1634 Eye Street, NW
Suite 1100
Washington, DC 20006

Statement of **Michelle Kathleen De Mooy**
Deputy Director, Consumer Privacy Project
Center for Democracy & Technology

Before the United States House of Representatives Committee on Science, Space, and Technology, Subcommittee on Research and Technology, Subcommittee on Oversight

Can Americans Trust the Privacy and Security of Their Information on HealthCare.gov?

February 12, 2015

Chairman Smith, Ranking Member Johnson, Chairwoman Comstock, Chairman Loudermilk and members of the Committee:

Thank you for the opportunity to testify today on behalf of the Center for Democracy & Technology. CDT is a nonpartisan, non-profit technology policy advocacy organization dedicated to protecting civil liberties and human rights on the Internet, including privacy, free speech, and access to information. I currently serve as the Deputy Director of CDT's Consumer Privacy Project, which focuses on developing privacy safeguards for consumers through a combination of legal, technical, and self-regulatory measures. Ensuring that services are designed in ways that preserve privacy, establishing protections that apply across the life cycle of consumers' data, and giving consumers control over how their data is used are key elements of protecting privacy in the digital age.

We welcome the attention the Committee has given to the pressing issues of consumer data privacy and security through the lens of data sharing on HealthCare.gov. CDT's testimony today will briefly describe current data collection and information sharing practices, how HealthCare.gov employs collection and sharing, and describe the associated privacy and security concerns. I will finish with policy and technical recommendations.

I. Data collection and sharing online

There are several layers of communication taking place each time an individual accesses a website. Some of these layers happen behind the scenes, without a user's express engagement, and some are more direct. Direct website interaction includes filling out a forms or signing into accounts. These interactions typically give consumers a fairly commonsense notice of the information they are sharing. Not all direct interactions are quite this clear. For example, a consumer may not know that user names or email logins may be used to link consumers visits across different websites.

A less obvious, but similarly straightforward, communication occurs when individuals chose to visit a website. The action of clicking on a link or typing in an address triggers a message from your browser to the intended website's server. This action essentially announces your arrival, while sharing basic information like your IP address—just like your phone number is your address on the telephone network, your IP address is your address on the Internet—in order to correctly load the site on your browser. Information exchanged during this process serves a utilitarian purpose—for example, the server needs to know which language you speak and what kind of graphics your computer will allow you to see in order to load the site correctly. Often, the basic information exchanged in this process is used to recognize you and customize your experience in subsequent visits. Information about users is often sent via a referer header, which acts as a kind of sign that people unknowingly carry around online as they surf. This sign lists the last websites that the person has visited and is used both by websites themselves and third parties, such as advertising companies. The information that is exchanged is called the refering URL and it sometimes includes browsing and search information that has directly been encoded into the web link.

On a level less visible to consumers, websites use tracking technology to get a more detailed look at them. To do this, they employ different methods to record a user's behavior as he or she navigates that particular site and even on other websites. Generally speaking, technologies on a website that record behavior and track users across visits (and across different websites) are what we mean when we refer to tracking technology.

There are a many types of tracking technologies, each with slightly different properties[1] but all serving the same general purpose of identifying an individual website visitor across time – an important distinction is made between first party tracking, or the capture of information by the website itself, while third party tracking is when other entities, typically unknown to the consumer, are contracted by the website to do analytics or other purposes. The most well known example of a tracking technology is a cookie, or a small file containing identifying information, that is stored on a computer at the request of a website's server – depending on your browser settings, you may be asked for permission for the server to do this but you may not, and it's fair to say that many times users are unaware it is occurring. Cookies are often used to improve the online experience by reducing loading speed and storing preferences like login information or remembering abandoned shopping carts. And when cookies from the same company appear on multiple websites—such as when an analytics company or

[1] For general descriptions of tracking software, see "*Know your Elements,*" a website by Ghostery. http://www.knowyourelements.com/. Visited on Feb 10, 2015. Some pieces of tracking software are more easily blocked by users, such as those with the ability to clear cookies from a browser. This has prompted an arms race of sorts with increasingly sophisticated tracking tools, such as super cookies, being downloaded by unsuspecting users.

ad network services several distinct sites—that company can correlate your activity across multiple different web contexts in ways that consumers might find unwanted or surprising. The information conveyed by browsing habits is used to develop a marketing profile of an individual: this might include the types of websites and pages a a consumer visits, as well as any web searches such as those for information on particular diseases or pharmaceuticals. This information can them be combined with offline data such as address, income, marital status, and prescription drug history to form a dossier. In this way, information about health-related information can be collected and interpreted solely in the context of a person's website browsing and searching habits.

It's important to note that the presence of tracking software may be justified, depending on the circumstances—many websites collect this type of information in order to observe the profile of visitors to their own site, something referred to as web analytics. CDT doesn't use cookies or third parties to perform analytics, but we do look at the log files generated by our servers to get a sense of what content people are interested in and where our visitors come from. Many other commercial and non-commercial sites feel comfortable using third party analytics providers; this results in sharing information about site visitors with companies with which the user has no awareness or relationship.

Whether the site itself or a service provider collects the data, performing web analytics are a key part of the online ecosystem. They allow websites to be responsive to their users interests and intentions in using their website – for example, HealthCare.gov may use web analytics to determine if visitors want to learn information eligibility right away and be directed instead to information about plan rates. The goal of digital analytics is to optimize the site so visitors will want so that they will stay on their site longer, viewing more advertising or buying more products in the case of e-commerce sites, or making it easier for people to enroll in an insurance plan in the case of HealthCare.gov.

Retargeting, also known as remarketing, is a cookie-based advertising technology that allows entities to promote their content they had previously engaged with on other sites around the web. For example, if you looked at a certain pair of shoes on Zappos.com, you might later see a remarketing ad for those same shoes on a different site. To serve these ads, a cookie is placed on a website visitor's computer when they visit a certain site. When these users browse online, this cookie allows that site, and any ad networks with which they do business, to serve them ads based on what they previously did on the original site. The cookie also allows website operators and their advertising partners to know specific details about the visitor such as what products they may have looked at and what they may have placed into a shopping cart. The idea behind retargeting is engaging users in a website by using advertisements that remind them of the products and services they were interested in and converting them into buyers.

II. What happened on HealthCare.gov?

Several weeks ago, the security firm Catchpoint Systems found that user health information was being shared with 50 or more third party entities on HealthCare.gov, without user knowledge or permission. The ensuing media firestorm attracted the attention of privacy and security advocates alike, as well as lawmakers from both sides of the aisle.

When citizens visit HealthCare.gov to learn more about the programs offered to them under the Affordable Care Act, they are asked to give certain pieces of personal information in order to shown which health insurance plans they qualify for in their state. Surprisingly, HealthCare.gov then sent a referer URL to an array of third parties that included, unbeknownst to users, includes some of the information submitted to the site such as parental status, zip code, state and annual income.

Administration officials have said that the referer URL was directed to third parties in order to give consumers a "simpler, more streamlined and intuitive experience" and this is doubtless true. It appears that the designers of HealthCare.gov contracted with third parties primarily with the intention to gain insight into the way the site was being accessed and used. Officials have also said these technologies were used "to get visibility into when consumers are having difficulty, or understand when website traffic is building during busy periods."[2]

It is true that the technology used on the site can help achieve these internal goals; however, contracting with third parties requires a two-way exchange of information. The government's decision to work with outside vendors allowed private companies to access user health information without knowledge or consent, and without the readily available and easy ability to avoid this exchange. Ad tracking technologies can be used to help advertisers, such as insurance brokers or other health or medical companies, tailor targeted ads solely to people who have visited government healthcare sites and add them to profiles indicating their interest in health insurance or in specific health and medical services. This type of tracking is not just happening on HealthCare.gov – Ghostery recently found many third parties receiving user information on 16 state insurance exchange sites, including personal health information.

The use of re-targeting to increase awareness of and enrollment in available health insurance plans would have been an understandable goal for the government in this case – and it appears likely to have played a role[3]; however, an understandable goal is not a free pass for the government to share user

[2] Center for Medicaid and Medicare Services, Press Release, *Protecting Consumer Privacy on HealthCare*.gov. January 24, 2015. http://www.cms.gov/Newsroom/MediaReleaseDatabase/Press-releases/2015-Press-releases-items/2015-01-24.html

[3] Kaye, Kate. *HealthCare.gov and State Sites Still Crawling with Ad Trackers.* AdAge, February 5, 2015. http://adage.com/article/privacy-and-regulation/healthcare-gov-state-sites-crawling-ad-trackers/296982/

information and characteristics with an array of third-party commercial entities without permission from users themselves.

III. Privacy concerns

Sharing of personal information with third parties is a privacy concern for several reasons. People who visit government websites often do not have a choice. They must visit a designated online place in order to access specific government products and services, such as those on HealthCare.gov. For this reason, the government should have been extremely cautious in its approach to third party sharing. Without an easy to implement option to opt-out, users were effectively coerced into agreeing to share personal health information, a clear violation of their expectations. At a minimum they should be given a timely and meaningful understanding of how their data is being collected and used by the website and by any third parties, and they should be given a choice about whether or not this is acceptable, with alternative access to comparable information and services if they choose to opt out.

Because there is a universe of companies that hold volumes of data about individuals, the addition of health information such as pregnancy status rounds out a data profile that can be used for profit. Health information is sold for a high premium on underground markets – some experts estimate as much as $40-$50 a record[4] – because it is fairly easy to monetize for criminals seeking to bill expensive medical items to Medicaid for example or to commit medical identity theft. Unlike financial details about a person, which can be reissued when compromised, health information is more valuable because it changes less often and is not as easy to reissue. Health information is not monitored routinely in the same way that banks monitor financial activity and thus it is harder to recognize theft and harder for consumers to seek redress. Individuals can get a new credit card but it is not as easy to change or obtain a new medical profile.

Some though not all citizens that lack health insurance are from disadvantaged communities, and thus the calculus for deciding on the use of third parties should be weighed towards privacy and away from sharing. Consumers in disadvantaged communities face more potential for harm such as being profiled in data banks as "Rural and Barely Making It," "Ethnic Second-City Strugglers," and "Retiring on Empty: Singles."[5], categories which a recent Senate Commerce

[4] Hu, Elise. *Anthem Hacks Renews Calls for Laws to Better Prevent Breaches*. National Public Radio, February 5, 2015. http://www.npr.org/blogs/alltechconsidered/2015/02/05/384099135/anthem-hack-renews-calls-for-laws-to-better-prevent-breaches

[5] Senate Committee on Commerce, Science, and Transportation, Office of Oversight and Investigations Majority Staff. *A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes,* December 18, 2013. Page ii. http://www.commerce.senate.gov/public/?a=Files.Serve&File_id=0d2b3642-6221-4888-a631-08f2f255b577

Committee report found. These characterizations may then prompt advertising of the type of subprime mortgage loans and other predatory lending that perpetuates the cycle of poverty.

The online environment is rife with this kind of data collection and sharing and while some companies behave responsibly with user data, many do not. As a steward for consumer protection, we believe the government's online activities should be held to a very high standard. The government should be constrained about the sharing of personal data, should be highly transparent, and should consider doing analytics or retargeting of any kind in-house in order to minimize privacy and security risks.

IV. Security concerns

The number of third-party content providers loading code into the browser of visitors to HealthCare.gov poses serious security issues. Researchers have pointed to third-party content as one of the primary ways for websites to be infected with malware.[6] Compromising the integrity of third party content providers can accomplish a wide range of attacks, from simply changing the content of the page to capturing user information and credentials like passwords.[7] There is no evidence that personal information from HealthCare.gov has been misused, but the number of outside parties that can load content (essentially code executed in the browser) and that can see personal health information about users is troubling. Vendors without a direct relationship (and accountability) to the user are often the weakest link in the privacy and security chain.

Malicious code was uploaded to the website in July of 2014[8], meaning that the web portal was successfully hacked, though authorities maintain that no personal information was stolen at that time. In September of 2014, a Government Accountability Office (GAO) report warned that "increased and unnecessary risks remain of unauthorized access, disclosure or modification of the information collected and maintained by HealthCare.gov." As of February 2015, the GAO's six specific recommendations to improve HealthCare.gov privacy and security appear to not have been fully implemented.

[6] Provos, Niels, McNamee, Dean, Mavrommatis, Panayiotis, Wang, Ke and Modadugu, Nagendra. *The Ghost In The Browser Analysis of Web-based Malware*. April 2007. https://www.usenix.org/legacy/events/hotbots07/tech/full_papers/provos/provos.pdf

[7] Grossman, Jeremiah. *Third-Party Web Security FAQ*, July 1, 2010. http://jeremiahgrossman.blogspot.com/2010/07/third-party-web-widget-security-faq.html.

[8] Yadron, Danny. *Hacker Breached HealthCare.gov Insurance Site.* Wall Street Journal, September 4, 2015. http://www.wsj.com/articles/hacker-breached-healthcare-gov-insurance-site-1409861043

From the perspective of US Government federal information privacy guidance, there are very few standards or other sources of guidance from agencies like the National Institute of Standards and Technology (NIST) that could be useful for entities like CMS when setting up a complicated information service like healthcare.gov. The most relevant material to privacy guidance is Appendix J of NIST Special Publication 800-53,[9] a catalog of privacy controls that can be employed beyond security measures to ensure privacy violations are minimized. However, a list of controls without any guidance or framework as how to apply them is limited in value and application. Comprised of a menu of privacy-enhancing tools that federal agency privacy technical folks should consider using in their systems, organizations, and deployments, they are useful but without a framework for practical implementation. There is an ongoing and important NIST effort to create standards for privacy engineering[10] – which would provide the guidance necessary around the controls in Appendix J of SP 800-53 – around a risk assessment framework. While this is a very important effort, it is not yet operational such that federal government designers and engineers could use it while designing and deploying information systems.

V. HealthCare.gov's privacy policy

We believe that HealthCare.gov should have been designed to strictly limit third party data sharing. The practice of sharing with various third parties was, in this case, exacerbated by poor disclosures in the HealthCare.gov privacy policy. HealthCare.gov's privacy policy is quite broad and overly vague, allowing for essentially unlimited user data to be shared with third parties.

Importantly, personally identifiable information (PII) is not defined in the policy. Although the National Institute of Standards and Technology (NIST) has identified data points that should be considered PII, there is no requirement that government agencies or companies adopt NIST's definition. This creates a loophole that, without guidance from HealthCare.gov's privacy policy on what constitutes PII, may allow for some personal information to fall outside the site's policy protections. As the FTC has described, individuals possess an interest in potentially identifiable information beyond "PII,"[11] but the Healthcare.gov privacy

[9] See Appendix J (starting at p. 437) of NIST SP 800-53, "Security and Privacy Controls for Federal Information Systems and Organizations," available at: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf.

[10] *Privacy Engineering at NIST* homepage. Accessed February 9, 2015. http://csrc.nist.gov/projects/privacy_engineering/index.html

[11] Federal Trade Commission, Staff Report. *Self-Regulatory Principles for Online Behavioral Advertising*, February 2009. http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-staff-report-self-regulatory-principles-online-behavioral-advertising/p085400behavadreport.pdf

policy does not describe or acknowledge the possibility that personal information may be collected without their knowledge through cookies and web logs.

The site policy states that it “uses a variety of technologies and social media services to communicate and interact with citizens” but it is unclear from this policy how extensive these communications are and what citizen information is collected and by whom. The privacy policy should at the least note what information, if any, is typically collected on citizens through third party interactions and how this information is used, stored and shared by HealthCare.gov. Furthermore, the description of use of cookies is, at best, confusing, by conflating first and third-party cookies. HealthCare.gov notes it does not collect personal information through cookies, but it is unclear whether third parties do have access to a HealthCare.gov users’ personal information through cookies. Further, the policy does not place limits on how long collected data may be retained. The policy states that it will keep data “as long as needed to support the mission of the website”. This essentially allows for limitless retention of citizens’ data, which increases the data sets’ vulnerability to hacks.

HealthCare.gov’s privacy policy states ”CMS conducts and publishes a Privacy Impact Assessment (PIA) for each use of a third-party website and application (TPWA) as they may have a different functionality or practice. TPWA PIAs are posted for public view on the HHS website at http://www.hhs.gov/pia.”

Presumably, any entity that participates in data flows should be subject to a PIA when installed and when changed materially in function, especially if the parties will be involved in directly handling sensitive health information, as was the case here. Therefore, PIAs for all 50 entities found to be sharing information should have been available on HealthCare.gov’s privacy policy; if these PIAs were conducted, they are not readily discoverable on HHS’s PIA website.

The privacy policy also claims that a user should “…review the third-party privacy policies before using the sites and ensure that you understand how your information may be used,” a direction that is both unrealistic and overly burdensome for consumers, as well as being somewhat disingenuous since many consumers are not aware at all of the third party collection of their data on the site.

Two memorandums from the Office of Management and Budget (OMB) provide clear guidance for federal agencies using analytics technology, including those supported by third parties, which in this appears to have been ignored by website developers. According to the OMB’s 2010 “Guidance for Online Use of Web Measurement and Customization Technologies,” web measurement or customization technologies must not “compromise or invade personal privacy.”[12]

[12] Office of Management and Budget. *Memorandum For The Heads Of Executive Departments And Agencies* June 25, 2010. Page 4. *“Federal agencies are forbidden from using technologies that: 1) track user individual-level activity on the Internet outside of the website or application from which*

The OMB further requires agencies to provide “clear, firm, and unambiguous protection against any uses that would compromise or invade personal privacy.” Additionally, OMB requires that agencies using this technology provide an easy method for the public to opt-out such that the information available to individual users is equal.[13] The third party sharing practices on HealthCare.gov appears to have violated these guidelines, as it is not clear if, as the agency has stated, turning off cookies would have sufficed to stop this type of sharing.

OMB’s “Guidance for Agency Use of Third-Party Websites and Applications” states “when information is collected through an agency’s use of a third-party website or application, the agency should collect only the information necessary for the proper performance of agency functions and which has personally identifiable information (PII) is collected, the agency should collect only the minimum necessary to accomplish a purpose required by statute, regulation, or executive order.” HealthCare.gov is also in violation of these rules. The government could have chosen to restrict information sharing to only that needed for the functionality of the site, running its analytics internally. Though it’s not clear if the site used retargeting to reach consumers who failed to complete a transaction, it’s dubious whether such a purpose is *necessary* under the OMB guidance.

VI. Recommendations

The privacy and security missteps taken by HealthCare.gov were avoidable. Not only did the OMB offer sound and easy-to-implement guidance on third party sharing scenarios that the website designers ignored completely, there are workable alternatives to third party sharing, such as performing analytics using only first party data collected on HealthCare.gov via software that does not send personal user information to the software maker. Another option would be

the technology originates; 2) share the data obtained through such technologies, without the user’s explicit consent, with other departments or agencies; 3) cross-reference, without the user’s explicit consent, any data gathered from web measurement and customization technologies against PII to determine individual-level online activity; 4) collect PII without the user’s explicit consent in any fashion; or for any like usages so designated by OMB.”

[13] Office of Management and Budget. *Memorandum For The Heads Of Executive Departments And Agencies* June 25, 2010. Page 5. *“Clear Notice and Personal Choice. Agencies must not use web measurement and customization technologies from which it is not easy for the public to opt-out. Agencies should explain in their Privacy Policy the decision to enable web measurement and customization technologies by default or not, thus requiring users to make an opt-out or opt-in decision. Agencies must provide users who decline to opt-in or decide to opt-out with access to information that is comparable to the information available to users who opt-in or decline to opt-out.” “Clear Notice and Personal Choice. Agencies must not use web measurement and customization technologies from which it is not easy for the public to opt-out. Agencies should explain in their Privacy Policy the decision to enable web measurement and customization technologies by default or not, thus requiring users to make an opt-out or opt-in decision. Agencies must provide users who decline to opt-in or decide to opt-out with access to information that is comparable to the information available to users who opt-in or decline to opt-out.”*

creating sharing buttons that direct users to social media without sending user information to these sites.

A careful implementation of privacy principles could have prevented the problems with HealthCare.gov. Specifically the site should have used only the data needed for functionality, restricting data sharing with third parties unless absolutely necessary, and adhered to rules that allow for user opt-outs or opt-ins and provide access to information without data sharing. As a general rule, one supported by the recent data breach of Anthem, government agencies and other organizations involved in health information should stop using Social Security Numbers as patient identifiers, encrypt data in transit and at rest, and institute a culture of data privacy and security that includes comprehensive training. We would hope that in the future when a third-party web application or analytics service is installed on HealthCare.gov that 1) at a minimum, a PIA has been conducted and is easily available to visitors via the healthcare.gov privacy policy page; and, 2) that only non-sensitive personal information will be exchanged, intentionally or not, with these third-parties.

The government should address and fix the problems identified in the GAO report. It should also adopt a policy of third party sharing only when necessary for site functionality. It should strictly follow the practical and privacy-protective guidance offered by OMB and should rewrite HealthCare.gov's privacy policy to make it responsive to these recommendations. Ultimately, Congress can best protect consumer information by strengthening legal incentives for companies to better safeguard data and by enacting comprehensive data privacy legislation to give users more insight and control over how their information is collected and used.