October 18, 2017

Department of Homeland Security, Privacy Office

On behalf of the Center for Data Innovation (datainnovation.org), we are pleased to submit comments in response to the Department of Homeland Security's (DHS's) plan to include social media data in its immigration records.[1]

The Center for Data Innovation is the leading think tank studying the intersection of data, technology, and public policy. With staff in Washington, DC, and Brussels, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as predictive analytics, open data, cloud computing, and the Internet of Things. The Center is a non-profit, non-partisan research institute affiliated with the Information Technology and Innovation Foundation.

Social media data may prove to be a valuable screening aid to help verify a person's identity, occupation, living situation, or other factors, and it is understandable why DHS would be interested in collecting this data to aid the immigration process. Indeed, the Center has supported past efforts to explore how social media might enable DHS to expedite its review of travelers to the United States.[2] However, unlike prior proposals, this initiative would not be voluntary, and thus would likely have a significant chilling impact on online speech for those subject to its provisions, as they could be denied entry into the country based on the content of their posts. Moreover, DHS has failed to include any methods for evaluating whether collecting this data is worth the potential harms and costs it would generate for individuals, the federal government, and U.S. businesses. DHS should not move forward with its plan to collect this data until it develops the necessary oversight mechanisms and carefully considers the impact collecting this data could have on the private sector.

---

[1] "Privacy Act of 1974; System of Records," Department of Homeland Security, September 9, 2017, National Science Foundation, January 12, 2017, https://www.federalregister.gov/documents/2017/09/18/2017-19365/privacy-act-of-1974-system-of-records.

[2] Daniel Castro and Joshua New, "Comments on Proposed Information Collections Pursuant to the Paperwork Reduction Act of 1995," Center for Data Innovation, August 22, 2016, http://www2.datainnovation.org/2016-dhs-social-media.pdf.

Sincerely,

Daniel Castro
Director
Center for Data Innovation
dcastro@datainnovation.org

Joshua New
Policy Analyst
Center for Data Innovation
jnew@datainnovation.org

## BACKGROUND

DHS has announced that it plans to modify its system of records for monitoring foreign-born people legally residing in the United States to include "social media handles, aliases, associated identifiable information, and search results" in the immigration records it keeps for each person. DHS plans to collect this data on all immigrants, including naturalized citizens, permanent residents, and green card holders, to support its mission, which could include immigration applicant screening and determining eligibility for federal immigration benefits.

Disappointingly, DHS submitted this plan to the Federal Register on September 9, 2017 and stated that it will go into effect on October 18, 2017, which is also the deadline for public feedback.[3] This timeline is concerning because it indicates that DHS intends to move forward regardless of any potential criticisms or recommendations raised by legal, national security, or technology experts, or by other stakeholders in the public.

## SOCIAL MEDIA DATA CAN BE USEFUL, BUT DHS'S PLAN LACKS NECESSARY OVERSIGHT

An individual's online activities, particularly on social media, can reveal important information about that person. This is one reason why organizations use social media data to make important decisions, such as whether to hire a job applicant, extend a loan to a potential borrower, or admit a student to a university.[4] Similarly, DHS potentially could use social media data such as Twitter activity, Facebook postings, or LinkedIn profiles, to more effectively vet people that wish to live in the United States or receive federal benefits. These online profiles can contain corroborating information about a person's identity, occupation, living situation, and other factors relevant to the immigration process. This is not the first time DHS has recognized the potential of social

---

[3] Ibid.

[4] Kimberlee Morrison, "52% of Recruiters Check Social Media During the Hiring Process [Infographic]," *Social Times*, March 24, 2015, http://www.adweek.com/digital/52-of-recruiters-check-social-media-during-the-hiring-process-infographic/; Stephanie Armour, "Borrowers Hit Social-Media Hurdles," *The Wall Street Journal*, January 8, 2014, http://www.wsj.com/articles/SB10001424052702304773104579266423512930050 ; Emmanuel Felton, "The New Tool Colleges Are Using in Admissions Decisions: Big Data," *PBS NewsHour*, August 21, 2015, http://www.pbs.org/newshour/updates/new-tool-colleges-using-admissions-decisions-big-data/.

media data as a screening aid; it proposed in June 2016 to use social media data in the border screening process.[5]

However, as in the June 2016 proposal, DHS has failed to include any mechanisms for evaluating whether collecting social media data is actually useful in practice, rather than just in theory.[6] DHS did not, nor does it plan to, conduct a pilot program to determine whether including social media data in immigration files can increase agency efficiency, productivity, or effectiveness. Given that DHS plans to implement this proposal immediately after the comment period closes, it is unlikely that DHS would now start with a pilot instead. However, DHS has not included any performance metrics in this proposal either, meaning it will have no way of knowing whether the costs of investing in this data collection are offset by the benefits. Furthermore, DHS does not disclose exactly how it intends to use social media data in the immigration process. For example, there is no indication whether it will use this data exclusively in the application process for immigration or immigration benefits or continuously monitor this data after these processes have completed.

DHS should be particularly conscious about the need to use data to evaluate program effectiveness just as it uses data in other operations: The DHS Office of the Inspector General published a report in February 2017 criticizing its previous attempt to use social media data in immigration screening because it "lack[ed] criteria for measuring performance to ensure they meet their objectives."[7] It is unclear why DHS decided to pursue a similar initiative without taking any of the recommendations of its Inspector General to heed, especially because the report also noted that "neither the private sector nor the U.S. Government possessed the capabilities for large-scale social media screening."[8] Collecting and using social media data effectively poses many challenges. For example, a significant amount of social media data would be unusable because it is posted privately and would not be accessible to the U.S. government (unless the government also demanded users' login credentials). Additionally, there are a wide variety of sites and services that may be considered social media platforms, including social networks, micro-blogging sites, video sharing platforms, commenting tools on websites, gaming

---

[5] Daniel Castro and Joshua New, "Comments on Proposed Information Collections Pursuant to the Paperwork Reduction Act of 1995," Center for Data Innovation, August 22, 2016, http://www2.datainnovation.org/2016-dhs-social-media.pdf.
[6] Ibid.
[7] DHS' Pilots for Social Media Screening Need Increased Rigor to Ensure Scalability and Long-term Success, Department of Homeland Security Office of Inspector General, February 27, 2017, https://www.oig.dhs.gov/sites/default/files/assets/2017/OIG-17-40-Feb17.pdf.
[8] Ibid.

communities, listservs, and dating websites. DHS has not clearly specified which social media platforms it is interested in, and the quality, format, and availability of data across these different platforms will vary widely.

That DHS still intends to collect this data—despite warnings from its Inspector General that it is incapable of carrying out this task—absent any method for performance evaluation is concerning from an accountability standpoint, as well as for the potential negative consequences for the general progress of data innovation.

## CHILLING EFFECT ON DATA INNOVATION

Social media data, which users generate as they create online profiles, have conversations, interact with businesses, and so on, has high value in the private sector. By collecting social media data in the immigration process without a clear understanding of how to use it or whether it will even be worthwhile, DHS threatens to considerably reduce the quantity and value of this data available to the private sector. If people are concerned that providing this data to DHS could negatively impact them, rightly or wrongly, there are two likely outcomes—people will either reduce or cease altogether their social media activity to reduce the amount of data available for scrutiny, or people will self-censor, lie, or otherwise manipulate their statements thereby distorting this information. Either result would diminish the value of this information for commercial purposes or government needs. DHS has no method to ensure that that social media data people provide is accurate, objective, or timely, because social media pages are easily falsified. This phenomenon would be magnified by the fact that any social media data collected about immigrants could incidentally include social media data about other people, natural-born citizens or others, such as their conversations, tagged photos, friend lists, and other interactions. Those people may also be inclined to reduce the amount of social media data they create or to falsify it.

## CONCLUSION

DHS should not move forward with its plan to collect this data until it develops the necessary oversight mechanisms and carefully considers the impact collecting this data could have on the private sector. DHS should take time to develop methods for performance evaluation, preferably through a voluntary pilot program. DHS can analyze these metrics, and if this data collection program proves to be beneficial, then DHS should publicize the results of its analysis and move

forward with scaling up the program. However, DHS should also weigh the potential economic consequences this data collection could have for U.S. businesses that rely on social media data that is both abundant and accurate. Additionally, DHS should clearly state how it intends to use the social media data it collects, as well as whether this would be a one-time collection or continuous monitoring. If the latter, DHS should clearly state the relevance to homeland security this monitoring would have, as it could be both costly and create a larger, ongoing chilling effect on social media participation and use.