

Authors: Stephen Buckner and Zack Schwartz

Trust & Safety team: A 2020 Census Initiative to Mitigate the Impact of Misinformation on Civic Activities

Over the past decade, the exponential increase in use of mobile phones, social media, and digital platforms for communication has reduced the barriers to public civic engagement and thus the ability for an individual to make a significant impact on a civic process. But this phenomenon has likewise lowered the bar to entry for both bad actors and uninformed individuals to spread online information-based threats, including disinformation, targeting civic processes. Government agencies entrusted with carrying out these civic processes, like the Census Bureau, have relished the opportunities this heightened digital engagement has provided to reach a larger portion of the public more easily. But few have proven readily capable to counter and manage the bad that accompanies this good.

In August 2019, the Census Bureau sought to address the risk that information-based threats could impact participation in the 2020 Census by setting up the federal government's first Trust & Safety team. For the past 15 months, the Trust and Safety Team's cross-functional network of experts has conducted 24/7 monitoring of public content to detect rumors, maintained robust partnerships with social media companies to identify and combat rumors, and proactively engaged with the public to both dismiss inaccurate information and communicate accurate information.¹

Strategy

In preparation for 2020, for the first time in nearly 230 years, the Census Bureau planned to introduce a major change to the primary method of census response, shifting from physical paper forms to online and phone response. Bureau officials recognized that this change in procedures combined with the exponential increase in digital engagement since the 2010 Census presented complex challenges, namely that the success of the 2020 Census count would hinge, more than ever, on the public's trust in the Bureau's reputation, security, and brand.²

The **first priority** in protecting the Census process from information-based threats was to **set up an active monitoring program**. The Trust & Safety team integrated state-of-the-art social listening tools to detect and investigate (in near-real time) content of concern, like violent threats to Census employees and misinformation about the use of Census data. The team also implemented advanced algorithms to analyze trends over time and signal if new developments arose that the tools were unable to detect. The Bureau also created a public tip-line (rumors@census.gov) for the members of the American public and other key partners to report incidents that would have otherwise gone unnoticed.

While monitoring and detection are essential elements in the protection of any civic process from mis- and disinformation, it was a tall order for a small team to combat reputational threats and dangerous rumors at the scale required to protect a democratic process. So, our **second priority** was to **establish a robust partnership network** to augment the work of our Trust and Safety team. Our team has relied on a network of more than 300,000 partners across multiple sectors to enhance our defense online and on the ground.³ Our partners included technology companies, civil society fact-check organizations, government agencies, and consumer advocacy groups.

- Technology partners like Facebook, Google, Nextdoor, Twitter, Pinterest, and YouTube established rapid response channels that allowed us to quickly detect and report content. These direct reporting mechanisms enabled our technology partners to more successfully mitigate the

spread of mis- and disinformation on their platforms. Since January 2020, we have reported more than **950** posts that the technology partners further assessed.

- Working with Facebook's certified International Fact-Checking Network (IFCN) enabled our team to more effectively limit the spread of false content. Fact-check organizations rapidly identify false content and refer the information to technology partners which, in turn, apply their policy guidelines to the content in question. Our partners published **17** articles dispelling various census rumors and helped the Census Bureau amplify accurate information directly to the public.
- Finally, our Trust & Safety team also engaged with civil society and consumer advocate organizations to directly reach groups like Hard to Count Communities (HTCs) and other members of the community most vulnerable to scams or privacy concerns.

Our **third priority** to protect the 2020 Census count was **proactive public outreach**. Our team provided customer service by proactively establishing direct communication channels and publishing up to date resources.⁴ Through this approach we utilized intake channels like rumors@census.gov for our team to interact directly with the public and quickly clarify the origins of rumors like the Home Affairs Hoax. In addition to rumors@census.gov, our team hosted Facebook live discussions and Reddit AMAs. Our team also published content like blogposts and pages on our Fighting Rumors page (<https://2020census.gov/en/news-events/rumors.html>) to dispel common rumors, and we worked with partners to update information as operations changed.

Success Stories

Two significant incidents underscore the Trust & Safety team's successful effort to address and mitigate the impact of disinformation.

Department of Home Affairs Rumor: In September of 2019, shortly after establishing the team, Census Bureau staff escalated a tip from the public about a false rumor on WhatsApp. The rumor perpetrated allegations that individuals posing as workers for the "Department of Home Affairs" were going door-to-door to confirm that everyone has a valid ID for the 2020 Census — and committing crimes at homes they are pretending to canvass. Although there is no "Department of Home Affairs" in the United States, individuals sharing the screenshot expressed fear and some made threats against the Census Bureau. The hoax was further perpetuated on public social media platforms, like Facebook and Nextdoor, by members of the public and local organizations. Most concerning was local police departments sharing the warning, unbeknownst that it was a hoax. This gave the rumor credibility in the eyes of the public.



Image of Home Affairs Flyer from WhatsApp

To mitigate the false rumor, our team activated a comprehensive response strategy. 1) We developed language for Fighting Rumors page to set the record straight.⁵ 2) We worked with tech companies to flag content for review and potential removal.⁶ 3) We engaged fact check groups to help us amplify accurate content. 4) We quickly updated our monitoring algorithms with the hoax's language for faster detection. 5) Finally, to address on the ground mitigation, we worked with civil society partners and law enforcement agencies to encourage them to respond to the spread of this rumor.

Stimulus Check Misinformation: The stimulus check rumor is an instance of misinformation that could have arguably increased Census participation, but the falsehoods actually eroded trust in the Census Bureau. Specifically, when Congress developed its first stimulus package in response to the pandemic, a rumor circulated that those who complete their 2020 Census form would receive \$1,200. This falsehood triggered a nearly 300 percent increase in inquiries to Rumors@census.gov.⁷ This rumor was dangerous because it could have led the public to falsely conclude that the Census Bureau was sharing personally identifiable information with other agencies, fueling data privacy concerns. The Trust & safety team debunked the rumor on 2020Census.gov and responded directly to inquiries. Our quick action cut inquiries to rumors@census.gov dramatically.^{8,9}

Lessons Learned

While the Home Affairs Hoax and Stimulus Check rumor are two of the most prominent and persistent threats the Trust & Safety team faced, they represent only a fraction of the thousands of mis- and disinformation incidents that the team has worked to combat. By combining previously disparate Census Bureau groups and integrating their siloed functions, the Trust & Safety team has made the Census Bureau much better equipped to handle the fast-paced and complex threats affecting its operations. These experiences and observations underscore the need for a holistic response to mis- and disinformation. Without a holistic approach, misinformation would spread beyond a team's capacity and capabilities to effectively mitigate the threat reducing trust in government and potentially participation.

We also believe that this example is not unique to the Census Bureau, as many government agencies increasingly face information-based threats that erode the public's confidence in its key functions and ability to carry out its mission. Government agencies engaging in state, local, and federal level civic activities should establish Trust & Safety teams to holistically respond to mis- and disinformation that could impact participation in civic activities and trust in government. These teams should be enterprise-wide, including cyber security teams. While other agencies have similar functions, the Census Bureau's Trust & Safety team provides a blueprint for how to effectively mitigate information-based threats online and on the ground.

¹ Bureau, U. (2019). Trust & Safety Team. Retrieved August 19, 2020, from <https://2020census.gov/en/trust-and-safety.html>

² Pew Research. (2020, June 05). Demographics of Social Media Users and Adoption in the United States. Retrieved August 19, 2020, from <https://www.pewresearch.org/internet/fact-sheet/social-media/>

³ Bureau, U. (2020, February 25). Ambitious 2020 Census Community Partner Goal Reached Ahead of Schedule. Retrieved August 20, 2020, from <https://www.census.gov/newsroom/press-releases/2020/partner-goals.html>

⁴ Bureau, U. (2020, February 25). Ambitious 2020 Census Community Partner Goal Reached Ahead of Schedule. Retrieved August 20, 2020, from <https://www.census.gov/newsroom/press-releases/2020/partner-goals.html>

⁵ Bureau, U. (2019). Are People Posing as "Department of Home Affairs" Workers and Checking IDs for the Census? Retrieved August 19, 2020, from <https://2020census.gov/en/news-events/rumors/home-affairs-hoax.html>

⁶ Schenk, M. (2020, January 30). Fake News: NO Urgent Community Notice, NO Census Security Alert From Department of Home Affairs. Retrieved August 19, 2020, from <https://leadstories.com/hoax-alert/2019/10/fake-news-census-warning-department-of-home-affairs.html>

⁷ Team, Trust & Safety. (2020). Rumors Inbox Maintenance Tracker, 2019-2020 (XLSX version)

⁸ Bureau, U. (2020). Census Responses Not Tied to Potential Stimulus Packages. Retrieved August 19, 2020, from <https://2020census.gov/en/news-events/rumors/stimulus-packages.html>

⁹ Team, Trust & Safety. (2020). Rumors Inbox Maintenance Tracker, 2019-2020 (XLSX version)