Geocoder Service

Does the CFPB use the information to benefit or make a determination about an individual?	No.
What is the purpose?	Provide an address lookup function for compliance, research, and analysis of markets and consumer data.
Are there controls to enforce accountability?	Yes, all standard CFPB privacy protections and security controls apply.
What opportunities do I have for participation?	This information does not have a public participation component.

Consumer Financial Protection Bureau

Overview

The Dodd-Frank Wall Street Reform and Consumer Protection Act ("Act"), Public Law No. 111-203, Title X, established the Consumer Financial Protection Bureau ("CFPB" or "Bureau"). The CFPB implements and enforces Federal consumer financial laws consistently to ensure that consumers have access to markets for consumer financial products and services, and that such markets are fair, transparent, and competitive. One of the primary functions of the Bureau is to publish information relevant to the functioning of markets for consumer financial products and services to identify risks to consumers and the proper functioning of such markets.

The Home Mortgage Disclosure Act of 1975, 12 U.S.C. 2801 *et seq*. ("HMDA"), requires most mortgage lending institutions to collect, report to federal regulators, and make public certain data about mortgage loan applications and originations and purchases of mortgage loans. The Federal Financial Institutions Examination Council ("FFIEC") has made this loan-level data, with certain fields redacted to protect applicant and borrower privacy, available to the public since 1991.

HMDA authorizes the Bureau to develop or assist in the improvement of methods of matching addresses and census tracts to reduce the compliance burden for depository institutions. 12 U.S.C. 2806(a)(1). HMDA requires financial institutions to record and report the geographic location of the property to which a loan or application relates, including by Metropolitan Statistical Area ("MSA") or Metropolitan Division, state, county, and by census tract, if the institution has a home or branch office in that MSA or Metropolitan Division. Currently, for some institutions, finding and reporting the accurate MSA, county, or Census Tract can be resource intensive. The deployment of the Bureau's new publicly available geocoder service tool (Geocoder), is expected to reduce burden and facilitate compliance with HMDA requirements. "Geocoding" is the process of converting addresses into geographic coordinates, which can be used to place markers or determine position on a map, and associating those geographic coordinates with other data, such a Census Tract or MSA. Users of the Bureau's Geocoder will be able to input an address and, in return, obtain the address's location (e.g. latitude and longitude) and other information regarding geography containing the address (e.g. county or census geography). As a public resource, the Geocoder can be used by members of the public, other agencies, and even the Bureau itself for geocoding needs.

Geocoding as a function has several categories of data:

- Input data (e.g. an address), which a user is intending to locate;

- Base data, which is used to find and return the latitude and longitude of the input data; and
- Ancillary data, which is used to associate the latitude and longitude of the input data to higher-level geography (e.g. Census Tracts).

A user will enter input data into the Geocoder tool and will receive back either the latitude and longitude of the location, or an associated geography of the location, or both. The input data is not stored by the Geocoder, nor is identifying information about the user stored in the system. The system will track the number of queries along with other administrative details such as error reports. The base and ancillary data are collected by the Bureau and used to power the Geocoder. This data is constantly curated by the Bureau. These data are all covered under common controls through normal data security processes at the Bureau.

Creation of the Bureau's Geocoder requires the Bureau to obtain certain information about addresses (e.g the base and ancillary data discussed above). The Geocoding PIA covers these types of information as well as the tool used to provide the geocoding service. The intake of the geocoding data is authorized by Sections 1011, 1012, and 1022 of the Dodd-Frank Act, and the Home Mortgage Disclosure Act. Its collection is compliant with applicable federal laws, including the Dodd-Frank Act, the Paperwork Reduction Act, the Right to Financial Privacy Act, and the Privacy Act of 1974. The Privacy Act of 1974 does not require a system of records notice for the Geocoder because the Geocoder does not retrieve records using the name of an individual or other identifying particular. In addition, OMB approval for the Geocoder is not required under the Paperwork Reduction Act because the Geocoder does not contain information collections within the meaning of the Act.

Privacy Risk Analysis

The primary privacy risks associated with data covered by the Geocoding PIA are risks related to:

- Data Quality and Integrity, and
- Data Minimization.

Data Quality and Integrity: The Bureau will collect a significant amount of information and could on occasion obtain out-of-date or incorrect information. The input data submitted by the public will not be maintained by the Bureau other than to process the geocoder request and

return the geocoded location. The data used to power the Geocoder (base data) has two primary sources, the US Census Bureau and individual states that publish local authoritative address data. Since these sources update those data on their own cycles (from daily to annually), the Bureau plans to refresh the data on a regular basis, likely sub-annually, thereby making the Geocoder as current a function as possible. The data used to link the geocoded location to higher level geography (ancillary data) is updated on a regular basis. Because the Bureau does not use any information collected through these types of interactions to provide or deprive an individual of a right or benefit, the privacy risks related to data quality and integrity that are associated with these collections are minimal. Some of the information the Bureau may obtain may include information from third-party sources, some of which may be publicly available. In cases where information is obtained from non-public sources, the Bureau collects such information in accordance with applicable law and pursuant to applicable agreements governing the sharing of such information (e.g. Memoranda of Understanding, Memoranda of Agreement).

Data Minimization: The Bureau reviews all collections of data in an effort to minimize the amount of directly identifying PII to the greatest extent possible, while still allowing the Bureau to complete its objectives. This may be done by stripping collections of direct identifying PII, aggregating data, or other means of minimizing such collection. For the Geocoder, the Bureau will be collecting the street addresses and location of all addresses in the United States, where available from an authoritative source like State or local government. This is the base data used to power the Geocoder. Much of this information is publicly available from State and local governments. Address information is considered PII by the Bureau, because it may indirectly identify an individual. The names of individuals living in those homes will not be collected as part of this effort, nor will any other data associated with individuals be collected. Although the addresses will be collected to power the Geocoder, the tool will minimize the data publicly disclosed to XY coordinates, Census Tract, County, State, or Metropolitan Statistical Area. Ancillary data is used to link XY coordinates to higher level geography and does not typically include PII. Because the Bureau necessarily collects a significant amount of PII, it consequently utilizes appropriate technical, physical, and administrative controls relative to the risk of the data. These controls are discussed in the subsequent sections of this PIA.

The technical, physical, and administrative controls implemented to promote individual participation, minimization, and accountability are appropriate.

Privacy Risk Management

1. Describe what information the CFPB collects, how the information is collected, and the sources from which the information is collected.

The Geocoder project collects location information for addresses in the United States. The CFPB limits the intake of PII to address because it is the only PII that is necessary for the purpose of creating the Geocoder. When a member of the public inputs an address into the Geocoder, it will return the location of the querried addresses from the collection, typically providing longitude and latitude. The primary purpose of the collection of addresses is to provide financial institutions a tool to facilitate compliance with existing regulatory requirements to submit certain property location information related to mortgage applications and loans.

The Standard Field format for address includes:

Num	Field	Description
1	Number	Household number, possibly includes prefix and suffix
2	Address	Complete street address. Includes secondary address unit designators (apt., bldg, unit, etc)
3	Alt_address	Alternate address, if appropriate
4	City	City
5	State	State
6	ZIP	5 or 10 digit ZIP Code where appropriate
7	Latitude	Latitude of Address
8	Longitude	Longitude of Address

The information may be collected from third-party partners, public sources or even procured data if required. Most commonly information is collected from:

- U.S. Census Bureau Topologically Integrated Geographic Encoding and Referencing (TIGER) information
- Publicly available State data from online systems
- Publicly available State data which is not available online
- Publicly available local government data (e.g. County-level data)

Information is also likely to be collected from the National Address Database, a database currently under development by the U.S. Department of Transportation and the U.S. Census Bureau. It will consist of data from the U.S. Federal Geographic Data Committee.

The above listed sources are all open authoritative sources of address information. We choose these sources based on the following criteria:

- Which sources produce comprehensive open data about addresses?
- Which sources are authoritative (e.g. government produced data of a government function)?
- Which sources provide the CFPB with the most flexibility in designing a system to meet business needs? In particular which sources do not have limited or overly restrictive use rights, and provide the best opportunity to not tie internal data with derivative use restrictions?

In cases where the information is derived from non-public sources, such as other Federal agencies or data brokers, the Bureau obtains such information using contracts, information sharing agreements, or other similar agreements or processes, and in accordance with applicable law.

2. Describe CFPB's objective for the information.

The CFPB is collecting the address information to power the Geocoder and to then release the geocoded coordinates to the public. The CFPB will not retain or make publicly available information entered by the public into the Geocoder. Geocoding is the result of translating an address into a location (X,Y), typically longitude and latitude. This process can be accomplished by using different methodologies and reference data. In general, there are three distinct ways to accomplish this translation:

- Using address point reference data, where a database contains the address (e.g. 123 Main St) and the longitude and latitude. In this case the geocoding function is a lookup of the requested address in the database and returning the associated location;

- By interpolating a location along a geographic line segment that contains the street name as well as the address range that it covers, then returning the associated loaction.

- By calculating the centroid of a known area (i.e. zip code)

The entire process requires two fundamental components: a) the search algorithm to perform the lookup functions described above, and b) the data (base and ancillary) upon which to perform the searches.

The Geocoder will be available to the public on consumerfinance.gov.

3. Describe how CFPB shares, for compatible purposes, any of the information with third parties, e.g. federal or state agencies, the general public.

The Bureau is making the Geocoder publicly available to facilitate compliance with existing regulatory requirements to submit certain property location information related to mortgage applications and loans. However, as a public resource, the Geocoder can be used by members of the public, other agencies, and even the Bureau itself for geocoding needs in other contexts as well.

4. Describe what opportunities, if any, individuals to whom the information pertains have to (a) receive notice regarding the CFPB's use of the information; (b) consent to such use; (c) access the information that pertains to them; or (d) obtain redress.

The Geocoder contains information that is not collected directly from individuals. Address information will be collected from authoritative sources, but no other PII will be collected. These authoritative sources are predominantly the US Census Bureau, and state and/or local government open data clearinghouses (e.g. data.gov). Notice to individuals is provided through this PIA. There are no opportunities for the individual to consent to such use, access the information that pertains to them, or obtain redress.

5. Explain the standards and relevant controls that govern the CFPB's—or any third party contractor(s) acting on behalf of the CFPB—collection, use, disclosure, retention, or disposal of information.

The CFPB complies with the Privacy Act of 1974, Right to Financial Privacy Act, and E-Government Act of 2002; adopts Office of Management and Budget privacy-related guidance as best practice;[1] and applies National Institute of Standards and Technology risk management processes for privacy.

The CFPB uses the following technical and administrative controls to secure the information and create accountability for the Bureau's appropriate collection, use, disclosure, and retention of the information:

- Audit Logs and Reviews,
- CFPB Personnel Privacy Training, including annual and role-based training,
- CFPB Privacy Incident Response and Recovery Plan and contractual obligations for third parties to support CFPB Privacy Incident Response and Recovery Plan,
- Compliance with CFPB cybersecurity policy and procedures,
- Information Quality and Integrity Checks,
- Extract logging and 90-day reviews,
- Policy and Standard Operating Procedures,
- Role-based Access Controls,
- Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies,
- Records Schedule Submitted to/Approved by National Archives and Records Administration (NARA): Records will be disposed of according to the applicable records schedule. Information in the Infrastructure GSS is covered by CFPB specific records schedules as well as general records schedules. Some records schedules are awaiting NARA approval, and
- Personnel Security supported through due diligence screening.

[1] Although pursuant to Section 1017(a)(4)(E) of the Consumer Financial Protection Act, Pub. L. No. 111-203, the CFPB is not required to comply with Office of Management and Budget (OMB)-issued privacy guidance, it voluntarily follows OMB privacy-related guidance as a best practice and to facilitate cooperation and collaboration with other agencies.

Information Quality and Integrity Checks are particularly relevant for the Geocoder. Because some sources provide more information than CFPB requires, such as different names for the same field type or addresses broken out into 5-10 fields rather than 1, CFPB retains only the data that is in the CFPB required format.

The CFPB may use contractors to help support the collection, use, disclosure, or retention of information covered by this PIA, and those contractors are subject to similar controls. Contractors with access to PII are required to report suspected or confirmed privacy incidents to the CFPB immediately and no later than one hour after discovery. Other requirements placed on contractors may include training on privacy and compliance with federal privacy requirements and Federal Acquisition Regulations.

6. Discuss the role of third party(ies) that collaborate or partner with the CFPB, if any. Identify any controls used to protect against inappropriate collection, use, disclosure, or retention of information. (This does not include third parties acting on behalf of the CFPB, e.g., government contractors discussed in Question 5.)

The Bureau looks to collaborate with third parties as appropriate. Collaboration in this context predominantly refers to the generation of base data, which makes the Geocoder more effective and efficient. Collaboration parties can include, but are not limited to, other Federal Agencies, State and local governernments, and where appropriate private data providers that add value to publically available authoritative data.

In all of these instances, controls are put in place to protect against inappropriate collection, use, disclosure, and retention depending on the type of sharing or data involved. Controls might include:

- Compliance with CFPB cybersecurity policy and procedures,
- Data Quality and Integrity Checks,
- Extract logging and 90-day reviews,
- Policy and Standard Operating Procedures, and
- Role-based Access Controls.

Document control

Approval

_____

Ashwin Vasan

Chief Information Officer

June 6, 2016

_____

Claire Stapleton

Chief Privacy Officer

June 6, 2016

_____

Michael Byrne

HMDA Operations Lead

June 6, 2016

Change control

Version	Summary of material changes	Pages affected	Date of change