December 21, 2018

Consumer Financial Protection Bureau
1700 G Street NW, Washington, DC 20552
Request for Information Regarding Bureau Data Collections
Docket ID CFPB-2018-0031

To whom it may concern:

This letter is in response to the request issued by the Consumer Financial Protection Bureau ("the CFPB") and listed as Docket No. CFPB-2018-0031 in the Federal Register.

The National Community Reinvestment Coalition (NCRC) has long been a proponent of the idea that data drives the movement for social justice. Data allows the public, both through agencies such as the CFPB and individually, to ensure the market health of our economy and the ability of all people to participate in that economy fairly, and without fear of predatory abuse.

In service of those goals, NCRC will address each of the questions in the request for information (RFI) issued September 25, 2018. The context of these answers should be stated at the outset. NCRC feels that broad access to data that is already in the public realm or which is easily available to regulated entities on the private market is to be encouraged at the CFPB. That data cannot be considered proprietary and the collection if it requires no additional burden on the regulated entities since it is either already released pursuant to other regulations or it is sold by the entities themselves. In addition, NCRC supports the ability of the CFPB to request and use non-public data, as long as that collection and storage adheres to the data governance policies in place for all Federal agencies[1].

In this letter we will discuss the CFPB's data collection and use practices with the understanding that, outside of its enforcement functions, the research produced by the CFPB must be replicable by outside researchers to the extent that is possible. The ability to replicate, improve, and expand research methods is a cornerstone of quantitative analysis. Therefore, work produced using data that is not available to the public, unless it has a specific research or enforcement purpose, is often of less use than publicly available data.

What follows are responses to the areas of interests outlined in the RFP.

1. Aspects of the CFPB's Data Governance Program:

---

[1] https://www.consumerfinance.gov/data-research/research-reports/sources-and-uses-data-bureau-consumer-

Beyond the necessity of protecting private consumer information that could be used to harm individual consumers, such as Social Security numbers, personal telephone numbers or other contact information (email, social media, etc.) much of the data used by the CFPB does not typically enjoy an expectation of privacy on the part of the consumer. Data such as property ownership, lien filings (mortgages), bankruptcies and credit report data are available either free or for purchase by the public. These data points are also sold on a secondary market, often by entities that the CFPB is charged with regulating. Furthermore, although Acting Director Mick Mulvaney has discussed the CFPB and potential security lapses in the past, the nature and scope of this appears to be both small and unsubstantiated. These data breaches also pale in comparison to those disclosed by entities regulated by the CFPB, in particular Equifax (140 million credit profiles with sufficient detail to allow identity theft). Given the lack of verifiable scope and impact, and that the data collected by the CFPB differs little from data sold by regulated entities themselves in most cases, there is no reason to suspect that the current policies on data governance published by the CFPB are not sufficient.

2.   The CFPB's Data Collection practices related to privacy:

The CFPB has already addressed privacy issues related to their data collection in a number of ways. These include converting easily identifiable items like borrower names to numeric identifiers, withholding data from their public releases that would offer enough detail to easily identify a specific person, and aggregating data in such a way to preserve its usefulness while obscuring links to individuals.

From a research perspective, transaction level data is ideally used without text data, such as names or specific addresses. Aggregating data also helps researchers work with data while removing personally identifiable traits from individual records. The level of aggregation is an important consideration. Census tracts are an acceptable aggregation level for spatially distributed data. For mortgage related products the census tract where the home is located is currently used and is sufficiently aggregated to make specific borrower identification difficult but still allows for research. For non-property related financial products, auto loans, credit cards, etc. the borrower's mailing address can easily be geocoded and assigned to the appropriate census tract by any of a number of commercial products. This is a common feature of private datasets, and regulated entities frequently sell such data through a variety of channels. Non-spatial numeric information can be effectively aggregated into buckets based on size as well. Including borrower income, loan amounts, or age, these data are often aggregated by rendering them in 000s format (for dollar amounts) or ranges.

NCRC feels that the steps that the CFPB has taken with HMDA data allow sufficient privacy for individuals but maintain enough detail to enable public oversight of the loan process. These policies and practices should be encouraged for use with other data that the CFPB deems necessary for them to pursue their mission.

3.   Changes the CFPB should make to how it sources, uses, stores data and if that data can be used to serve other functions of the CFPB:

By its nature the financial industry is a fluid one, with new products, capabilities and threats evolving constantly. Within the statutory limitations under which it was created, the CFPB has broad latitude in how it investigates and addresses activities that unfairly harm consumers, intentional or not. Within that scope, and in accordance with published policies on data collection and security, the CFPB should have similar latitude in using data to address these evolving threats. While regulated entities might assert that flexibility in how data is used is unfair, those entities should be evaluating their activities constantly for their impact, not for the chance of being caught. Therefore, allowing the CFPB flexibility in how it uses the data it collects should not present any additional risk or burden to those regulated entities.

4.   How and when data collected primarily for one Bureau function should, or should not, be used for other Bureau functions consistent with applicable law.

As the financial and banking sectors evolve and develop new processes and products of their own the CFPB and other regulators face an increasingly complex and shifting environment to regulate. This means that new methods of analysis must constantly be created and tested. Data collected within the established processes laid out in the CFPB's data governance document should be open for use in ways demanded by the marketplace, just as regulated entities would use such data. To do otherwise would run counter to Section 1021(b)(ii)(iii)(iv) whereby the CFPB is authorized to exercise its authority to ensure that consumers are protected from "unfair, deceptive, or abusive acts and practices and from discrimination". The CFPB is also empowered to reduce the burden of unnecessary reporting and to promote greater fairness and competition in the marketplace, both goals that would be better served if the CFPB can use data collected for one function to inform another function without demanding additional data from its regulated entities.

5.   Ways to improve Data Collection processes that reduce reporting burden without hindering the CFPB's ability to accomplish statutory objectives:

The Home Mortgage Disclosure Act (HMDA) requires that the regulated entities submit mortgage loan application records each year. The CFPB has developed and deployed a system that streamlines and automates this process far better than existed previously. This model would be ideal for application across a variety of data sets, including small business loan data, call report data, community investment data, and bank branch location data. Each of these subject areas are now collected by individual regulatory agencies. The current methods of collection are diverse and burdensome to the regulated entities and their availability to the public or other agencies is haphazard at best.

In particular, the CFPB has developed these processes using current methods of software design, file format, and compression. This allows regulated entities and their third party collaborators to streamline their internal processes as well. This work has successfully united multiple goals laid out under Section 1021(a) of the Dodd-Frank Act. Specifically, they have done so by reducing the reporting burden of HMDA while increasing the transparency of this data. This was accomplished by publishing data at an earlier point in the year, in multiple formats and channels for public consumption. Collection of the data is in a standardized format via a secure 'dropbox' that regulated entities can incorporate into their compliance process. As one vendor to the financial industry advertised, submission of HMDA data requires "no specialized training" and

user reviews indicate that data moves seamlessly from the lender's loan origination software to the CFPB. This is a substantial improvement over a system that formerly relied on painstaking reviews of loan records, error correction, and submission via CD-ROM, FTP or paper files.

The experience with HMDA data collection should be a model for use in other data collection efforts, such as small business data, branch location data, loan originator data or other possible submissions.

6.   Changes the CFPB could make to existing Data Collections, or potential new Data Collections the CFPB could collect, consistent with its statutory authority, to more effectively meet the statutory purposes and objectives as set forth in section 1021 of the Dodd-Frank Act:

The CFPB has already mapped out a process for collecting data from regulated entities, maintaining the necessary level of privacy such data requires, and providing data to the public in a manner that is both fair and transparent. Additional data collections from regulated entities that would further the established goals of Section 1021(a) would include the following:

1. A crosswalk file that would allow researchers to link together datasets from various regulatory agencies would multiply the impact of those datasets with a relatively small investment. Datasets that we have linked include Community Reinvestment Act (CRA) small business data, HMDA, and bank branch data.
2. Loan level data on business lending including the loan amount, date, location, business NAICS code, ownership structure, and loan terms aggregated sufficiently to make difficult the identification of specific businesses. The SBA-7(a) loan program provides an excellent model of how to report such data.
3. Bank branch location data is now made public via the FDIC and records maintained by the regulated entities themselves. These records are prone to error and lack any information on regulated entities that are not also FDIC depositories. Mortgage and business lending are clearly linked to physical office locations and where available those locations would enhance the transparent and fair nature of CFPB data in accordance with Section 1021(a).
4. Other datasets commonly used in private research and available for purchase from a variety of sources would include, loan level note ownership and servicing agreement data, loan originator information, automated valuation model data, flood and hazard insurance costs, MLS data, deed and parcel data including information on cash sales and property taxes. Each of these datasets improves the ability of the CFPB to monitor credit markets for patterns of predatory behavior in a number of ways, including foreclosure and servicing abuse, falsified appraisals, and disruptive investor activity.

These additional data collections could be useful in supporting the statutory requirements of Section 1021(a) by replacing diverse, burdensome and outdated collection techniques in place today. In addition, the data collected would assist with guaranteeing the fairness, transparency, safety and soundness of the financial market for participants and consumers. Regulated entities benefit from understanding that all market participants are adhering to an identical set of guidelines and standards. Furthermore, since the data is

either already collected by the regulated entities in the course of business or is sold by those regulated entities to private data aggregators already there is no question of additional burden to those entities.

7. Other activities that the CFPB could engage in to make the Data Collection requests from financial institutions more effective and efficient:

The CFPB has excelled at an open and collaborative mentality that should be encouraged and expanded. Regular meetings and roundtables with advocate and industry researchers and compliance staff allows for processes to be developed which meet the requirements of all market participants. A standing committee structure that is managed by the CFPB but allows for a wide membership of regulated entities, advocates, software providers, and staff from other government regulators would allow new processes for data collection and use to be developed with the input from all segments of the market.

8. Areas where the CFPB has not exercised the full extent of its Data Collection authority; where Data Collections would be beneficial and align with the purposes and objectives of the applicable Federal consumer financial laws; and/or where the CFPB can better leverage data as a strategic asset to increase effectiveness.

The CFPB continues to lag the statutory requirements of the Dodd-Frank Act with respect to expanding the reporting of business loan data (Section 1071). In addition, the more time that elapses between passage of the Dodd Frank Wall Street Reform and Consumer Protection Act and the public disclosure of enhanced HMDA data required by Dodd Frank continues to hamper both advocates and market participants from identifying and eliminating occurrences of steering, disparate impact, and predatory lending.

Sincerely,

Jesse Van Tol
Chief Executive Officer
National Community Reinvestment Coalition