data, Data Everywhere, But Can It Be Used?

**Errors using inadequate data are much less than those using no data at all, said Charles Babbage.**

Since the collapse of the mortgage markets and the revelation that the data used to price and rate those securitizations was bad, one of the primary focuses of the industry has been on data. On top of our own concerns, regulators have identified that the transmission of accurate data is one of their risk management issues and they also see it as one of the best ways to effectively monitor mortgage lenders and servicers. It would be wonderful to wave our magic wands and resolve all these data issues. However, that just isn't going to happen. As an industry we need to take a look at all the issues and the opportunities that are present today and decide individually and collectively how to best attack the problem.

**BY BARBARA PERINO  
AND REBECCA WALZAK**



## WHAT ARE THE CURRENT DATA ISSUES?

The place to start of course is in the identification of what data we need. While loan origination systems collect a large amount of data about the borrower, little of it is used past the origination process. Furthermore, data that is needed to analyze, evaluate and securitize these loans is not collected. For example, while we now know if the borrower is self-employed, we still don't have a place to capture the type of legal structure under which the company is organized. Without this it is more difficult to evaluate income. And speaking of income, should we be capturing better information on "other" types of income? Many times the loan applications that I review simply show one lump sum of this data. Without being able to break it down into the various components it is very difficult to ensure that it is continuing and stable. If the borrower receives child support there is no place on the application to indicate the duration of the income. Lenders have to rely on the fact that someone reviewed the divorce documents and substantiated this information.

Servicing data has traditionally been focused on information that is relevant to collecting payments, paying escrows and performing default management functions. However when it comes time to analyze borrowers for modifications or other issues, the original information about the loan is missing. HUD has always required that servicers identify loans where servicing information conflicts with the origination data in order to determine if there was any fraud in the file. However, this has been impossible with the limited amount of data transmitted.

Now of course we are faced with another challenge; the data required for regulator reviews. Of course the primary focus of servicing data is the information necessary to perform the investor reporting functions. However, the regulators have turned that approach up-side-down and now are expecting data to be focused on ensuring that the consumer's issues are ad-

ressed. This involves standardizing and collecting much of the information that previously was stored in the individual "note" screens and as we all know, reporting on free form data is impossible. Data that identifies a consumer's question or inquiry as well as a compliant must be tracked and reported on. Resolution types must be collected as well. In addition to these data fields, servicers want and/or need to have data that helps them address any operational problems and control the risk of non-compliance with the new regulations. While MISMO has indicated that they will have all the new data requirement fields defined and formatted by the end of the year, this does not mean they will be included in the servicing and/or origination systems. This brings us to the next issue. Collecting reliable data.

In the past few years one of the best ways to get a conversation started with mortgage bankers was to ask the question "How do you handle data

find a data field labeled "product type" to contain a code for "property type" or loan amounts to vary throughout the data set. Boarding loans onto a servicing system is a great way to test the reliability of the data into various fields. However servicers are not without their own data quality issues. Once again, manual input in many of the fields is most likely the source of the data issues. While some of the data issues are being addressed with electronic feeds of information this approach does not solve all the problems. One of the suggestions heard to address this problem is to implement data consistency checks on each loan application and download.

Of course, even with recognizing the risks of incorrect data input, the issues of having data inconsistent with the documentation in the file still exists. Having reviewed numerous files, it is not uncommon to see that income information is calculated incorrectly and input into an AUS system inac-

The leaders who will **emerge successfully** from the problems of today will be the ones that know how we use data to make their organization **more efficient, effective and profitable.**

integrity?" Everyone has an answer but unfortunately, they are all different and this opens up every lender to the risk of poor quality data. Despite the fact that MISMO has done an excellent job of defining and formatting industry data, the reality is that not all data is accurate. This is primarily because so much of it is input through a manual process. It is not unusual to

curately. As a result, the AUS result is suspect and the integrity of the data is once again also suspect. With Fannie Mae's recent announcement that they will only accept "perfect loans" this type of data integrity issue becomes more critical.

Servicing is not without similar problems, especially in the area of modifications, principal reductions

and forbearance. Inputting inaccurate information during this process can result in approving those that should not be approved or denying those that should. In addition, timing of the process is a critical factor for regulators. Inputting inaccurate information can delay meeting the deadlines for response.

The only way to really address these problems is through a quality control review during the entire processing, underwriting, closing and servicing of the loan. While manually conducting these reviews will alleviate some of these risks, the associated cost in terms of time and personnel

sion of data through the Internet and DVDs, provides multiple opportunities for individuals to abscond with this data. While we have many experts in the area of data security to utilize as a resource, in the end it is up to each individual lender to make sure that they have the means to ensure the data is secure.

#### **USING DATA**

The use of the data we have always collected has traditionally been limited to transmissions involving the secondary market and to investors. Now regulators, the secondary agencies and other lenders are becoming more active in seeking and using data

or agency tells me I have problems that are worse than others, how do I know that is true? With the focus on consistent, reliable and common data elements lenders and servicers will have the opportunity to conduct their own comparative analysis. The industry has long been able to see overall delinquency rates and analyze by product, rate and location, but expanding this data into the area of Operational Risk is just beginning.

Mortgage True View, a company focused on collecting data from servicers and originators, provides not only the ability to analyze the Operational Risk within your own organization, but to compare it to others in the database. Of course, the comparative data is comprised of multiple participants without any type of lender, servicer or borrower information included. While it has always been intended as a means to establish the regulatory standard variation rate in a company, it also provides numerous other opportunities for lenders and servicers to identify potential issues and make changes either before it becomes a problem or to take advantage of a market opportunity before others.

Knowing that the data transmitted to investors is accurate and reliable will also support the revitalization of the secondary market. Investors have been reluctant to get back into the market and want some assurances that this time we will give them correct information about the risk within the mortgage files. If every loan contains the results of a data and operational review that establishes the overall reliability of the file, investors would have the opportunity to make fact-based decisions on what level of risk is acceptable to them.

#### **ROADBLOCKS**

If all of this focus on data is good, what is preventing the industry from achieving this goal? While there are multiple reasons, there are three that stand out. First is the breadth of the data itself. As mentioned earlier, the amount of data that lenders need to capture continues to grow. While the defining and formatting process are

**Investors have been reluctant to get back into the market and want some assurances that this time we will give them correct information about the risk within the mortgage files.**

make it prohibitive. The only real solution is an automated QC review that identifies the issues and allows changes to be made as issues occur.

Data security is also a big concern. Since we do obtain and maintain a large amount of the borrower's personal information, we are charged with ensuring that it is safe from unintended leakage and intrusion. Most lenders have incorporated firewalls into their system structure, but as we have all seen, these are not impenetrable. Failure to update passwords, sharing of information and just plain carelessness creates numerous risks for those whose job it is to keep the data secure. Furthermore the widely accepted use of laptops and transmis-

to identify and establish operational, credit risk and regulatory risk.

The CFPB announced from its early beginnings that it would be a "data driven" organization. Other regulators, including state audit groups followed suit. Unfortunately due to all the issues previously identified, the initial attempt to evaluate lenders and servicers was not successful. However they continue to pursue this approach, which is an additional incentive for lenders and servicers to focus on the overall quality and integrity of their data.

This attention to data also has its benefits. One of the issues that always arises is how does my company compare to others. If an investor, regulator



being addressed, the addition of these data fields into origination and servicing systems will take more time. Add to that the tendency of the regulators and agencies to want more and more data, it is a daunting task to get a complete data set incorporated into any industry system.

A second issue delaying the implementation of this data is the industry's historical lack of focus on their operations as a whole rather than as a group of separate and distinct processes. In order to make sense of all the Operational Risk that a comprehensive set of data would expose, management must be able to comprehend the statistics that are presented to them. For example, if a QC report tells them that there is an error rate of 5%, do they know how to determine the validity of that statement? Or better yet, does the QC staff know how to position the results within a confidence interval so that management does not go off on a wild goose chase to resolve a problem that is in fact just random variation. Understanding how to present data as well as how to determine what actions should occur is a task that management and staff need to undertake. Continuing to focus on issues on a loan-by-loan basis, which is where we are most comfortable, is not helpful and limits our ability to effectively manage operations.

The third issue which I have heard

from numerous speakers and industry members is fear. It is not unusual for anyone to fear the unknown and that is where we are today. The future of

have an election, so why spend the time and energy to deal with them. Unfortunately, whether or not they change, one thing is certain. The in-



#### ABOUT THE AUTHOR

Barbara Perino is Director of Business Development for Direct Valuation Solutions a new cloud based technology platform that is designed to streamline and enhance the appraisal fulfillment experience. Barbara has 23 years of extensive sales and sales management experience along with operational knowledge from all facets within the residential property valuation industry. She is also a professional certified coach.

#### ABOUT THE AUTHOR

rjbWalzak Consulting, Inc. was founded and is led by Rebecca Walzak, a leader in operational risk management programs in all areas of the consumer lending industry. In addition to consulting experience in mortgage banking, student lending and other types of consumer lending, she has hands on practical experience in these organizations as well as having held numerous positions from top to bottom of the consumer lending industry over the past 25 years.



the industry, what will be demanded of us, the costs associated with meeting all these data requirements, is still unknown. This causes a lot of stress and hesitancy in taking on new issues and dealing with old problems. In fact it is not uncommon to hear managers say that these new requirements are going to go away once we

dustry has changed and at the heart of that change is data, what we collect, how we manage it is now a top priority for anyone in the business. The leaders who will emerge successfully from the problems of today will be the ones that know how we use data to make their organization more efficient, effective and profitable. ❖