Comments of the World Privacy Forum to the Bureau of Consumer Financial Protection

Regarding Disclosure of Certain Credit Card Complaint Data, CFPB-2011-0040

Via Regulations.gov

Monica Jackson
Consumer Financial Protection Bureau
1500 Pennsylvania Avenue NW.,
(Attn: 1801 L Street)
Washington, DC 20220

January 30, 2012

Re: Regarding Disclosure of Certain Credit Card Complaint Data, *CFPB-2011-0040*

The World Privacy Forum appreciates the opportunity to submit comments to the Consumer Financial Protection Bureau's (CFPB) proposed policy statement about the CFPB's proactive disclosure of credit card complaint data. The proposed statement appeared in the Federal Register on December 8, 2011 at 76 Federal Register 76628, http://www.gpo.gov/fdsys/pkg/FR-2011-12-08/pdf/2011-31153.pdf and at https://www.federalregister.gov/articles/2011/12/08/2011-31153/disclosure-of-certain-credit-card-complaint-data.

The World Privacy Forum is a nonprofit, non-partisan public interest research group, with a focus on research and analysis of privacy issues, along with consumer education. For more information on WPF activities, see generally http://www.worldprivacyforum.org.

The WPF generally supports making information about consumer complaints public to the extent possible without violating the privacy of consumers or without discouraging consumers from making complaints. As an example, we note that the WPF has made effective use of aggregate consumer complaint data from the FTC's Consumer Sentinel database to produce a map showing the location of all medical ID theft complaints. We relied very heavily on the city field for locating a criminal pattern. Without the city in the complaint, we could not have created this data visualization, and the data would have been much less useful, as zip code only does not narrow the information enough in rural areas, or in some cases, less populous states in general. See: http://www.worldprivacyforum.org/medicalidentitytheft-map.html.

Our pioneering work on this map showed for the first time a strong geographic clustering of medical identity theft in Florida, California (especially southern California), New York, Arizona, and Texas. The map offers useful information that allows consumers to evaluate medical identity theft risks, health care providers and insurers to make better choices about appropriate responses to health care fraud, and federal and state agencies to make more informed decisions about the best allocation of resources for investigation and prosecution of medical identity theft cases and health care fraud activities in general. This data offers a window on the medical identity theft threat, and it illustrates how third parties like the WPF take unused data from a federal agency and produce new information that the agency itself did not provide.

Having compiled the medical identity theft data, the WPF is well aware that the quality of the underlying data and non-random nature of the complaints limit the value of the resulting analysis. Indeed, on our website we take note of the limitations of the data, observing that "we see this map as one additional piece of information about the geography of a crime that is difficult to detect and analyze." We cannot, however, allow the lack of data perfection to limit the use and analysis of the data that is available.

The CFPB appropriately takes note these general problems of the quality and representativeness of complaint data in its discussion of issues in its Federal Register notice. While we agree that any complaint data is far from perfect, it remains essential that data be available whenever possible. Everyone is free to comment on the shortcomings of any conclusions drawn from the data and to add additional data to the marketplace of ideas. However, if data is withheld because it is not perfect, important discussions may not occur at all or discussions will be more poorly informed when they do occur. We can tell you unambiguously that without the FTC Consumer Sentinel data that extremely important information about medical identity theft would never have been revealed, such as the geographic modalities of the crime. Additionally, our first report on medical ID theft in 2006 made use of Consumer Sentinel data, it was crucial to our documentation of this crime in what was the first-ever report on the issue.

The credit card industry has information about consumer complaints, but it has no incentive to share that data with outsiders. In effect, the industry has something akin to a monopoly of information about consumer complaints. The limited data collected and made available by agencies and, sometimes, by consumer organizations undermines that monopoly. The Internet is increasingly allowing consumers to talk among themselves, and this too usefully increases public discussion and knowledge.

Under any regime, industry is always able to add its own information to public discussions, but it may have no incentive to do so unless there is other information available that starts the discussion. Any differences of opinion about the nature and significance of data will add to the public debate and inform all participants. Recognizing the need to protect consumer privacy, we nevertheless have no difficulty in reaching the conclusion that more data (and more resulting public debates) will best serve the public interest.

We do not have the technical expertise to comment in detail about the specific data fields that can be released without threat to the privacy of individual consumers. However, we welcome the opportunity to offer some general observations.

First, the ability to identify consumers from scraps of data – often data without any unique individual identifiers – is already widespread and increasing. Professor Latanya Sweeney, now at Harvard University, has shown that 87% of Americans are uniquely identified by their date of birth, gender, and five-digit zip code. So much personal information resides in the files of credit bureaus, data brokers, and others who profit from the exploitation of consumer data that the threats to privacy from the release of non-identifiable and even aggregate data increase every day. However, statisticians are actively developing techniques that allow for the disclosure of useful data that is nevertheless protected from re-identification. Professor Sweeney is a leading voice on these techniques, and we refer you to her work (http://latanyasweeney.org/work/identifiability.html) for additional ideas. Others are also involved in developing techniques that support disclosure and protect privacy.

Statisticians have a bagful of tricks that can satisfy the mildly conflicting goals of preventing re-identification and allowing information to be public so that the number and flavor of the complaints to be known and trackable over time. Professor Sweeney, for example, developed a technique for locating and replacing personally identifying information in unrestricted text that might have some application to the narrative field that CFBP reasonably proposes to withhold at present. See http://dataprivacylab.org/projects/scrub/index.html. CFPB should explore these and other creative ideas that support both data sharing and privacy.

Second, where public disclosure of data cannot be justified, it may still be possible to share data with legitimate researchers under conditions that provide accountability. An increasingly standard method for preventing re-identification of consumer data is a data use agreement. The privacy rule issued under the authority of the Health Insurance Portability and Accountability Act offers an interesting, if partially flawed, model for data use agreements involving health data. See 45 C.F.R. § 164.514(d).

The terms of these agreements typically require data recipients to agree to various privacy and security obligations, including not attempting re-identification. Privacy expert Robert Gellman recently published an article that proposed interesting legislative (or potentially regulatory) ways to employ data use agreements. See *The Deidentification Dilemma: A Legislative and Contractual Proposal*, 21 Fordham Intellectual Property, Media & Entertainment Law Journal 33 (2010). http://iplj.net/blog/wp-content/uploads/2010/11/C02_Gellman_010411_Final.pdf. The article is to offer a contractual solution for the sharing of deidentified personal information while providing protections for privacy. A data discloser and a data recipient can enter into a voluntary contract that defines responsibilities and offers remedies to aggrieved individuals. When data cannot be made freely public because of privacy concerns, data use agreements can allow for the reasonable use of data for analysis while limiting threats to privacy of data subjects.

Third, the CFPB notice discusses the possibility of the disclosure of narratives with consumer consent. We would be cautious here. Whether choices are offered via opt-in or opt-out, consumers often do not understand the terms, the stakes, or even the basic choices. In most contexts, consumers will accept the default option (whether for or against disclosure) approximately 95% of the time. Given the broad range of consumer capabilities, we are disinclined to rely on a notice and choice system here.

Both consumers and CFPB will face threats and consequences if consumers agree to public disclosure of personally identifiable data. It is not hard to imagine that crooks will exploit public data, contact consumers "on behalf of" CFPB or otherwise, and bilk consumers out of money or more data. It would only take a news story or two for public and congressional pressure to stop these disclosures. A defense that the consumer agreed to the disclosure – no matter how thoroughly CFPB warned the consumer of the risks – would not be effective.

We do not necessarily agree with CFPB's assertion that "the privacy risks cannot be systematically assessed other than by reviewing the complaints and issuer responses that we receive." The FTC has done an excellent job of balancing this issue in its Consumer Sentinel database, and we think the CS database serves as an excellent model here. We think that statistical and other techniques may possibly allow automated review of complaints for purposes of public disclosure.

Overall, we support the CFPB proposed policy statement and methodology for disclosing consumer complaint information. It strikes a reasonable balance between somewhat conflicting objectives. We enthusiastically support the need for further study of disclosures for finding ways to increase disclosure while fully protecting the privacy of those who file complaints.

Thank you for the opportunity to comment on the CFPB policy statement.

Respectfully submitted,

/s

Pam Dixon
Executive Director
World Privacy Forum
www.worldprivacyforum.org