Can You Prove YOUR DATA INTEGRITY?

BY LLOYD BOOTH AND ANDY YOUNG

EVERYONE WANTS TO BE RISK AVERSE AND ENSURE THE INTEGRITY OF THEIR DATA, BUT FEW KNOW HOW TO ACTUALLY DO IT.

In today's regulatory environment, lenders are asked not only to assure the integrity of the data in their system, but also to prove it. Everyone is concerned with data integrity, but there is little information on how to address it. How does the lender manage their systems, integrate with a myriad of service providers, layer on their own operational idiosyncrasies, and still maintain certainty in the information? This article presents four key issues that can threaten the overall integrity of the data. For each issue, we will offer ideas on how to identify and remedy data integrity obstacles. The result is a guide to assist in the assessment and management of the data integrity standards in your organization. The purpose of an assessment may differ for each reader; but the concepts and strategies are relevant for any business.

Data integrity must be achieved and demonstrated within real production, and also proven to be built into the systems, processes and organizations. Accurate, quality data is a result of equal parts technology, service provider standards and organizational procedures. We are focusing on the issues the executive staff needs in order to understand and have faith in the information produced by their company. The issues we'll discuss include:

- **Disparate Expectations**– (consistent interpretations of quality.)
- **Data Evolution** – (how the data is perfected over time.)
- **Data Equivalence** – (consistent data for all participants.)
- **Process Influences** – (workarounds and interactions that affect data.)

The perspective of each organization will determine the priority of these issues. Understanding the details in each issue is critical to creating environments and processes that can improve data integrity.

This article will cover the first two issues and next month a follow-up article will cover the last two.

DISPARATE EXPECTATIONS

If you were asked to describe a quality meal you could quickly describe the attributes of what quality means. There are differences from one person to the next, but everyone still agrees on what constitutes a quality meal. Defining data integrity is not so straight-forward. Data quality, data integrity, data transparency and data fidelity have all been used both synonymously and distinctly to describe our information. These terms are used interchangeably but not always with the same definitions. Understanding what we mean with these terms is crucial.

So, what is the difference between data accuracy and data validity? When a data element passes all business rules that may be applied, it is considered valid. Whereas; accuracy is a reference to whether the information is correct given the current perspective of the data.

> Data integrity must be achieved and demonstrated within real production, and also proven to be built into the systems, processes and organizations.

Consider 11/03/02 as the date of a loan application. If the field is entered erroneously as 03/11/02, the field is still valid since it passes any business rules, but it is not accurate since it is the incorrect date. From a business standpoint the data needs to be both valid and accurate. Ideally data accuracy must warrant the data to reflect "the whole truth, and nothing but the truth." However the "truth" depends on many factors:

- **Loan Stage as available information evolves** - Mortgage lending transaction data is meant to evolve over time as the information is perfected.
- **Perspective of lending participants** – Each participant can have a different view on what the data means and how it is represented in their systems.
- **Best faith resolution of data discrepancies** – The mortgage lender must reconcile the multiple versions of the data from different sources and different points in time.
- **Supporting information trail that can be audited** - In the past, organizations were able to rely on paper files with handwritten notes in the margins. In the virtual data world, these notes must be systematized in the technology.

Historically any "fuzzy truths" in the paper loan file might be explained away based on undocumented factors and intuition. Today, technology systems need to provide accessible, high-quality data. A supporting employment document might use "Bill" instead of "William" for the borrower's name and be ignored in the paper file. Perhaps an irrelevant discrepancy, but technology systems must explicitly reconcile the variation.

Data transparency is often used to infer data integrity. Transparency refers to the ability to access and modify data regardless of the system or organization that created it. The problem is an expectation that the accessible data is actually coming from the official source; transparency doesn't necessarily ensure data integrity.

Another consideration is data fidelity. This refers to a system's ability to faithfully reproduce all information that is shared with that system. Data fidelity is compromised when a system receives a set of data and discards any information that is not specifically required by it. A lender system provides all of the detailed employment information to a vendor, and the vendor is only concerned with the aggregated totals and so it discards the details that made up the total. In a one-way interface this "filtering" of data may be acceptable. However, if the vendor passes aggregate information to a separate service, there is a possibility of degraded fidelity (and consequently integrity).

WHAT CAN WE DO?

We must recognize that lenders are responsible for assessing the accuracy of all data captured for the loan. Vendors should play a role, but the lender needs to develop standards that hold the vendor accountable for their part in the overall data integrity strategy.

The lender must understand what data is the most critical and which information provides either the most protection or the most risk. This understanding is best managed through a lender defined data dictionary. This data dictionary is focused on creating a shared business understanding of the

important fields, the "human" meaning of them, their usage, dependencies, and what systems or people impact them.

This dictionary should be shared with vendors and updated as new fields, procedures and requirements are discovered. Most vendor interface projects start with a set of required information that is well described by industry standards such as MISMO. However, the impact and dependencies of fields on each other may be missing. If the interface requires a field that only gets updated when a certain loan program is selected, is the value valid when the loan program changes? With a proper data dictionary these questions can be sorted out during the design and testing of the interface. In addition, this dictionary helps people within the lender's organization that have various backgrounds and have developed different understandings for "common" terms.

DATA EVOLUTION

Everything that we do in the manufacturing of a mortgage loan is designed to perfect the data of the loan. Income estimates are replaced with exact dollar amounts, additional information clarifying property values is added, and so on. A great example of the impact of data evolution is the change in procedures brought by the Mortgage Disclosure Improvement Act ("MDIA"). This act caused LOS vendors to provide mechanisms to understand and track the difference between the loan costs provided to the borrower at application, and the loan costs presented to the borrower at closing. Business rules and workflow can then be applied based on any differences. With the MDIA requirements, systems now have a partial mechanism to accommodate data evolution.

WHAT CAN WE DO?

As part of the data integrity assessment, we must understand and investigate the issues directly resulting from data evolution. Data evolution is quite normal, but the fact that we can't depend on the values in our system can create problems during the production of a loan. These problems manifest themselves in critical ways:

- **Data Ghosting** – This occurs when the value of a data element is modified but the data associated (and often automatically entered) with that element are not updated. An example could be a loan at application entered as a FHA loan, and later switched to a conventional mortgage. To accurately reset all of the loan parameters, the system of record must update all fields that describe either the old or the new loan program. If the related loan data isn't correctly updated, FHA specific dates and amounts could be erroneously present. This ghosted data could be shared with third parties leading to issues with their interpretation of that data. The data is accurate but not valid.
- **Non-reproducible Transaction Sets** – When multiple fields are shared with trading partners or third party interfaces (automated underwriting engines for example), the data shared is often referred to as a "Transaction Set." It contains all of the fields shared with the third party at the time a service is requested. Many systems of record can't reproduce the values that were shared at the time of ordering.
- **Indirect Data Updates** – Most LOS databases are designed to destroy data. While this statement might provoke some debate, in fact, these systems only allow a single value of any given database field. As a database record is added or updated, the previous value of a field is replaced with the new value. Each time the borrower FICO score is updated, the system loses the prior value and only has the most recent.

Many loan origination systems provide the ability to track changes made to individual fields. These "audit systems" or "audit logs" include data about who made the change, when, the old value and the new value. These basic tracking systems are focused on isolated field changes after the data has been modified and are often blind to changes made outside the system. Audit logs do not reflect data changes as a group within the context of events in the lending process. If the current architecture doesn't adequately track data evolution, how can these systems deal with this complex issue?

One solution approach to look for during the assessment is a database architecture design known as Point in Time Architecture (PTA). PTA provides support for both history and an audit trail of relevant data. The basic goal of a PTA system is to provide a version of the data as it existed at a given point in time. Databases designed with PTA have been implemented to "roll back" a set of data values. Many accounting systems need to roll back data changes to a specified point in the billing cycle. For a mortgage lending system we are not as much interested in rolling back the data as we are in managing and reconciling how the values evolve throughout the lending process. For our purposes, two key design philosophies are important to be looked for in a lending PTA system:

- **Recording triggered by lending milestones.**
- **Definition of specific data to record.**

For the PTA enabled system to improve data quality, the versions of the data must be understandable

Accurate, quality data is a result of equal parts technology, service provider standards and organizational procedures.

in the context of lending milestones. PTA databases provide a mechanism to identify key milestones (triggers) within the lending process that affect data integrity. Examples might include delivery of the advanced disclosure, submission to an automated underwriting engine or a request to prepare loan closing documents.

Not every data field warrants historical tracking. Examples might include static identifier fields, system assigned dates such as the date the loan was entered or even a telephone number. In most instances it is not important to track that the telephone number has changed, only that the system reflects the most accurate version of the telephone number. Obviously some organizations will attach a greater level of importance to specific fields. A PTA system needs the ability for the organization to define which fields are important at which events.

Data "fields" to look for should logically include individual internal data elements, complete transaction data sets and document images received or generated. For instance, a trail of AUS transactions or TIL images is needed to understand the state of a mortgage loan at a point in the mortgage process.

In this half of the article, we have established the meaning of some key data integrity terms for mortgage lending and what needs to be included in a data integrity assessment within the organization. Next month, in the final half of the article, we will explore how data equivalence is lost between various internal and external participants who share data and what to look for in assessing this aspect of data integrity. ❖

About The Author

Lloyd Booth is the Owner of Mead Technology Solutions, LLC (MTS), a provider of technology consulting services to the financial services and mortgage lending industries. Prior to MTS Lloyd served as the President and COO of Blueberry Systems, LLC a mortgage lending technology solutions provider. At Blueberry he oversaw the design, development, marketing and implementation of the company's flagship product "Relay" an enterprise loan origination system. Prior to Blueberry Systems, Lloyd was a co-owner and CIO for ShadowNet Mortgage Technologies, an industry leading document preparation firm. ShadowNet was acquired by First American Financial where Lloyd continued on as the company's CIO and later division president. Lloyd has nearly 30 years of experience in the mortgage lending industry providing executive level technology leadership and management.

About The Author

Andy Young provides mortgage technology consulting services as the owner of Effective Business Systems, LLC. Mr. Young assists major lenders, vendors and other mortgage industry participants in creating and executing strategic mortgage technology initiatives in the origination, servicing, and default phases of lending. His clients include Fannie Mae, Bank of America, Pega Systems, First American, Irwin Mortgage, Temple-Inland Mortgage, GE Mortgage, Banco Hipotecario, and ICICI bank. In addition to consulting, Mr. Young has had a mortgage technology executive role with ISGN, Aurora-Lehman, Mortgage.com, CheckFree-London Bridge, CPI-Alltel(LPS), Fannie Mae Software, and Financial Software.

Index of Advertisers