KEEPING THE INTERNET
OPEN • INNOVATIVE • FREE

www.cdt.org

CENTER FOR DEMOCRACY
& TECHNOLOGY

1634 I Street, NW
Suite 1100
Washington, DC 20006

P +1-202-637-9800
F +1-202-637-0968
E info@cdt.org

CDT POSITION PAPER ON THE TREATMENT OF PSEUDONYMOUS DATA UNDER THE PROPOSED DATA PROTECTION REGULATION

May 23, 2013

In recent months, the debate around the proposed European Data Protection Regulation has increasingly focused on whether the law should afford special treatment to pseudonymous data. CDT has previously argued that the Regulation should be formulated to incentivize companies to keep data in less readily-identifiable forms, and different treatment of pseudonymous data does make sense in certain cases. At the same time, we believe that the definition and rules for the processing of "pseudonymous data" must be carefully constrained so that this exception does not swallow the rule that citizens have a right to the protection of their personal — including pseudonymous — data.

I. Three States for Data

CDT believes that providing for differentiated protections for pseudonymous data is appropriate, and we have previously argued in favor of a similar regulated structure in the United States.[1] At the same time, it is essential that the Regulation recognize that pseudonymous data remains a subset of *personal data* subject to the fundamental protections afforded to personal data by the European Convention on Human Rights.[2]

In our previous writings on the draft Regulation, CDT has advocated that the legislative text needs to clarify that the Regulation includes protections for pseudonymous data.[3] We noted that Recital 24 in the draft Data Protection Regulation implies that online identifiers such as cookies are "personal data" only insofar as they can be combined with other information to clearly identify individuals.[4] We urged that this language be clarified to ensure that data subjects possess a personal interest in data that is not readily linkable to real-name identity. That is, it is important to recognize that pseudonymous data are *still personal data* for two reasons: (1) unlike truly anonymous data, pseudonymous data could be tied back to an individual with new information, and (2) individuals

---

[1] Center for Democracy & Technology, *CDT Top-Level Analysis of the Commercial Privacy Bill of Rights Act of 2011*, April 27, 2011, https://www.cdt.org/files/pdfs/20110427_kerry-mccain_analysis.pdf.

[2] Convention for the Protection of Human Rights and Fundamental Freedoms as amended by Protocols No. 11 and No. 14, June 1, 2010, http://conventions.coe.int/Treaty/en/Treaties/Html/005.htm.

[3] [3] Center for Democracy & Technology, *CDT Analysis of the Proposed Data Protection Regulation*, March 28, 2012, https://www.cdt.org/files/pdfs/CDT-DPR-analysis.pdf.

[4] Center for Democracy & Technology, *CDT Analysis of the Proposed Data Protection Regulation*, March 28, 2012, https://www.cdt.org/files/pdfs/CDT-DPR-analysis.pdf

can be singled out by having their activities monitored, altered, and personalized through pseudonymous data collection and use.[5]

At the same time, we put forth the proposition that the collection and retention of data in *less identifiable* forms should be given more permissive status under the Regulation, to incentivize companies to keep data in a form that is less likely to be tied to particular individuals.[6] Pseudonymization can be an important constraint upon the association of data sets to real-world identity; it is not perfect, but sufficiently valuable to encourage companies to store data in this form. Moreover, the impossibility of authenticating pseudonymous data sets (without collecting more identifying information in violating of Article 10 of the Regulation) means that certain of the Regulation's protections (such as access and data portability) are inappropriate, as the security threat from illegitimate interception of personal information outweigh the countervailing benefits.

II. How to Define Pseudonymous?

The first question is what exactly should qualify as pseudonymous data that merits less rigorous rules under the Regulation. In order to qualify as pseudonymous data deserving of intermediate levels of protection, the data must not be linkable to a particular *individual*, but can be tied to a particular *device*. On the other hand, to qualify as deidentified or anonymous data outside the scope of the Regulation, a data controller must reasonably process data in a way such that it cannot correlate that data to a person *or device*. CDT supports the test proposed by the Federal Trade Commission to define data outside the scope of data protection responsibilities: (1) the controller has taken steps to ensure that the data with reasonable confidence could not be tied to a person or device, (2) the controller publicly promises not to try to associate the data with a person or device, and (3) every party to which the controllers transfers the data is contractually prohibited from attempting to re-identify the data.[7]

Not all pseudonyms necessarily merit special status under the Regulation. Many pseudonyms are permanent and easily tied to real identity, and thus should not receive less stringent protection. For example, email addresses are one common form of pseudonym. However, email addresses are easily searchable through social networking sites or data brokers,[8] so that they can be tied to particular individuals with only a modicum of effort. Similarly, permanent device identifiers that are publicly shared and not configurable by a user are often not reliably divorced from real world identity, as many of the parties who have access to those pseudonyms could also have identifying information about the individual (such as websites and applications that require login or credit card information). For this reason, we believe that controllers that process pseudonyms that are *universal* or *persistent* should carry a greater burden to qualify for

[5] Center for Democracy & Technology, *Comments for the Center for Democracy & Technology Before the Federal Trade Commission in the Matter of Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Business and Policymakers*, February 18, 2011, https://www.cdt.org/files/pdfs/20110218_ftc_comments.pdf.

[6] *Id.*

[7] Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change*, March 26, 2012, http://ftc.gov/os/2012/03/120326privacyreport.pdf.

[8] Spokeo, *Spokeo Email Search*, http://www.spokeo.com/email-search.

intermediate pseudonymous treatment under the Regulation given that in many circumstances they may be more easily linked to specific individuals.[9]

On the other hand, domain-specific HTML cookies, like first-party cookies and online tracking cookies, could reasonably be considered worthy of intermediate privacy protection. These cookies are only readable by one company, so the pseudonym is not globally shared. The user also has control over the cookie, and can either delete it or prophylactically prevent parties from setting one. So long as the party setting the cookies does not collect personally identifying information, it has no readily available means to tie the pseudonymous profile to a particular individual.

IP addresses are a trickier issue, as they are in some cases static, and they are widely shared. Moreover, Internet service providers are required to retain identifying information about IP addresses pursuant to the Data Protection Directive, so IP addresses could be readily tied to an individual with an ISP's cooperation.[10] The European Data Protection Supervisor has previously argued that for these reasons, they should not be considered for lesser protection under the Regulation.[11] However, most residential IP addresses do in fact change regularly,[12] and websites typically do not have access to ISP records about subscribers. Moreover, much of the modern Internet is predicated on the processing and sharing of IP addresses; a visit to nearly any webpage will result in a user's IP addresses being processed by a number of companies who provide content for the page. An ordinary user would probably not want to have to consent for each company to obtain her IP address just to deliver content for the page. For these reasons, pseudonymous-level protection may be justifiable. To alleviate the uncertainty regarding pseudonymous data sets, the Regulation could forbid companies from providing identifying information to commercial queries for real-name information about pseudonymous identifiers. And while we are skeptical about the need for Commission rulemaking on several aspects of the Data Protection Regulation, it could make sense for the Commission to retain the ability to revise the categories of identifiers that qualify for pseudonymous treatment over time as technologies and business practices evolve.

---

[9] *See*, *e.g.*, Google, Measurement Protocol/SDK Policy, https://developers.google.com/analytics/devguides/collection/protocol/policy (stating rules for upload of data to Google's Measurement Protocol:

> You will not upload any data that allows Google to personally identify an individual (such as certain names, social security numbers, email addresses, or any similar data), or data that permanently identifies a particular device (such as a mobile phone's unique device identifier if such an identifier cannot be reset), even in hashed form.)

[10] Directive 2006/24/EC of the European Parliament and of the Council, March 15, 2006, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:105:0054:0063:EN:PDF.

[11] European Data Protection Supervisor, *Additional EDPS Comments of the Data Protection Reform Package*, March 15, 2013, http://www.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Comments/2013/13-03-15_Comments_dp_package_EN.pdf.

[12] Open DNS, *Knowledge Base: Networks with Dynamic IP Addresses*, http://www.opendns.com/support/article/61 ("The majority of home, small school, and small business networks typically are provisioned by ISPs that issue a dynamic IP address when defining each unique Internet network. Dynamic IP addresses change over time due to a variety of factors primarily centered around the ease of network administration for the ISPs."). There has been concern that this could change with the eventual transition to IPv6, where more IP addresses will be available to ISPs. However, so far, companies implementing IPv6 have committed to the use of privacy extensions that prevent long term user tracking. *See* Alissa Cooper, Center for Democracy & Technology, *Privacy in a Future that is Forever*, June 7, 2012, https://www.cdt.org/blogs/alissa-cooper/0706privacy-future-forever.

The essential element of any regime for pseudonymous data must be that a data controller *cannot* readily tie the data to an individual. It should not be sufficient that a party has the ability to link but does not intend to. For example, if a party hashes a personally-identifiable data set, but retains the cryptographic key it used to create the new data set, without more that data cannot reasonably be deemed pseudonymous.[13] If a controller hashes a data set, but retains the key used to create the hash, the controller (or an individual working for the controller) could undo the deidentification if properly motivated. Similarly, as noted earlier, if a pseudonymous data set could be tied to real name identity by using a third-party lookup service, that data also should not merit lesser protection under the Regulation. The key distinction should be that a controller cannot tie a pseudonymous identifier to a real person — not that it is merely disinclined to at this point in time.

III. What Incentives to Keep Data in Pseudonymous Form?

In deciding how to treat pseudonymous data under the Regulation, the rules will have to balance the desire to incentivize pseudonymous data collection over real-name data collection when possible, while still providing robust protection over what is unquestionably still personal data.

IV. Legal Basis

One potential rule for pseudonymous data could be less rigorous consent requirements for pseudonymous data. Obviously, the right formulation for pseudonymous data is contingent upon the final consent rules and treatment of "legitimate interest" for real identity-linked data sets. CDT has urged that a general requirement that consent be "explicit" is reasonable, but that for some categories of data, the "legitimate interest" justification paired with a robust right to refuse processing is appropriate.[14] It may be the case that pseudonymous data may be able to take advantage of the "legitimate interest" exception more easily, though there still must be a demonstrated interest — it cannot be the case that all pseudonymous collection and usage is deemed legitimate *per se*. Moreover, data subjects must retain the ability to opt-out or refuse processing of pseudonymous data collected under the legitimate interest exception, and if no other legal basis for processing exists, the collector must delete that data under data minimization requirements. For data collected by third parties with which the data subject does not have a direct relationship, some sort of persistent, global automated means (such as "Do Not Track") should be developed to allow users to refuse processing of pseudonymous data by all unknown parties.

V. Controller Obligations

For unauthenticated pseudonymous data sets, it may also be reasonable to excuse data controllers from obligations such as access rights and data portability. Otherwise, for shared devices, one user could obtain from a data controller all information the controller has about the device, potentially compromising the privacy of other users. It would be contrary to the text and spirit of Article 10 of the proposed Regulation to require controllers to collect and authenticate

[13] Ed Felten, Tech@FTC, *Does Hashing Make Data Anonymous?*, April 22, 2012, http://techatftc.wordpress.com/2012/04/22/does-hashing-make-data-anonymous/.

[14] Center for Democracy & Technology, *CDT Analysis of the Proposed Data Protection Regulation*, March 28, 2012, https://www.cdt.org/files/pdfs/CDT-DPR-analysis.pdf.

more precisely identifying information in order to determine whether to comply with an access or portability request for unauthenticated pseudonymous data. However, for pseudonymous accounts (such as a Twitter or other online account) that require users to authenticate with a password, access and portability rights may still be appropriate.

Finally, if a pseudonymous data set has been breached, it may be reasonable to excuse the breaching party from the obligation to notify the data subject. However, controllers cannot necessarily rely on the fact that they have not sought to tie a profile to an individual, as the data set itself may be intrinsically identifying of real individuals. Controllers should have an obligation to evaluate breached data to ensure that it reasonably could not be tied to particular individuals.

About the Center for Democracy and Technology

The Center for Democracy & Technology is a non-profit public interest organization working to keep the Internet open, innovative, and free. With expertise in law, technology, and policy, CDT seeks to enhance free expression and privacy in communications technologies. CDT is dedicated to building consensus among all parties interested in the future of the Internet and other new communications media.

For more information, please contact:

Jens-Henrik Jeppesen, Representative and Director for European Affairs, jjeppesen@cdt.org
Justin Brookman, Director of CDT's Project on Consumer Privacy, jbrookman@cdt.org