



Consumer  Mortgage Coalition



November 24, 2017

Monica Jackson  
Office of the Executive Secretary  
Consumer Financial Protection Bureau  
1700 G Street, NW  
Washington, DC 20552

**Re: Disclosure of Loan-Level HMDA Data; Docket No. CFPB-2017-0025**

Dear Ms. Jackson,

We appreciate the opportunity to provide comments on the Consumer Financial Protection Bureau's (CFPB or Bureau) proposed policy guidance, Disclosure of Loan-Level HMDA Data (Proposed Guidance).<sup>1</sup> Our associations, the American Bankers Association,<sup>2</sup> Consumer Bankers Association,<sup>3</sup> Consumer Mortgage Coalition,<sup>4</sup> Housing Policy Council of the Financial Services Roundtable,<sup>5</sup> and Mortgage Bankers Association,<sup>6</sup> represent the thousands of lenders who originate and purchase home loans in the U.S., and these lenders are committed to

---

<sup>1</sup> Disclosure of Loan-Level HMDA Data, 82 Fed. Reg. 44586 (Sept. 25, 2017).

<sup>2</sup> The American Bankers Association is the voice of the nation's \$17 trillion banking industry, which is composed of small, regional and large banks that together employ more than 2 million people, safeguard \$13 trillion in deposits and extend more than \$9 trillion in loans.

<sup>3</sup> The Consumer Bankers Association represents America's retail banks above \$10 billion in assets. We advance legislation and promote policies geared toward creating a stronger industry and economy. Established in 1919, CBA's corporate member institutions account for 1.6 million jobs in America, extend roughly \$3 trillion in consumer loans, and provide \$270 billion in small business loans. Follow us on Twitter @consumerbankers.

<sup>4</sup> The [Consumer Mortgage Coalition](http://www.consumermortgagecoalition.org), a mortgage industry trade association, dedicated to better-serving consumers by streamlining and improving the laws and regulations governing the mortgage industry.

<sup>5</sup> The Housing Policy Council (HPC) is a division of the Financial Services Roundtable (FSR). Our members are 31 of the leading national mortgage lenders, servicers, mortgage insurers, and title and data companies. HPC advocates for the mortgage and housing marketplace interests of its members in legislative, regulatory, and judicial forums. For additional information, visit: <http://www.fsroundtable.org/category/hpc/>.

<sup>6</sup> The Mortgage Bankers Association (MBA) is the national association representing the real estate finance industry, an industry that employs more than 280,000 people in virtually every community in the country. Its membership of over 2,300 companies includes all elements of real estate finance: mortgage companies, mortgage brokers, commercial banks, thrifts, REITs, Wall Street conduits, life insurance companies and others in the mortgage lending field.

ensuring fair treatment of consumers, equal access to credit, and a broad reach into all American communities.<sup>7</sup>

In the 2015 HMDA Final Rule, CFPB interpreted HMDA to require that CFPB use a balancing test to determine whether and how HMDA data should be modified prior to its public disclosure in order to protect applicant and borrower privacy while also fulfilling HMDA's disclosure purposes. The balancing test is whether the release of unmodified data creates risks to applicant and borrower privacy interests that are not justified by the benefits of such release to the public in light of the statutory purposes. The Proposed Guidance is CFPB's application of this balancing test. Under the Proposed Guidance, CFPB would exclude certain loan-level data points from disclosure, modify some data points prior to disclosure, and disclose other data points without modification.<sup>8</sup>

Consumers<sup>9</sup> provide a wide range of financial and personal data to our lender and vendor members – information that provides a clear financial and demographic profile of the borrower and the terms of the loans. As responsible keepers of this sensitive information, we feel compelled to articulate our profound concerns about the risks to consumer privacy, of identity theft, and fraud presented by the Proposed Guidance. Our concerns include:

- The CFPB has not engaged in the statutorily-required rulemaking process for modification of itemized data. To rectify this, CFPB must withdraw this Proposed Guidance and issue a formal Administrative Procedure Act (APA)-compliant rulemaking.
- The CFPB's balancing test fails to consider real threats to consumers, as re-identification under the Proposed Guidance would be a virtual certainty. Research has shown that re-identification is already highly possible, even using only the *current* public HMDA data. It becomes significantly more attainable with any increase in data disclosures lacking sufficient masking. With advances in technology enabling easier re-identification, this concern transforms from a risk to a virtual certainty. To guard against invasions of privacy, identity theft, and fraud, CFPB should disclose the new data only in aggregate form, carefully designed to protect the financial interests of individual consumers.
- Data security is of the utmost importance, and CFPB should take steps to ensure this sensitive data is protected, including providing a comprehensive update of its data security practices and taking measures to restrict access and ensure the data is properly protected.

---

<sup>7</sup> Our comments are representative of single-family HMDA reporters. Some of the undersigned trade associations may submit separate comments relative to multi-family HMDA reporters.

<sup>8</sup> See Appendix A for a detailed overview of the Proposed Guidance.

<sup>9</sup> While we use the term "consumers" throughout this letter, we recognize that some business purpose loans also are reportable under HMDA, and therefore some of these privacy concerns may also be applicable to businesses.

- The CFPB’s new data submission interface raises concerns about accessibility, functionality, data aggregation and storage, and liability. We have several suggestions for addressing these concerns, including providing for the ability to submit data incrementally and permitting the appending of earlier data.
- Finally, to improve the public understanding of why lenders ask for certain information, the obligation to submit that information to CFPB, and the statutory requirement for public disclosure we ask CFPB to engage in a public education initiative.

#### **I. The Dodd-Frank Act Requires CFPB to Use APA Rulemaking to Address the Privacy Issues Presented by Public Disclosure of the HMDA Data.**

The Dodd-Frank Act defines the terms under which CFPB is authorized to release to the public an expanded set of loan-level data. CFPB is required to “*develop regulations*” “*after notice and comment*” which: a) establish a method to evaluate whether the publication of specific data will violate consumers’ privacy interests (the balancing test) *and* b) describe how CFPB will “modify” any data that trigger this privacy breach. The letter and intent of the law is only satisfied by meeting both of these terms. The HMDA balancing test, as promulgated in the 2015 HMDA Final Rule, clearly fulfills the statutory obligation to establish a standard for evaluating the impact of disclosure of loan-level data. However, neither that regulation nor the proposed rule that preceded that Final Rule sets forth the method to mask the data fields deemed problematic. It is in this latter action, the *application* of the balancing test, that the impact of the Rule will be felt by all stakeholders, especially consumers. As modified by Dodd-Frank, the statutory requirement is to develop regulations “that modify or require modification”, not merely a conceptual framework for making that determination at a later date. It is this outcome that was the core concern for lawmakers, when mandating the rulemaking, not merely the means. The application of the test and the outcomes necessitate an APA-compliant rulemaking process.

Therefore, we disagree with CFPB’s conclusion that it has met the statutory requirement of issuing a rule regarding modification of itemized information. The fact that the Bureau has requested comment on the Proposed Guidance does not excuse the procedural shortcut: the requirements for APA rulemaking encompasses far more than notice and comment to promote fairness, transparency and fact-based decision-making.<sup>10</sup> Moreover, a final rule adopted through APA rulemaking has a degree of permanence that is not guaranteed with agency guidance.

---

<sup>10</sup> In part, the APA requires a notice of proposed rulemaking to include reference to the legal authority under which the rule is proposed and the substance of the proposal. There are other statutory requirements for a formal rulemaking as well, such as section 1022(b)(2) of Dodd-Frank, which requires the CFPB to conduct a cost-benefit analysis; the Regulatory Flexibility Analysis Act which requires an agency to consider the potential impact of its regulations on small entities, and the Paperwork Reduction Act, which requires an agency to seek the Office of Management and Budget approval for information collection requirements prior to implementation.

In fact, CFPB clearly states that the purpose of the Proposed Guidance is to “describe[] the loan-level HMDA data that the Bureau intends to make available to the public... including **modifications** that the Bureau intends to apply to the data.”<sup>11</sup> While the 2015 HMDA Final Rule articulated the balancing test, it was only through the Proposed Guidance, which sets forth the application of that standard, that they revealed the critical effect of “modify[ing]... itemized information.”<sup>12</sup> Simply stated, CFPB has not fulfilled Congress’ clear direction for a formal rulemaking process.

We believe that the Dodd-Frank Act further emphasizes the need for formal rulemaking in three significant ways.<sup>13</sup>

First, Dodd-Frank added the phrase “after notice and comment” to the rulemaking requirement, making it clear that Congress’s intent was to require CFPB to engage in an APA-compliant rulemaking.

Second, Dodd-Frank added that CFPB must issue a regulation, subject to notice and comment, which “modif[ies] or require[s] modification of itemized information... that is or will be available to the public.” Prior to the Dodd-Frank Act, this was not a requirement, and the Federal Reserve Board (Board) could make determinations about modifying HMDA information without a formal rulemaking. Knowing that this was the previously existing process, yet expressly adding this new language, Congress clearly intended a procedural change. CFPB is required “after notice and comment” to “develop regulations” regarding the modification of itemized information, for the purposes of protecting the privacy interests of consumers.

Third, Dodd-Frank imposed specific modification requirements itself, mandating modification of credit score data, age, and any other category of data the Bureau deems necessary to satisfy the purpose of protecting privacy interests.<sup>14</sup> Congress also required CFPB to “prescribe standards for any modification [of itemized information]... to effectuate the purposes of this chapter, in light of the privacy interests of mortgage applicants or mortgagors. Where necessary to protect the privacy interests of mortgage applicants or mortgagors, the Bureau shall provide for the disclosure of information... in aggregate or other reasonably modified form, in order to effectuate the purposes of this chapter.”<sup>15</sup>

Notably, CFPB also is required to engage in an APA rulemaking regarding the information that must be disclosed through individual institution loan application register (LAR) data. HMDA requires financial institutions that must disclose HMDA data to “make available to the public, upon request, loan application register information, *as defined by the Bureau by*

---

<sup>11</sup> Disclosure of Loan-Level HMDA Data, 82 Fed. Reg. 44586, 44611-12.

<sup>12</sup> 12 U.S.C. § 2803(h)(1)(E).

<sup>13</sup> See Appendix B for the relevant excerpts of HMDA, prior to and after the Dodd-Frank Act.

<sup>14</sup> 12 U.S.C. § 2803(h)(3).

<sup>15</sup> 12 U.S.C. § 2803(h)(3)(B).

*regulation.*<sup>16</sup> Further, CFPB must require, by regulation, such deletions as the Bureau may determine to be appropriate to protect any privacy interest of any applicant and to protect a depository institution from liability under federal or state privacy law.<sup>17</sup> CFPB has neither proposed a regulation nor addressed the data points that will be deleted from individual LARs through the Proposed Guidance.

For all of these reasons, we strongly urge CFPB to withdraw this Proposed Guidance and initiate a formal APA rulemaking regarding the modification of itemized information prior to publicizing the new HMDA data and the modification for LAR disclosure. These issues are too important to dispose of perfunctorily. The new data will not be disclosed until 2019. There is ample time to conduct this rulemaking properly, or if need be, delay public disclosure until privacy concerns are addressed. As part of the rulemaking process, the Bureau must demonstrate clearly that consumer interests will be protected adequately.

## **II. The Application of the Balancing Test Fails to Fully Consider the Real Threat to Consumers, and Aggregation of Certain Data would Better Protect Consumers.**

We are concerned that CFPB's balancing test as applied does not properly and fully account for significant potential risks to consumers that will result from this data being public. Aggregation of the new data that CFPB is proposing to make public would better protect consumers.

### *a. HMDA re-identification risk is real.*

Before the changes to the HMDA rule, the Board's economists recognized the ability to re-identify individual loan applicants and borrowers using publicly available loan-level HMDA data. Research in this area tells us that due to technological advances and the ability to scrape public data, it has become increasingly possible to either identify the mortgage transactions of individuals whose identity and property are known or to re-identify HMDA data by determining the names and addresses of the mortgage applicants and borrowers.

Assuming HMDA data available today is accurate, attaching a borrower's name and property address to HMDA data can be achieved in over 80% of all cases.<sup>18</sup> The addition of the new data fields required by the 2015 HMDA Final Rule raise the probability to virtually 100%. The combination of loan amount, house value, lender name, and census tract is sufficient to identify virtually every mortgage transaction, particularly when advances in computing and machine learning allow for rapid analysis and comparison of HMDA data fields, alone, in

---

<sup>16</sup> 12 U.S.C. § 2803(j)(1).

<sup>17</sup> 12 U.S.C. § 2803(j)(2)(B).

<sup>18</sup> See Anthony Yezer, *Personal Privacy of HMDA in a World of Big Data*, Institute for International Economic Policy Working Paper Series, IIEP-WP-2017-21 (2017) [hereinafter *Yezer Working Paper*].

conjunction with one another, and/or in conjunction with other publicly available data sets (such as real estate sales data and publicly recorded documents).<sup>19</sup>

Certain new data points the Bureau proposed to disclose will reduce privacy, as they are not otherwise available publicly. Other data points will reveal creditworthiness of applicants and borrowers. Information on house value, whether revealed directly or obtained as the quotient of loan amount and loan-to-value (LTV) ratio can be matched to property or appraisal records and will further increase identification rates.

The effects of the proposed additional HMDA data points on individual privacy are significant. If income and mortgage amount are disclosed along with the debt-to-income (DTI), the borrower's non-mortgage debt can be approximated. Even if the credit score is withheld as proposed, disclosure of additional information on the note rate, combined with other information such as the combined LTV and the ability to determine the payment-to-income ratio will allow observers to form a very accurate estimate of creditworthiness.<sup>20</sup>

*b. Re-identification risk invites adversaries that target unsuspecting consumers.*

It is clear that most consumers today can be identified through their HMDA data, should someone undertake the effort to do so. The question is whether it is appropriate to disclose even more sensitive information under these circumstances. As data matching techniques continue to improve, cross-referencing separate databases will become easier. Improved machine learning techniques will also expand opportunities to use this data. While this may have some benefits, it would be irresponsible to not also consider the negative implications.

For example, while CFPB has expressed an interest in protecting older Americans,<sup>21</sup> disclosure of detailed data puts those very same older Americans at greater risk. Individuals, with or without malicious intent, can make use of the HMDA data in connection with other publicly available data sets to identify targets and, using disclosed data, establish a level of trust with unsuspecting elders who may be uneducated about the depth and breadth of information disclosed through HMDA data.

Although the Bureau does not believe that the data it proposes to disclose could be used to open fraudulent accounts in the name of an individual consumer, the information invites all forms of predatory marketing. Re-identification aids in the discovery of credit scores, non-mortgage debt levels, and the identities of individuals susceptible to these forms of marketing. A private entity could potentially target applicants who were denied a home loan for possibly exploitative credit offers. And even if the disclosed data could not be used to open a fraudulent

---

<sup>19</sup> See Appendix C for a detailed discussion of the re-identification risk.

<sup>20</sup> Any proposal to reveal information on applicant credit scores would be inconsistent with the privacy protections granted under the Fair Credit Reporting Act, 15 U.S.C. § 1681 et seq.

<sup>21</sup> See CONSUMER FINANCIAL PROTECTION BUREAU, DRAFT CFPB STRATEGIC PLAN FY 2018 – 2022 (2017), p. 16-17.

account, it could be manipulated to be used as a tool to gain access to sensitive information about an individual consumer that could in turn be used to establish fraudulent accounts.

As discussed above, the increase in the types of data sought, and the form of data, combined with today's advances in technology, provide the potential to allow 100% re-identification of applicants and borrowers. Without proper masking, modification of the data, or sensitive handling of the information, many Americans would be placed at risk of having information made public that they would have otherwise assumed would remain private. With that information available to the public, unassuming Americans may become targets of scams or predatory actors or the victims of malicious parties.

- c. Public release of data should consider the privacy interests of not only individual borrowers or applicants, but also the broader community.*

CFPB's application of the balancing test appears to weigh the individual privacy interest of the applicant or borrower against a broad benefit to the public. However, the balancing test completely fails to account for the fact that aggregation of individual information creates a larger, more general public privacy risk. If each individual's nonpublic personal information is re-identifiable, the entire data set becomes more attractive to adversaries. While the privacy risk to a specific individual is a vital concern, attention should also be given to the wisdom of creating mechanisms to aggregate and transmit large data sets of sensitive information. It is clear that consolidation of information on the level provided under HMDA warrants additional masking, heightened security, and greater protections for consumers.

- d. Due to the risks to consumers, all new data that the Bureau has not proposed to restrict should be disclosed only in aggregate form.*

With re-identification a virtual certainty, the Proposed Guidance puts millions of consumers at risk for fraud and scams. This result conflicts with the intent of the balancing test and with a key objective of CFPB itself – ensuring that “consumers are protected from unfair, deceptive, or abusive acts and practices and from discrimination.”<sup>22</sup>

To better protect consumers while still serving the purposes of HMDA, all the new data fields financial institutions will begin collecting in 2018 should only be publicly disclosed at the aggregate level, not loan level. Specifically, the following data points should be disclosed only in aggregate:

- Regarding applicants and borrowers and the underwriting process:
  - Borrower or applicant age;
  - Name and version of the credit scoring model; and
  - Automated Underwriting System (AUS) name.
- Regarding the property securing the loan:

---

<sup>22</sup> 12 U.S.C. § 5511(b)(2).

- Property value;
- Manufactured housing secured property type;
- Manufactured housing land property interest; and
- Total units.
- Regarding the application or loan:
  - Application channel;
  - Whether the loan was initially payable to the financial institution;
  - Prepayment penalty term;
  - Introductory rate period;
  - Interest rate;
  - Total loan costs or total points and fees;
  - Origination charges;
  - Total discount points;
  - Lender credits;
  - Non-amortizing features (including balloon payment, interest-only payments, negative amortization);
  - DTI ratio;
  - Combined LTV ratio;
  - Open-end line of credit flag;
  - Business or commercial flag; and
  - Reverse mortgage flag.

This data should be aggregated at the Metropolitan Statistical Area (MSA) level. For numerical values, such as age, property value, DTI, and LTV, aggregated data points such as the mean, median, and average could be reported. For non-numerical data points, such as credit scoring model, AUS name, or reverse mortgage flag, those could be totaled by MSA. For example, the disclosed data would include how many applications were evaluated in an MSA using a certain credit scoring model. This would not affect the existing HMDA data that is disclosed at the loan level.

As CFPB will be administering the disclosure of individual LARs through its website, we believe that any data points that are disclosed in aggregate form should be deleted from the information disclosed in an individual institution's LAR before it is released.

Disclosure of the new data in the aggregate at the MSA level would meet the goals of HMDA and Regulation C while maintaining consumer privacy and protecting consumers from fraud. This aggregate data would: (1) help determine whether financial institutions are serving the housing needs of their communities; (2) assist public officials in distributing public-sector investment so as to attract private investment to areas where it is needed; and (3) assist in identifying possible discriminatory lending patterns and enforcing antidiscrimination statutes.<sup>23</sup>

---

<sup>23</sup> 12 C.F.R. § 1003.1(b)(1).

### III. Data Security is Critical for All Who Have Access to HMDA Data.

Data breaches in the private and public sector have resulted in millions of consumers having their privacy invaded and exposed to financial fraud. HMDA data contains confidential financial information. To protect consumers, it is the duty of all who handle HMDA data to maintain the highest level of data security. The industry and CFPB share the same goal of ensuring that consumers' sensitive personal information is kept secure.

To protect against data breach and help ensure public confidence in the security of this sensitive information, we ask CFPB to detail the types of general data security safeguards it will put in place. Additionally, we ask that CFPB provide a comprehensive update on its information security practices to Congress—perhaps in a closed session if there are concerns about publicly disclosing detailed information about security protocols.

As CFPB is aware, the agency's information security has been an ongoing area of concern. Just two months ago the Board's Office of Inspector General (OIG) cited "ensuring an effective information security program" as a major management challenge for CFPB.<sup>24</sup> The OIG has identified this issue as a major management challenge for CFPB for the last four years. The OIG correctly notes that "unauthorized access to or disclosure of sensitive information, through internal or external threats, could undermine the public's trust in the CFPB and limit its ability to accomplish its mission."<sup>25</sup> The OIG found that

The CFPB has not fully implemented processes, such as data loss prevention technologies, within its internal network that would enable the agency to detect and better protect against unauthorized access to and disclosure of its sensitive information. Likewise, the CFPB is in the process of implementing multifactor authentication for its internal system users.<sup>26</sup>

Data security is critical for all institutions that handle sensitive personal information, and we ask CFPB to provide these updates and details to help ensure this information remains secure.

Since the new data greatly expands the possibility of re-identification, as discussed above, access to any raw HMDA data must be restricted to protect consumers. It is the financial regulatory agencies who bear the ultimate responsibility for protecting that information. Access to the raw data should be limited to properly-accredited and duly authorized personnel, and this access should be carefully monitored and reviewed periodically to ensure that access is appropriate. Making the data accessible to the general public or even sectors of the general public who are not subject to the same data security mandates as government agencies may

---

<sup>24</sup> Memorandum from Mark Bialek, Inspector General, Federal Reserve Board of Governors and CFPB, to Director Cordray, CFPB (Sept. 27, 2017), <https://oig.federalreserve.gov/reports/cfpb-major-management-challenges-sep2017.pdf>.

<sup>25</sup> *Id.*

<sup>26</sup> *Id.*

significantly compromise consumers' financial well-being. Any third party permitted access to this raw data should be subject to stringent security protocols.

Finally, if any data breach were to occur at CFPB, we expect CFPB to commit to publicizing it promptly, both to notify the public and to underscore that it is through no fault of the financial institutions which submitted the data to CFPB. The same obligations that apply to other sectors handling consumer data should apply to the Bureau. When a data breach occurs, no matter what entity is breached, whether it be a government agency or a credit bureau, financial institutions often are implicated in such a breach. If the CFPB or any other entity is the cause of the breach, they should bear the responsibility and liability for all costs that result from the breach.

#### **IV. The CFPB's New Data Submission Interface Raises Concerns about Data Security as well as Accessibility, Functionality, Data Aggregation and Storage, and Potential Liability.**

While the privacy issues are paramount, we also wanted to raise some concerns with CFPB's new data submission interface (HMDA Platform).

##### *a. The CFPB's HMDA Platform introduces new vulnerabilities and risks.*

The recently released CFPB HMDA beta data submission interface raises concerns about accessibility, functionality, data aggregation and storage, and liability. Previously under the FFIEC software, lenders were capable of merging data from disparate software systems prior to submitting their final report. Under the new HMDA Platform, final submission must be made in a single file containing all the required data. While it is understood that for validation purposes data may be tested in sets, the requirement to consolidate data from multiple sources raises data security challenges and accessibility issues.

Consolidation of the required data will dramatically decrease the difficulty faced by a malicious actor trying to obtain otherwise secured data, while increasing the value of capturing the data. Each discrete system may have had different layers of protection and varied amounts of "desirable" data, reducing the ability and incentives to target them individually. In addition, the current functionality of the new HMDA Platform permits every user account created to retain identical authorizations. Each user account is capable of validating data and submitting finalized data. Combined with the lack of demarcation between validation and finalizing a submission, an increase in inadvertent errors that unfairly bind lenders is likely.

CFPB should properly test and assess the system for heavy traffic. Many lenders will transmit their data on the last day to file. Data submitted en masse may subject CFPB's servers to seizing under heavy load. This raises the potential for delays in the submission of data sets, which may inadvertently push a lender past the deadline. Responsibility for issues related to data

transmittal need to be addressed and testing should be performed to protect against unexpected downtime.

*b. Suggestions to improve the new HMDA Platform.*

Several suggestions have been already offered in regards to the issues raised above. Primarily, the ability to submit data incrementally and permitting the appending of earlier data would alleviate several issues. Categorizing user accounts with varying levels of authorization would provide for necessary oversight, avoiding accidental submission, enhancing training regimes, and allowing servicers to self-regulate their data submission practices.

While it is understood that the current practice is to accept a single file and the statute requires financial institutions to report data, the increase in the amount of data requested and the increasingly hazardous data security climate argues strongly for taking much needed precautions. In the past, Congress has taken steps to effectuate change to match developing and emerging technologies. As seen in recent legislation and hearings, data breaches that involve vast amounts of consumer information are a major concern. Dodd-Frank imbued the Bureau with the authority to promulgate rules in this area — to the degree of altering statutorily required HMDA reporting — in order to address the perpetually evolving technological landscape. The suggestions above are within the ability of CFPB to execute with the care necessary to protect the privacy of consumers and promote accurate reporting by lenders.

**V. Public Education is Needed.**

We have serious concerns that there is a lack of public understanding as to why financial institutions must ask for and collect this sensitive data, report it to CFPB, and have it publicly disclosed. As institutions begin to implement the new data collection requirements, our members are already hearing first hand questions from the public about why the institution is asking such detailed questions and what happens to the information. In addition, our members already are beginning to field complaints about intrusive questions seeking government monitoring information. Our members are the interface with consumers, and it is our members that also have to protect the privacy of the information entrusted to them by consumers and also surrender reams of consumer data to the government. While education regarding HMDA's implications for consumers has been a longstanding issue, it has become a top priority as: (1) the increasing number of data breaches is causing growing concern regarding maintaining and accessing sensitive information; (2) the new HMDA data to be collected and reported includes much more sensitive and personal information than ever before; and (3) there is an increasing and more sophisticated set of entities that compile and aggregate data for profit.

Currently, the HMDA portal provides a succinct explanation of the underlying authority of Regulation C and its importance. The explanation falls short of assuaging consumers concerns relative to the vast amount of information applicants and borrowers are required to provide and lenders are required to disclose. The nature of the disclosure itself, printed on the HMDA portal,

ABA, CBA, CMC, HPC, MBA  
November 24, 2017

only lends itself to consumers who actively seek out the portal on the Bureau website. Consumer concerns typically arise during the application process, providing information in the proper form prior to that sensitive moment is of great importance.

Furthermore, our members stress to us the need to train employees carefully to respond to consumer anxiety about the data that is required to be collected during the mortgage origination process. As associations, we acknowledge that we also must help our members with questions about the level of information collected and data being transmitted. It is not infrequent that customers ask that their data not be included or they search for ways to prevent personal information from being shared with third parties. Therefore, we ask that CFPB consult with the industry and engage in educational efforts regarding the purposes and requirements of HMDA. We and our members look forward to assisting these efforts.

We look forward to working with the Bureau as it moves forward with this important issue. If you have any questions, please contact Meg Burns at [meg.burns@fsroundtable.org](mailto:meg.burns@fsroundtable.org) or 202-289-4322.

Respectfully,

American Bankers Association  
Consumer Bankers Association  
Consumer Mortgage Coalition  
Housing Policy Council, Financial Services Roundtable  
Mortgage Bankers Association

## **Appendix A: Overview of Proposed Guidance**

### **HMDA, as implemented by Regulation C, and its purpose.**

HMDA requires certain financial institutions to collect, report, and disclose data about their mortgage lending activity on an ongoing basis to federal regulators and through the government to the general public.<sup>27</sup> As implemented by Regulation C, the purpose of HMDA is three-fold: (1) provide the public and government officials with sufficient information to enable them to determine whether financial institutions are serving the housing needs of the communities; (2) assist public officials in their determination of the distribution of public sector investments in a manner designed to improve the private investment environment; and (3) assist in identifying possible discriminatory lending patterns and enforcing antidiscrimination statutes.<sup>28</sup>

### **The balancing test under Regulation C.**

The 2015 HMDA Final Rule requires the Bureau to use a balancing test to determine whether and how HMDA data should be modified prior to its disclosure to the public in order to protect applicant and borrower privacy while also fulfilling HMDA's public disclosure purposes.

CFPB has asserted that in applying the balancing test, the risk to the applicant and borrower privacy interests arise only where at least one data field or a combination of data fields in the data set substantially facilitates the identification of an applicant or borrower, and at least one data field or combination of data fields discloses information about the applicant or borrower that is not otherwise public and may be harmful or sensitive.<sup>29</sup> The Bureau has further interpreted HMDA to require that data be modified only when the release of unmodified data creates risks to applicant and borrower privacy interests that are not justified by the benefits of such release to the public in light of the statutory purposes.<sup>30</sup>

### **HMDA data to be modified by exclusion.**

In light of these interpretations, CFPB proposes to modify the public loan-level HMDA data to exclude entirely:

- the universal loan identifier (ULI);
- the date of the application;
- the date action was taken by the financial institution;
- the address of the property securing the loan;

---

<sup>27</sup> Disclosure of Loan-Level HMDA Data, 82 Fed. Reg. at 44586.

<sup>28</sup> 12 CFR § 1003.1(b).

<sup>29</sup> Home Mortgage Disclosure (Regulation C), 80 Fed. Reg. 66128, 66134 (Oct. 28, 2015) (to be codified at 12 C.F.R. pt. 1003).

<sup>30</sup> *Id.*

- the credit score or scores relied on in making the credit decision;
- the Nationwide Mortgage Licensing System and Registry Identifier (NMLS ID);
- the result generated by the underwriting system; and
- free-form text fields used to report: applicant or borrower race and ethnicity, name and version of credit scoring model used, principal reason for denial (if applicable), and the AUS name.<sup>31</sup>

Having determined that the disclosure of an applicant or borrower's credit would likely reveal information about the applicant or borrower that is not otherwise public and may be harmful or sensitive, the Bureau's proposal excludes the credit score from disclosure as the risk would not be justified by the benefits of the disclosure. CFPB also had determined that modification strategies would not appropriately reduce the risk of harm or sensitivity. The proposal intends to exclude the ULI until information is available concerning how financial institutions use ULIs other than for HMDA purposes. It is believed that disclosure of the ULI would likely substantially facilitate the re-identification of an applicant or borrower, but CFPB purports that it is unlikely information disclosed would be harmful or sensitive.

The Bureau does not believe the application date to be inherently sensitive, but believes disclosure would likely substantially facilitate re-identification. The date an action is taken on the application was also found to likely substantially facilitate re-identification, though it is unlikely information disclosed would be harmful or sensitive.

The proposal notes that the disclosure of the address of the property securing the loan would likely substantially facilitate re-identification, but believes it to be unlikely that information disclosed would be harmful or sensitive. Disclosure of the NMLS ID would likely result in substantially facilitating the re-identification of an applicant or borrower. Disclosing the AUS result field would likely disclose information about the applicant or borrower that is not otherwise public and may be harmful or sensitive and this risk is not justified by the benefits of disclosure. With respect to free-form text fields, the proposal argues that due to the volume of the data it is not feasible for CFPB to review the individual contents of each field and thus they will not be disclosed.

#### **HMDA data that raises a risk, but warrants modification to provide a public benefit.**

CFPB acknowledges that disclosure of the following data would substantially facilitate re-identification and/or disclose information about the applicant or borrower that is not otherwise public and may be harmful or sensitive, but believes that this risk is justified by the benefits of disclosure warranting modification of the data. The data fields to be modified include:

- Loan amount;

---

<sup>31</sup> Disclosure of Loan-Level HMDA Data, 82 Fed. Reg. at 44586, 44587.

- Age of the applicant or borrower;
- DTI ratio; and
- Property value.<sup>32</sup>

The loan amount data field will be modified by disclosing the mid-point for the \$10,000 interval into which the reported loan amount falls and by indicating whether the loan amount exceeds the applicable GSE loan limit.<sup>33</sup> CFPB determined that while disclosure of the loan amount may substantially facilitate re-identification, the risk was justified by the benefits of disclosure.<sup>34</sup> The Bureau also asserts in its Proposed Guidance that property value will be similarly modified, as discussed below; therefore it is believed that there is minimal risk that adversaries will be able to use the combined LTV ratio to derive the loan amount.

Disclosure of the applicant or borrower's age would likely reveal information about the applicant or borrower that is not otherwise public and may be harmful or sensitive, but CFPB determined that this risk is justified by the benefits of disclosure.<sup>35</sup> Modification of this data field will be achieved by binning and top- and bottom-coding age. Specifically, the Bureau intends to bin reported values into the following ranges, as applicable: 25 to 34; 35 to 44; 45 to 54; 55 to 64; and 65 to 74; bottom-code values under 25; top-code values over 74; and indicate whether the reported value is 62 or higher.<sup>36</sup>

Disclosure of the DTI ratio relied on in making the credit decision would likely disclose information that may be harmful or sensitive and that, for certain ratios, the risk would not be justified by the benefits of disclosure.<sup>37</sup> However, CFPB believes that modification resulting in binning and top- and bottom-coding certain DTI values will sufficiently mask the information to justify the risk. Specifically, the Bureau intends to bin reported values into the following ranges, as applicable: 20% to less than 30%; 30% to less than 40%; and 50% to less than 60%; bottom-code values under 20%; top-code values of 60% or higher; and disclose, without modification, values greater than or equal to 40% and less than 50%.<sup>38</sup>

As mentioned above, property value would likely substantially facilitate the re-identification of an applicant or borrower and therefore requires modification. Modification will result in disclosure of the midpoint for the \$10,000 interval into which the reported property value falls.<sup>39</sup>

---

<sup>32</sup> *Id.*

<sup>33</sup> *Id.* at 44601.

<sup>34</sup> *Id.*

<sup>35</sup> *Id.* at 44604.

<sup>36</sup> *Id.* at 44612.

<sup>37</sup> *Id.* at 44606.

<sup>38</sup> *Id.* at 44612.

<sup>39</sup> *Id.* at 44607.

**HMDA data that will be disclosed to the public, unmodified.**

CFPB proposes to disclose the following data fields to the public without modification.

- Regarding applicants and borrowers and the underwriting process:
  - Income;
  - Sex;
  - Race;
  - Ethnicity;
  - Name and version of the credit scoring model;
  - Reasons for denial; and
  - AUS name.
- Regarding the property securing the loan:
  - Census tract;
  - State;
  - County;
  - Occupancy type;
  - Construction method;
  - Manufactured housing secured property type;
  - Manufactured housing land property interest; and
  - Total units.
- Regarding the application or loan:
  - Loan term;
  - Type;
  - Purpose;
  - Application channel;
  - Whether the loan was initially payable to the financial institution;
  - Whether a preapproval was requested;
  - Action taken;
  - Type of purchaser;
  - Lien status;
  - Prepayment penalty term;
  - Introductory rate period;
  - Interest rate;
  - Rate spread;
  - Total loan costs or total points and fees;
  - Origination charges;
  - Total discount points;
  - Lender credits;
  - HOEPA status;
  - Non-amortizing features (including balloon payment, interest-only payments, negative amortization);
  - Combined LTV ratio;
  - Open-end line of credit flag;

- Business or commercial flag; and
- Reverse mortgage flag.
- Regarding the lender:
  - Legal Entity Identifier (LEI); and
  - Financial institution name.<sup>40</sup>

CFPB asserts that disclosure would likely present low risk to applicant and borrower privacy for the data fields currently disclosed to the public without modification and those proposed to be disclosed without modification.<sup>41</sup> However, the Bureau has acknowledged that the LEI, financial institution name, census tract, income, action taken (where the loan is denied), and reasons for denial pose a risk to applicant or borrower privacy.<sup>42</sup>

CFPB believes that disclosure of LEI, financial institution, and census tract is likely to substantially facilitate the re-identification of applicants or borrowers. The disclosure of income, action taken (where the loan is denied), and reason for denial, would also likely create a risk of harm or sensitivity if HMDA data were re-identified. In light of these concerns, however, the Bureau believes that these risks are justified by the benefits of disclosure in light of HMDA's purposes.<sup>43</sup>

---

<sup>40</sup> Disclosure of Loan-Level HMDA Data, 82 Fed. Reg. at 44586, 44597.

<sup>41</sup> *Id.*

<sup>42</sup> *Id.* at 44597-98.

<sup>43</sup> *Id.* at 44598.

**Appendix B: Excerpt of HMDA Statute Prior to and After Dodd-Frank**

**12 U.S.C. § 2803(h) prior to the Dodd-Frank Act** (emphasis added)

"(h) **SUBMISSION TO AGENCIES.**—The data required to be disclosed under subsection (b)(4) shall be submitted to the appropriate agency for each institution reporting under this title. Notwithstanding the requirement of section 304(a)(2)(A) for disclosure by census tract, the Board, in cooperation with other appropriate regulators, including—

"(1) the Comptroller of the Currency for national banks;

"(2) the Director of the Office of Thrift Supervision for savings associations;

"(3) the Federal Deposit Insurance Corporation for banks insured by the Federal Deposit Insurance Corporation (other than members of the Federal Reserve System), mutual savings banks, and any other depository institution described in section 303(2)(A) which is not otherwise referred to in this paragraph;

"(4) the National Credit Union Administration Board for credit unions; and

"(5) the Secretary of Housing and Urban Development for other lending institutions not regulated by the agencies referred to in paragraphs (1) through (4),

**shall develop regulations** prescribing the format for such disclosures, the method for submission of the data to the appropriate regulatory agency, and the **procedures for disclosing the information to the public.** These regulations shall also require the collection of data required to be disclosed under subsection (b)(4) with respect to loans sold by each institution reporting under this title, and, in addition, shall require disclosure of the class of the purchaser of such loans. Any reporting institution may submit in writing to the appropriate agency such additional data or explanations as it deems relevant to the decision to originate or purchase mortgage loans."

**12 U.S.C. § 2803(h)(1) after the Dodd-Frank Act** (emphasis added)

**(h) SUBMISSION TO AGENCIES**

**(1) IN GENERAL** The data required to be disclosed under subsection (b) shall be submitted to the Bureau or to the appropriate agency for the institution reporting under this chapter, in accordance with rules prescribed by the Bureau. Notwithstanding the requirement of subsection (a)(2)(A) for disclosure by census tract, the Bureau, in consultation with other appropriate agencies described in paragraph (2) and, **after notice and comment, shall develop regulations that—**

(A) prescribe the format for such disclosures, the method for submission of the data to the appropriate agency, and the procedures for disclosing the information to the public;

(B) require the collection of data required to be disclosed under subsection (b) with respect to loans sold by each institution reporting under this chapter;

(C) require disclosure of the class of the purchaser of such loans;

(D) permit any reporting institution to submit in writing to the Bureau or to the appropriate agency such additional data or explanations as it deems relevant to the decision to originate or purchase mortgage loans; and

**(E) modify or require modification of itemized information, for the purpose of protecting the privacy interests of the mortgage applicants or mortgagors, that is or will be available to the public.**

...

### **(3) RULES FOR MODIFICATIONS UNDER PARAGRAPH (1)**

**(A) Application** A modification under paragraph (1)(E) shall apply to information concerning—

(i) credit score data described in subsection (b)(6)(I), in a manner that is consistent with the purpose described in paragraph (1)(E); and

(ii) age or any other category of data described in paragraph (5) or (6) of subsection (b), as the Bureau determines to be necessary to satisfy the purpose described in paragraph (1)(E), and in a manner consistent with that purpose.

**(B) Standards** The Bureau shall prescribe standards for any modification under paragraph (1)(E) to effectuate the purposes of this chapter, in light of the privacy interests of mortgage applicants or mortgagors. Where necessary to protect the privacy interests of mortgage applicants or mortgagors, the Bureau shall provide for the disclosure of information described in subparagraph (A) in aggregate or other reasonably modified form, in order to effectuate the purposes of this chapter.

### **Appendix C: Re-Identification Risk**

CFPB's current position, as it relates to the currently available HMDA data, is that individual privacy is preserved. However, published research has established methods for linking names and addresses of individuals to their mortgage transactions disclosed to the public in HMDA data. This research conclusively demonstrates that more than 80% of borrowers can potentially be re-identified under old HMDA data disclosures. This is particularly true for minorities, where the possibility of re-identification is even higher. Under the new HMDA data disclosures, these rates would potentially rise to nearly 100%.

Information on property transactions necessary to record ownership and deeds is now available online; this data is searchable, scrapable and contains names of those involved in the transaction, loan amounts, and property addresses. Research in the mid-90s was capable of a 52% success rate in linking conventional loans to old HMDA data, using the quality of data and programming techniques known at the time.<sup>44</sup> A re-identification rate of 68% was reported by matching property records from the Cuyahoga County Recorder with HMDA data for several years.<sup>45</sup>

In 2011 researchers performed a massive 27 million loan match of HMDA data to available data from lenders and noted that unique matching is more difficult when factors raise the volume of lending by a given lender in a particular census tract.<sup>46</sup> The inverse relationship between lender volume and re-identification probability is easily understood. As a lender produces more volume in a concentrated area, re-identification decreases. However, most lenders reporting under HMDA are small. Because smaller lenders typically make loans in only a few counties with smaller census tracts, the number of mortgages per census tract becomes more concentrated which inevitably increases the re-identification rate. However, even in the most challenging scenario for re-identification, such as a lender like a large depository where there are as many as 20 or 25 purchase mortgages in a given tract, the differential effect of borrower race and ethnicity can easily increase the re-identification rate.<sup>47</sup>

Property records in Massachusetts matched with HMDA data is especially notable for a number of reasons. The match rate from property records to HMDA data was 60% for every year between 1998 and 2006; while the match rate from HMDA data to property records was 70% in

---

<sup>44</sup> See Yezer Working Paper, *supra* note 13, at 11; see also, Anthony Pennington-Cross and Joseph Nicols, *Credit History and the FHA-Conventional Choice*, 28(2) *Real Estate Economics* 307-336 (2000).

<sup>45</sup> See Yezer Working Paper, *supra* note 13, at 12; see also, Claudia Coulton et al., *Pathways to Foreclosure: A Longitudinal Study of Mortgage Loans, Cleveland and Cuyahoga County, 2005-2008*, Center on Urban Poverty and Community Development, Mandel School of Applied Social Sciences, Case Western Reserve University (2008).

<sup>46</sup> See Yezer Working Paper, *supra* note 13, at 11; see also, Debbie Gurnestain Bocian et al., *Lost Ground, 2011: Disparities in Mortgage Lending and Foreclosures*, Center for Responsible Lending (2011).

<sup>47</sup> See Yezer Working Paper, *supra* note 13, at 17-18.

1998, rising to 75% in 2001.<sup>48</sup> The difference is notable, as property records include all transactions involving liens on real estate and illustrate how HMDA data can be paired with this information to re-identify most borrowers.

In each example, re-identification rates were achieved using computer algorithms that matched based on a limited range of information because the research purpose was to assemble large data sets that included HMDA information to supplement other data. With advances in machine learning and data science, more precise algorithms will raise match rates by identifying for issues, such as differentiating between “2<sup>nd</sup> St” and “Second St.”<sup>49</sup>

The 75% match rate mentioned above should be considered a lower bound, as it was achieved using only a fraction of the data points available in the old HMDA data. Furthermore, algorithms for matching surnames to ethnic or racial identities are capable of potentially raising match rates to 100%. HMDA has a full set of racial and ethnic identifiers. Simply based on the differing sizes of these ethnic and racial groups, the difficulty of matching may be reduced by nearly 94% where a certain group only represents 6% of the responding population.

In other words, the addition of this new set of data into the public domain significantly exacerbates a pre-existing problem, making it far easier for predatory actors to verify not only basic identity, but also the financial profile of an individual. This problem exposes everyone, but particularly the most vulnerable segments of the population.<sup>50</sup> With the financial condition of citizens now on public display, scam artists can easily tailor sham products and services to appear to meet the needs of these consumers. In addition, these new data factors will increase the susceptibility of applicants and borrowers to malicious actors who may make use of the information to prey on individuals by gaining their confidence with detailed knowledge of their loan terms and financial condition. Furthermore, providing additional detail on the cost of credit, in the form of interest rate and discount points paid, along with LTV and DTI ratios allows imputation of the credit worthiness of the borrower.

---

<sup>48</sup> See Yezer Working Paper, *supra* note 13, at 12; see also, Kristopher S. Gerardi and Paul S. Willen, *Subprime Mortgages, Foreclosures, and Urban Neighborhoods*, (Federal Reserve Bank of Boston, Public Policy Discussion Paper No. 08-06, 2008).

<sup>49</sup> See Yezer Working Paper, *supra* note 13, at 13.

<sup>50</sup> Small geographic areas can product significant re-identification risk, see, e.g., Khaled El Emam et al., *A method for managing re-identification risk from small geographic areas in Canada*, BMC Med Informatics and Decision Making (2010).