



Manufacturing the **PERFECT MORTGAGE:** An **INNOVATION Challenge**

Thought Paper on Proposed Expansion of the Uniform Mortgage Data Program

Team:

Innovation Challenge Data Validation Team

Randy Gillis - Black Knight Financial Services, *Thought Paper Lead*

James Grizzle - Informatica Corporation, *Data Validation Team Chair*

1/1/2014

Table of Contents

1. Executive Summary	3
2. What is UMDP?.....	3
3. Core Data Set Recommendations.....	4
3.1 Core Data Set Definitions	4
3.2 Core Data Set Data Validation.....	4
3.3 Core Data Set Ownership	4
3.4 Critical Exchanges of the Core Data Set	4
3.5 Data Validation of the Critical Data Exchanges	5
3.6 Core Data Set Maintenance and Change Control	5
3.7 Separation of Core Data Set from Operational Data Set	5
3.8 Components of the Core Data Set.....	5
4. Validation of the Core Data Set Recommendations	7
4.1 What is Validation of the Core Data Set?.....	7
5. UMDP / CDS Development Process Possible Next Steps.....	8
6. Out of Scope	9
6.1 Definition of Operational Data	9
6.2 Ownership of Operational Data	9
7. Appendix –Data Validation Innovation Challenge Team Members.....	10

Proposed Expansion of the Uniform Mortgage Data Program

1. EXECUTIVE SUMMARY

The recommendations outlined in this Thought Paper were developed by the CIO Forum's Innovation Challenge Data Validation team. The team proposes expanding the Uniform Mortgage Data Program (UMDP) to include additions to:

- The initial residential loan **Core Data Set (CDS)**
- Subsequent **data validation/quality assurance steps** during critical information exchange points across the mortgage life cycle.

We recommended that these proposals be implemented through the Mortgage Industry Standards Maintenance Organization (MISMO) and be included in the MISMO work products. Regarding implementation, these changes should be included in the MISMO Implementation Guides and announced through MISMO's standard communication channels.

These proposals are foundational elements of a mature origination, servicing, and securitization data validation processes. One of long term goals of the UMDP data validation recommendations is to ensure that the process is Capability Maturity Model Integration (CMMI) Level 5 (Optimizing) compliant.

2. WHAT IS UMDP?

The UMDP is a multifaceted, ongoing program in which the government-sponsored enterprises (GSEs) develop and implement mortgage data standards via MISMO for the single-family loans they purchase and/or securitize. The UMDP benefits the industry and housing market by standardizing mortgage delivery data through the Uniform Loan Delivery Dataset (ULDD) and appraisal data through the Uniform Appraisal Dataset (UAD). Additional projects under the UMDP umbrella include:

- **The Uniform Mortgage Servicing Dataset (UMSD)**
A standard dataset that facilitates data exchanges between servicers and investors with standardized definitions, formats, and valid data values.
- **The Uniform Closing Dataset (UCD)**
A common industry dataset to support the Consumer Financial Protection Bureau's (CFPB) proposed Closing Disclosure form.

3. CORE DATA SET RECOMMENDATIONS

3.1 Core Data Set Definitions

Currently, UMDP and MIMSO¹ do not provide concise descriptions of mortgage products; instead their focus is on broad facets of the mortgage life cycle. We recommend that UMDP should clearly define the critical entities and attributes of the Core Data Set and incorporate those definitions into the MIMSO Logical Data Dictionary (LDD).

3.2 Core Data Set Data Validation

We propose that data validation of the core data set be expanded to include:

1. The business and technology rules and the corresponding validation rules for the core data elements.
2. The business and technology exchange rules using the MIMSO Business Rules Exchange (BRE) specification.
3. The Source of Trust (SOT) for each element in the core data set.
 - The SOT is the system, document, process or location of absolute certainty for the value of each core data element.
 - The SOT for a given data element may be different at different data exchange points.

3.3 Core Data Set Ownership

Currently, ownership of the CDS is not clearly defined; however, we believe that the mortgage product CDS should have clear ownership. Without clear responsibility for the data and a “master” representation, there will continue to be multiple conflicting data sets. We recommend that responsibility for the core data set should be assigned to the investor upon purchase of the product since the core data set reflects the asset itself. Because of the legal ramifications, we believe that ownership rules should be established by legislation or by regulatory decree at the federal level.

3.4 Critical Exchanges of the Core Data Set

Critical exchange points among stakeholders are specified by UMDP with regard to the format and structure. These points are also defined within the MIMSO work products and/or standards. However, we believe that further standardization and better definitions are needed. Critical data exchanges contain some, but not necessarily all of the CDS. We propose that the critical exchange points listed below be incorporated into the process:

1. Originator to Investor (ULDD for the GSEs and Other Investor Uploads)
2. Originator to Servicer (Loan Boarding Specification)
3. Servicer to Investor (Investor Reporting Specification, Default Reporting Specification)
4. Servicer to Servicer (Loan Transfer Specification)

¹ MIMSO uses a set of defining parameters to define a mortgage product.

3.5 Data Validation of the Critical Data Exchanges

Each of the critical data exchanges must validate the core data set using a set of standardized rules. We recommend that MISMO and UMDP better describe these rules and make them part of the freely available MISMO work products.

These rules should be provided via the MIMSO Business Rules Exchange Specification. Once the CDS is validated and incorporated into the Securitization Platform, subsequent exchanges of the CDS should be validated by comparing the exchanged version of the CDS to the authoritative copy within the CSP or by using the standardized rules.

3.6 Core Data Set Maintenance and Change Control

Maintenance and change control procedures for the CDS is lacking in the industry. The ideal place for this function may well be in the investor securitization platform or similar technical solution. Therefore, we propose that the securitization platform should:

- Contain the authoritative copy of the Core Data Set during the loan's life cycle.
- Synchronize the authoritative copy on a routine basis with the core data set that supports the operational management of the loan
- Maintain all versions of the core data set and all associated metadata describing any changes made during the loan life cycle.
- Contain the core data validation rules as maintained and published by the MISMO standards development process.
- Send a change control notification to the Investor for approval anytime a CDS value is changed.

3.7 Separation of Core Data Set from Operational Data Set

The definition of the core data set should include a clear separation of the residential loan CDS elements from the elements reflecting the management of the mortgage product throughout its life cycle. This is to emphasize the ownership and maintenance roles related to the core data set.

3.8 Components of the Core Data Set

The Core Data Set represents the critical characteristics of the loan product. The CDS consists of four specifically related components:

1. The Container

The container is metadata that describes certain characteristics or constraints (i.e. rules) about data validation of the CDS both in business and technical terms. Data quality and validation begin with these definitions.

Containers are further enhanced by specific data container taxonomy as generally recognized by the entire Technology Industry. This taxonomy (classification scheme)

with support a specific set of data validation constraints finely tuned to each type of container. Containers can be expressed as entities or attributes:

- **Entities** are described as things of interest to the mortgage process. CDS Entities have data attributes, which include at least two required attributes: the Entity Identifier and the Entity Description. CDS entities should be fully populated during the Origination Process. Examples of entities include:
 - *Product* which describes the loan type and terms
 - *Account* which describes balances and status
 - *Borrower* which describes the party who promises to pay the loan
 - *Property* which describes the collateral for the loan
- **Attributes**, also referred to as elements, describe the entities.
 - Each attribute will have a designation of its Source of Truth (SOT).

2. The value within the container.

The value is the persisted fact about the loan that is observable and actionable by the stakeholders in the loan life cycle. It is described and certified by the Source of Truth (SOT).

3. The description of the CDS metadata within the MIMSO Logical Data Dictionary (LDD)

4. The description of CDS values within the MIMSO LDD.

4. VALIDATION OF THE CORE DATA SET RECOMMENDATIONS

4.1 What is Validation of the Core Data Set?

The Core Data Set is validated by business rules and the governance of those business rules. These rules represent the basis of trust in the critical characteristics of the loan product.

We recommend that the proposed CDS Data Validation consist of four specifically related components:

1. Business Rules expressed in metadata

- *Entity business rules* (constraints) are expressed as Identifiers and relationships.
 - Each *Entity* should be further described by key attributes.
- *Identifiers* uniquely describe each occurrence of an entity.
- *Relationships* describe existence constraints such as required, optional or conditional. Relationships are also described as cardinality such as One product is described within one and only one account (a one-to-one relationship) or One account must have one or more borrowers (a one-to-many relationship).
- *Entity Attribute metadata business rules* (constraints) are expressed as metadata characteristics such as data type, length, precision, business description and allowed enumerations.
- *The SOT* is an attribute metadata characteristic describe in textual form that builds the foundation of trust in the value contained by the attribute.
 - It will ensure each value is traceable to the foundation.
 - The foundation could be a field on a mortgage loan related document, a key business process, a key content provider, etc.
 - The SOT designation should be subject to the data validation governance process.

2. Business rules that cannot be expressed in Entity and Attribute associated metadata.

- Many critical data quality rules cannot be expressed in metadata. Instead, these rules can be expressed in a consistent way using an English language form and include life cycle influences such as:
 - authoritative copies and copy management,
 - security at rest and during delivery,
 - retention and versioning,
 - archival and contextual preservation over a long period of time,
 - critical attribute change management and control,
 - digital object identifiers and unique identifier management,
 - business rule exchange between parties,
 - and much more.

- Values can be constrained by other values. These include existence constraints and content constraints (i.e. value of one attribute is determined by the value of another attribute)

3. Business Rule Exchanges

- The business rules describing the CDS should flow with the CDS.
- This is possible through the exchange of business rules from the authoritative source.
- During any critical CDS information exchange, the CDS can be validated by the authoritative business rules.

4. The Governance of business rules, containers and values.

- Governance is the cornerstone of TRUST and must describe the existing rules used to determine validity of the data as well as the process used to manage changes to rules.
 - Data Quality and validation change over time.
 - Governance ensures the change life cycle maturity as well as its audit-ability.
- MISMO has a governance process that represents the current standard of TRUST.
- This process must be adapted to the declaration and management of the CDS.

5. UMDP / CDS DEVELOPMENT PROCESS POSSIBLE NEXT STEPS

Establish MISMO Support via a CDS workgroup to:

1. Collect the current descriptions of the CDS critical exchanges (ULDD, Loan Boarding, Servicing Transfer, Default Reporting and Investor Reporting Specifications)
2. Assess and incorporate any planned changes to the above critical exchanges.
3. Identify the entities and attributes that exist within all of the critical exchanges (intersection), thereafter the CDS V.01
4. Publish the CDS V.01 using the MISMO public commentary process, and include as identified:
 - The business and technical rules validating the CDS
 - The source of truth for each attribute in the CDS

6. OUT OF SCOPE

6.1 Definition of Operational Data

The data points associated with managing the core data set change often according to market, regulatory and policy influences. This white paper does not attempt to describe operational data.

6.2 Ownership of Operational Data

Ownership of the core data set does not imply ownership of operational characteristics related to managing the life cycle of the core data set. This ownership emphasizes certain responsibilities of other stakeholders relative to the product life cycle and key communications.

7. APPENDIX –DATA VALIDATION INNOVATION CHALLENGE TEAM MEMBERS

- Louis Abbott FNC, Inc.
- Sue Baker D+H Mortgagebot LLC
- Rod Bechdoldt Advest Mortgage Company
- Keith Beijer Mortgage Cadence / Accenture Credit Services
- Brett Benson CoreLogic, Inc.
- Tom Booker CoreLogic, Inc.
- Darius Borzogi Veros Real Estate Solutions
- Kris Bryant Xerox Corporation
- Charlie Burke Federal Home Loan Bank of Chicago
- Brian Camper Sierra Pacific Mortgage Company, Inc.
- Jim Connell Sierra Pacific Mortgage Company, Inc.
- Bill Corbet RBS Citizens Financial Group, Inc.
- Uday Devalla Stearns Lending, Inc.
- Patrick Elmore Mortgage Cadence / Accenture Credit Services
- Paco Fitzgerald Informatica Corporation
- Pebbles Folsom a la mode, inc.
- Rich Gagliano Lender Processing Services, Inc. (LPS)
- Randy Gillis Lender Processing Services, Inc. (LPS) – *Data Validation Thought Paper Lead*
- James Grizzle Informatica Corporation - *Data Validation Team Chair*
- Limin Hu Ellie Mae, Inc.
- Dave A. Hurt CoreLogic, Inc.
- Felice Kesselring CoreLogic, Inc.
- Christian Leask FNC, Inc.
- Ken Marlin Xerox Mortgage Services
- Dan Massey SunTrust Mortgage, Inc.
- Jennifer Miller a la mode, inc.
- Mike Mitchell FNC, Inc.
- Terry Mitchum a la mode, inc.
- Jonas Moe Ellie Mae, Inc.
- Monica Orluk Fiserv, Inc.
- Kurt Raymond DataVerify
- Craig Rebmann Lender Processing Services, Inc. (LPS)
- Ginger Ricchetti Regions Bank
- Liz Rudolph PHH Mortgage
- Parvesh Sahi Ellie Mae, Inc.
- Shawn Telford FNC, Inc.
- Lance Thoet DataVerify
- Greg Yeager Xerox Corporation