Americans for Financial Reform
Association for Neighborhood and Housing Development
Center for Responsible Lending
Consumer Action
Empire Justice Center
National Community Reinvestment Coalition
National Consumer Law Center (on behalf of its low-income clients)
National Coalition for Asian Pacific American Community Development
New Economy Project
National Fair Housing Alliance
Ohio Fair Lending Coalition
Reinvestment Partners
Western New York Law Center
Woodstock Institute

Ren Essene
Manager, Policy Section of the Chief Data Office
Consumer Financial Protection Bureau
1275 First St. N.E.
Washington, D.C. 20002

Re: Public disclosure of new HMDA data points

Dear Ms. Essene:

Thank you for the opportunity to provide feedback on how the Consumer Financial Protection Bureau (CFPB) should balance the compelling public interest underlying Home Mortgage Disclosure Act (HMDA) data disclosure with privacy concerns.

The undersigned groups hereby urge the CFPB to develop a comprehensive policy for meaningful public disclosure of HMDA data. Public disclosure of HMDA data is critically needed to expose housing and lending discrimination, to assess whether lenders are meeting the housing needs of local communities and to inform public policy and investment decisions. Data that the CFPB requires mortgage lenders to collect and report under the new HMDA rules will allow much richer and more accurate analyses of pressing questions about lending in the U.S., but only if the data are made public.

We appreciate that the CFPB must balance the public's need for comprehensive HMDA data disclosure with the risk of releasing previously-unobtainable data that might violate loan applicants' privacy. Most of the new HMDA data points, however, are already available for purchase or in publicly-available data sets. Moreover, the CFPB can take straightforward steps to mask or otherwise limit access to sensitive data points. Therefore, the CFPB can and should make most new HMDA data public at the loan-applicant level.

Congress enacted HMDA in 1975 to provide the public with information to help assess whether lending institutions are serving the housing needs of their communities, to help public sector agencies identify areas in need of public investment which would then attract private sector investment, and to identify possible discrimination.[1] In 2010, Congress passed the Dodd Frank Wall Street Reform and Consumer Protection Act (the Dodd-Frank Act), which made critical updates to HMDA, including expanding the data lenders are required to report and authorizing the CFPB to require additional reporting. The Dodd-Frank Act also directed the CFPB to assess whether any new data elements should be aggregated or otherwise reasonably modified before being made public to protect loan applicants' privacy interests while also furthering HMDA's purposes.

In its recent HMDA rulemaking, the CFPB took important steps to ensure that HMDA data will include all data points necessary to fulfill HMDA's purposes in the current mortgage lending landscape. This letter explains the substantial benefits and minimal risks of the CFPB's making most of the new HMDA data points public.

The CFPB is well aware of the benefits of publically available HMDA data. The undersigned groups support improvements to HMDA including public data availability that will help to better shine a light on discrimination and promote community reinvestment. Recently, NCRC and several of its member organizations negotiated public benefit agreements with KeyBank and Huntington National Bank during their acquisitions of other banks. NCRC and its members used HMDA data to document each bank's lending patterns and negotiate with the acquiring banks to ensure that the combined banks' home mortgage lending would not be cut back after the acquisitions. As a result, KeyBank and Huntington National Bank committed to provide more than $32 billion in loans and investments over a multi-year time period after the mergers. These commitments would not have been possible without public HMDA data. Using HMDA data ensured that both community organizations and the banks were working from the same facts, which facilitated negotiations aimed at a mutually beneficial outcome for banks and communities.

Another illustration of HMDA's benefits is the regular use of HMDA data in the fair housing planning process. Jurisdictions receiving HUD funding for affordable housing and community development initiatives are required to develop Assessment of Fair Housing plans that identify racial disparities in access to wide array of opportunities including employment and housing. Access to credit can critically affect a neighborhood residents' ability to secure decent housing and opportunities for employment. Accordingly, HMDA data analysis can be used in the AFH plans to identify communities in need of additional public and private sector investment as a means of improving the affordable housing stock and economic condition of communities.

HMDA data have also proven central to fair lending enforcement. The CFPB and the Department of Justice have recently concluded redlining and price discrimination settlements in which HMDA data were key to the detection of the discrimination. With the new Dodd-Frank data elements, HMDA data will become more effective in combating discriminatory lending by

[1] http://www.ffiec.gov/hmda/history.htm

enabling improved documentation of disparities in pricing and in loan terms and conditions among equally qualified applicants who only differ by their race, ethnicity, age, or gender.

The CFPB has acknowledged and reinforced the public benefits of HMDA data in its comprehensive rule by significantly increasing data reported by lenders on borrower characteristics and loan terms and conditions. At the same time, the CFPB has stated that it will decide how and whether to publicly disseminate the new data elements by implementing a balancing test. This test states that, "considering the public disclosure of HMDA data as a whole, applicant and borrower privacy interests arise under the balancing test only where the disclosure of HMDA data may both substantially facilitate the identification of an applicant or borrower in the data and disclose information about the applicant or borrower that is not otherwise public and may be harmful or sensitive."

It is our contention that most of the new HMDA data fields can be made public in a way that will not substantially facilitate the identification of applicants beyond what is already possible in the existing public HMDA data. Moreover, the new HMDA data will provide very little information that is not otherwise readily available from private sector entities, data brokers, or in marketing lists from credit bureaus. In fact, HMDA is unlikely to be useful as a data source for marketers or data brokers because there is a lag of more than a year between lending decisions and when the HMDA data on applicants become publicly available and it is time-consuming and difficult to connect HMDA data to names and addresses in a systematic way. Actors seeking to invade applicant privacy or market harmful products are much more likely to turn to other easier-to-use and timelier data sources that reveal lending decisions or other useful marketing information.

HMDA is not well suited as a mass identification data source because it is very difficult to identify particular borrowers using HMDA. The HMDA database has no personally identifiable information such as street address or Social Security numbers. Thus, marketers will need to use other information they obtained on applicants in order to identify them with certainty in the HMDA data. For example, if they know that a family of a particular race bought a house in a particular neighborhood, they would need to sort through HMDA data for the census tract representing that neighborhood to identify borrowers of the particular race. For a number of tracts, there might be several borrowers of the particular race, rendering HMDA data cumbersome for identifying individual borrowers. In other census tracts, there might fewer borrowers of a particular race, enabling a marketer to potentially identify an individual borrower. However, HMDA data in this case can only be used to identify relatively few individuals, not large numbers of people. If a marketer wanted to mass market harmful products based on loan rejections or loan terms and conditions to large numbers of recent borrowers, they would find HMDA data's ability to identify relatively few people with a time lag not useful.

Existing Data Sources Similar to HMDA

In contrast to HMDA data, a number of readily available databases from the private and public sectors can be used more effectively to identify and market to specific individuals on a large scale. Although we do not condone the use of this data for marketing or related purposes, the fact is that the data are already widely available. Therefore, making the new HMDA data fields public does not cause any further harm. Some examples of these databases include:

Realtytrac

Realtytrac is a private sector real estate data service that offers a wide variety of data on neighborhood housing stock and price trends.[2] Users can view data for individual homes or summary data on a zip code level for free, for a limited period trial, or by purchasing data from Realtytrac. NCRC obtained sample data from Realtytrac in excel format that costs about 8 cents per record. The data includes detailed information about homeowners including name, street address, loan characteristics, and property characteristics.

The full list of variables available in Realtytrac for each homeowner is the following:

- Name of borrower
- Whether property is in foreclosure
- Which bank owns the foreclosed property
- Street address, city, state, zip code, county
- # Bedrooms, # bathrooms, square footage
- Lot size
- Year of construction
- Purchase date
- Purchase amount
- Number of open loans
- Total amount of open loans
- Estimated market value
- Loan to value ratio
- Equity
- Loan date
- Loan amount
- Loan type – conventional or non-conventional
- Loan interest rate
- Lender name
- Lender name for loan two
- Bankruptcy

Legitimate marketers and scammers can easily obtain Realtytrac data that includes real-time information on mortgage lending with many of the same loan term and property fields as HMDA data. Variables about foreclosure status, bankruptcy, loan-to-value ratio, amount of total loans (first and subordinate liens) can enable marketers to readily identify borrowers in distress and target them for abusive refinance loans or loan modification or foreclosure scams. Realtytrac data exceeds information about properties provided in HMDA data and is therefore more useful to marketers.

[2] http://www.realtytrac.com/

Public Land Records

Cities and counties also maintain recorder of deeds records that can be obtained either in-person or online. The District of Columbia recorder of deeds, for example, enables a user to establish an account for free and then type in names of property owners in the District.[3] The user can then visually inspect several documents including the original mortgage, the deed, and notices of foreclosure or deed-in lieu. This information can enable marketers to identify borrowers in distress and market harmful products to them. The information can also enable marketers to find other borrowers that are not in distress but due to the length of time in their house might be vulnerable to offers of abusive, equity-stripping refinance or home equity loans. New York City, Massachusetts counties, and many other cities, counties and towns have similar online land record systems.

Credit Bureau Prescreening and Marketing Reports

The nationwide consumer reporting agencies (credit bureaus) sell information derived from people's credit and other consumer reports. Credit bureaus collect and hold data on most adults in the United States, and have aggressively marketed their data products to lenders, insurance companies, and even employers. Credit bureaus sell prescreening reports that include or expose data similar to the new data fields that will be reported under HMDA.

As the CFPB knows, prescreening reports are detailed marketing lists meant for use by lenders and insurance companies that intend to make firm offers of credit or insurance to the people listed in the report. These reports include individuals' names and addresses and are available in lists broken out by credit score, other credit history data, including delinquencies or negative information, and/or personal characteristics. Equifax advertises that their prescreening reports can be customized using 1,500 credit-related data points.[4] Although prescreening reports are available only to lenders and insurance companies making firm offers of credit or insurance, many predatory lenders and scam companies fall into that category and can easily purchase information on individuals segmented by credit score or credit history for marketing.

Credit bureaus also sell summary information on household, census block, census tract and zip code levels that can enable predatory marketing. Equifax, for example, sells aggregated FICO scores on a zip + 4 code level that enables financial institutions to market products targeted to customers with a specified credit risk profile. Zip + 4 codes are small areas consisting of approximately seven to ten households.[5] Credit score data for zip + 4 units is available for sale for less than two cents per record, and data sellers will match the aggregate credit score data with names and addresses. People living in zip codes where residents have low aggregate FICO scores can be easily targeted by marketers of risky and abusive products.

---

[3] https://gov.propertyinfo.com/DC-Washington/
[4] https://www.equifax.com/business/prescreen
[5] See http://www.equifax.com/assets/USCIS/aggregated_FICO_scores_ps.pdf and https://www.teamupturn.com/static/files/Knowing_the_Score_Oct_2014_v1_1.pdf

The credit bureaus advertise that these zip + 4 products are not subject to the Fair Credit Reporting Act (FCRA) because they use aggregate rather than individual data, but we believe that these products should be subject to FCRA and treated as prescreening reports. Even if, however, access to zip + 4 data products is restricted to lenders and insurers eligible to purchase prescreening reports, subprime lenders and other finance companies seeking to target people with poor credit for high-cost loans would still have full access to this low-cost and easily available data.

Data Brokers

Financial institutions and other marketers can also purchase the services of data brokers that obtain detailed data from a variety of sources, including combing the internet for information on people's purchasing habits and hobbies. The Federal Trade Commission (FTC) in a report called, *Data Brokers: A Call for Transparency and Accountability*, documents that sophisticated brokers can help their clients target people for marketing pitches. Data brokers provide their clients with detailed information on customers including age, gender, net worth, real property attributes, household income, and credit card usage. The brokers conduct data analysis to segment customers into groups with names like "Underbanked" or "Financially Challenged." These names suggest that the marketing pitches may take advantage of people's financial vulnerabilities. The data brokers can also group consumers into buckets based on estimates of their creditworthiness or utilization of bankcards. Finally, the FTC suggests that the data brokers do not monitor their clients' use of the data carefully, merely asking their clients to read the terms and conditions on their websites.[6]

Balancing Data Disclosure and Loan Applicants' Privacy

As the above examples show, currently available data sources from the private sector include comprehensive financial information about most Americans and offer information that is timelier and more readily available than HMDA data. HMDA data would be an inferior data source for predatory lenders, scam companies or others engaging in abusive marketing. In our groups' experience, we have not heard accounts of HMDA data being used for widespread predatory marketing. Moreover, we have asked staff from the Federal Reserve Board and CFPB if they were aware of any unsavory use of HMDA data. They have yet to provide us with any occurrences of concrete harms associated with inappropriate use of HMDA data. Nevertheless, to decrease the small chances of harmful use of HMDA, we offer the following recommendations regarding how to disclose the new variables mandated by the Dodd-Frank Act, which the CFPB and others may consider sensitive in terms of privacy.

Although most new data fields reported under HMDA do not pose privacy risks under the CFPB's balancing test, and should be reported as collected at the loan level, for a small set of variables we recommend that the CFPB use proven privacy-enhancing data manipulation techniques to further safeguard loan applicants' personal data when making loan-level data

[6] Federal Trade Commission, Data Brokers: A Call for Transparency and Accountability, May 2014, via https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

public. In particular, the CFPB should explore using a combination of data masking, data swapping and differential privacy techniques, largely based on those used by the Census in its Public Use Microdata Sample (PUMS). PUMS data includes detailed income, occupation, education, housing and other data at relatively small geographies, mirroring HMDA data disclosure quite closely.[7] The HMDA data that the CFPB requires lenders to report under the new rule include a small number of new fields that may raise privacy concerns, either alone or in combination with other fields. The fields include: universal loan ID, address, detailed nationality, age, credit score, and debt-to-income ratio. Each field requires a different approach to best balance preserving HMDA data's utility in fulfilling HMDA's statutory purposes with loan applicants' privacy interests.

In addition to the data safeguards outlined below for each field, the CFPB should consider using data swapping to protect the privacy of loan applicants whose demographic data points are particularly unusual. Following the methodology used by the U.S. Census in its Public Use Microdata Sample, the CFPB could switch records for similarly-situated but demographically unusual loan applicants between nearby census tracts, making it nearly impossible to reconnect individual loan applicants with public land records, but maintaining the utility of HMDA data including for users doing analysis at the neighborhood level.[8] The CFPB should explore which of the data swapping techniques the U.S. Census uses will best balance protecting loan applicants' privacy and allowing users to do meaningful analysis of HMDA data at the census tract level.

Universal Loan ID

The CFPB should suppress the universal loan ID variable from the public HMDA data if the universal loan ID will become available in mortgage recording documents. The ID would make it much easier to merge the two data sources with confidence and would allow even relatively unsophisticated data users to join HMDA data with loan borrowers' name and addresses. Because the universal loan ID is an important tool for tracking loan sales and other mortgage outcomes, the CFPB should make loan-level data with universal loan IDs available upon request. The CFPB should share HMDA data with universal loan ID included for public interest uses under a data use agreement that would prohibit using the data to re-identify individuals, using the data for marketing, or otherwise acting against the public interest. The CFPB should provide the data electronically in a timely manner and not unreasonably deny access to community groups, researchers, public officials or others who plan to use the data for public policy purposes. The data requester should be able to receive the data in electronic format from the CFPB and must not be required to travel and use the data in only certain public agency offices.

Address Data

The public HMDA data have never included loan applicants' address information, and we support the CFPB's suppression of address data to reduce re-identification risk. We encourage the CFPB instead to provide an alphanumeric code that is not tied to geography for each property

---

[7] See page 2 of https://www2.census.gov/programs-surveys/acs/tech_docs/pums/accuracy/2014AccuracyPUMS.pdf and https://www.census.gov/srd/papers/pdf/rr96-4.pdf

[8] http://www.stat.cmu.edu/~fienberg/DLPapers/Fienberg-McIntyre-LNCS-2004.pdf

that will remain consistent from year to year. This approach will allow users to track home flipping, repeated refinancing and other signs of predatory lending over time.

Disaggregated Nationality Data

The CFPB's inclusion of disaggregated nationality data for Asian and Latino loan applicants is a major step forward in documenting and addressing long-standing fair lending and fair housing concerns. Studies by Chhaya CDC, National CAPACD and other groups have uncovered pervasive targeting and discrimination against specific communities, such as South Asians or Hmong immigrants, that are not exposed by the broad race and ethnicity data now reported in HMDA.[9] National Council of La Raza has documented significant economic differences between Puerto Ricans who have recently moved to Florida and Latinos living in Florida overall.[10] Making public this detailed nationality data is critical to ensure that HMDA serves its stated purpose in Asian and Latino communities. Given that loan applicants choose to report their nationalities for inclusion in the HMDA data voluntarily and loan application forms clearly state that race and ethnicity data are used in government monitoring of lending patterns, it will not violate loan applicants' privacy to include this data in the public HMDA data.

Disaggregated nationality data, may make it easier, however, to re-identify individual loan applicants, for example, if they are the only people of a particular nationality in their census tract who applied for a mortgage. Rather than suppressing this critical field, the CFPB should use data swapping to reduce re-identification risk and should ensure that sensitive data fields are appropriately masked so that people are not harmed by re-identification if it does occur.

Age

The CFPB should mask loan applicants' exact ages to help prevent re-identification, but should include age ranges in the public HMDA data. The CFPB should report age data in 5-year ranges and should be sure to split categories of seniors into those who do not and do meet the age eligibility requirement for borrowing a reverse mortgage (currently 62 years) so that it is clear which people between 60 and65 years of age are eligible for a reverse mortgage.

Credit Score and Debt-to-Income Ratio

It is critical that the CFPB include in the public data meaningful information about loan applicants' credit scores and debt-to-income ratios. These variables are key to identifying potential lending discrimination and understanding longstanding racial disparities in mortgage lending. In addition, the variables assist regulatory agencies and the public in identifying unaffordable and unsustainable lending before it becomes a widespread problem. Although it is already possible for marketers and other companies to purchase information on people or households with particular credit score ranges or debt burdens, the CFPB should modify these

[9] See http://chhayacdc.org/wp-content/uploads/2014/05/Deepening-Roots-and-Creating-Space1.pdf
[10] See http://publications.nclr.org/bitstream/handle/123456789/1578/PR_Migrant_FL.pdf?sequence=5&isAllowed=y

fields before making them public so that loan applicants' exact credit scores and debt-to-income ratios are not visible.

The CFPB must not, however, aggregate these fields across census tracts or other geographies. Doing so would render the data useless for fair lending analysis because the data would essentially reflect existing inequities and segregation in our housing and economic system, and any analysis relying on aggregate data could be used to justify discrimination. That is, lenders have for years attempted to explain away mortgage redlining by stating that people in communities of color have low credit scores and therefore most do not qualify for conventional mortgage loans. Indeed, the average credit score in communities of color is often low, as a result of decades of housing segregation, unequal access to jobs and quality education, government and bank disinvestment, and targeted predatory lending. However, based on their creditworthiness, many people in communities of color do, or should, qualify for mortgages. If the CFPB reported critical variables such as credit scores and debt-to-income ratios as neighborhood averages or medians, then public HMDA data could not be used to analyze this critical fair lending issue within census tracts.

Instead, the CFPB should normalize the credit score data reported each year and report loan applicants' credit scores either as z-scores, a measure of a credit score's place in the overall distribution of credit scores for loan applicants that year, or in percentile ranges based on the distribution of loan applicants' credit scores. Z-scores have the advantage of being useful for statistical analysis. Percentile ranges are much easier to understand (e.g., a loan applicant's score was in the 55^{th}-60^{th} percentile for all loan applicants this year) without revealing information about a loan applicant's numerical credit score. To keep credit score information confidential, the CFPB would have to take care not to publish information about the underlying distribution of credit scores for applicants within a given year.

The CFPB should report debt-to-income ratios in top and bottom coded ranges. Ranges of less than 20%, 20% - 30%, 30% to 35%, 35% - 45% and 45% or more should be used. Top and bottom coding will protect the privacy of people with particularly high or low debt-to-income ratios. A high debt-to-income ratio is already included as a reason for denial in HMDA, so no additional information about people with high debt-to-income ratios is revealed by making ranges public.

Making HMDA Data More User-Friendly

The CFPB should expand its user-friendly HMDA data site to allow analysis of the un-masked HMDA data by building in basic privacy screening tools. The CFPB has already built a powerful query system that allows users to create summary statistics for data filtered by lender, geography and loan and applicant characteristics. The CFPB should expand this system by implementing technological solutions to protect loan applicants' privacy, such as not returning results for queries with less than a certain number of records to be summarized. While updating the system, the CFPB should also add functionality to allow users to track loan sales using the universal loan ID and group loans made by multiple lenders or multiple subsidiaries of a lender. Such a system would help increase the accessibility of HMDA data for people and community groups not seeking to analyze the full HMDA data set using statistical software but that do want to understand local lending patterns and uncover potential discrimination or opportunities for additional local investment.

Conclusion

The Dodd-Frank Act requires the CFPB to apply a balancing test in determining which HMDA data points may be made public and in what format. We urge the CFPB, when applying the test, to take into account the pivotal role of HMDA data in public policy interventions needed to address discrimination, encourage lenders to meet communities' housing needs, and identify opportunities for needed investment. Most new HMDA data fields are already available for purchase from private vendors. Moreover, the CFPB will be able to employ proven techniques to protect loan applicants' privacy when data are not already widely available. The CFPB should therefore make most of the new data fields public. For more information or any questions about the issues raised in this letter, please contact Josh Silver at National Community Reinvestment Coalition or Alexis Iwanisziw at New Economy Project.

Sincerely,

Americans for Financial Reform
Association for Neighborhood and Housing Development
Center for Responsible Lending
Consumer Action
Empire Justice Center
National Community Reinvestment Coalition
National Consumer Law Center (on behalf of its low-income clients)
National Coalition for Asian Pacific American Community Development
New Economy Project
National Fair Housing Alliance
Ohio Fair Lending Coalition
Reinvestment Partners
Western New York Law Center
Woodstock Institute