

RON WYDEN, OREGON , CHAIRMAN

DEBBIE STABENOW, MICHIGAN	MIKE CRAPO, IDAHO
MARIA CANTWELL, WASHINGTON	CHUCK GRASSLEY, IOWA
ROBERT MENENDEZ, NEW JERSEY	JOHN CORNYN, TEXAS
THOMAS R. CARPER, DELAWARE	JOHN THUNE, SOUTH DAKOTA
BENJAMIN L. CARDIN, MARYLAND	RICHARD BURR, NORTH CAROLINA
SHERROD BROWN, OHIO	ROB PORTMAN, OHIO
MICHAEL F. BENNET, COLORADO	PATRICK J. TOOMEY, PENNSYLVANIA
ROBERT P. CASEY, Jr., PENNSYLVANIA	TIM SCOTT, SOUTH CAROLINA
MARK R. WARNER, VIRGINIA	BILL CASSIDY, LOUISIANA
SHELDON WHITEHOUSE, RHODE ISLAND	JAMES LANKFORD, OKLAHOMA
MAGGIE HASSAN, NEW HAMPSHIRE	STEVE DAINES, MONTANA
CATHERINE CORTEZ MASTO, NEVADA	TODD YOUNG, INDIANA
ELIZABETH WARREN, MASSACHUSETTS	BEN SASSE, NEBRASKA
	JOHN BARRASSO, WYOMING

JOSHUA SHEINKMAN, STAFF DIRECTOR
GREGG RICHARD, REPUBLICAN STAFF DIRECTOR

United States Senate

COMMITTEE ON FINANCE
WASHINGTON, DC 20510-6200

October 4, 2021

The Honorable J. Russell George
Treasury Inspector General for Tax Administration (TIGTA)
1125 15th Street, NW
Suite 700A
Washington, DC 20005

Dear Inspector General George:

On June 8, the Senate Finance Committee held a hearing with IRS Commissioner Rettig titled “The IRS Fiscal Year 2022 Budget,” in which the Commissioner advocated for significant and permanent increases in funding for the IRS, partly on the basis of reducing the tax gap. Immediately prior to the start of the hearing, advocates at ProPublica began disclosing confidential, private, and legally-protected taxpayer information and named particular individuals in an ongoing series of articles. ProPublica claimed that it had obtained a “vast trove of Internal Revenue Service data on the tax returns of thousands of the nation’s wealthiest people, covering more than 15 years.”¹

Despite announcements of various inquiries and investigations, nothing is known about the source of the massive leak or hack. In response to inquiries we made to the IRS Commissioner on August 10 about a major data-security breach at the IRS, the Commissioner responded on September 13, “We do not yet know whether there has been a data breach or a threat of a breach.”

In light of the significant threats to privacy associated with the IRS request for monitoring of taxpayers’ financial accounts, and unknown vulnerabilities at IRS, we write to ask for an investigation into IRS research activities and security protocols surrounding those activities. In particular, the investigation should focus on the activities of the Research, Applied Analytics, and Statistics (RAAS) division of the IRS and the use of contractors.

We believe this audit is warranted for the following reasons. Clearly, taxpayer information is vulnerable, as shown by the apparent leak or hack of confidential information. One longstanding

¹ “The Secret IRS Files: Trove of Never-Before-Seen Records Reveal How the Wealthiest Avoid Income Tax,” ProPublica, June 8, 2021, at <https://www.propublica.org/article/the-secret-irs-files-trove-of-never-before-seen-records-reveal-how-the-wealthiest-avoid-income-tax> .

way in which confidential taxpayer information is utilized is through research done in-house at the IRS, and by working with outside academics and researchers. As long as little is known about the source of the information published by ProPublica, every avenue by which taxpayer data is accessed should be investigated. Even legitimate efforts to anonymize data for research purposes may not adequately protect individuals from being identified by the resulting data. The articles published by ProPublica are themselves a product of unknown individuals performing analysis of large amounts of complicated data, which suggests a possible connection to legitimate research functions. Others in the research community have expressed concern that the leak or hack could have implications for researchers.

Existing IRS Programs Using Taxpayer Information for Research Deserve Scrutiny

As background, on April 13, the Senate Finance Committee held a hearing with IRS Commissioner Rettig titled “The 2021 Tax Filing Season and 21st Century IRS.” At the hearing, the Commissioner speculated that, in his personal opinion, the “tax gap” may be more than \$1 trillion per year. Part of the basis for the Commissioner’s claim of a tax gap significantly larger than previously documented was a largely unvetted National Bureau of Economic Research (NBER) working paper titled, “Tax Evasion at the Top of the Income Distribution: Theory and Evidence.”² The Commissioner cited that paper and, along with additional discussion of the tax gap, argued for significant increases in IRS funding.

The research paper was authored by two permanent IRS employees at RAAS in the IRS, along with professors hired by RAAS as special IRS employees through the Joint Statistical Research Program of the Statistics of Income Division of the IRS, under agreements made possible by the Intragovernmental Personnel Act of 1970 (5 U.S.C. 3371-3376). One of the paper’s authors has not been, to our knowledge, hired as either an IRS employee or under the IRS Joint Statistical Research Program of the Statistics of Income (SOI) division within RAAS at the IRS, yet must have utilized IRS data on individuals, perhaps in anonymized form to protect taxpayer identities.

According to the research paper for which access to private, legally-protected, personal taxpayer data was provided: “All data work for this project involving confidential taxpayer information was done at IRS facilities, on IRS computers, by IRS employees, and at no time was confidential taxpayer data ever outside of the IRS computing environment.”

In a written question for the record submitted pursuant to the Finance Committee’s April 13 hearing, Ranking Member Crapo asked Commissioner Rettig if he could substantiate NBER’s claims that confidential taxpayer information was protected. Commissioner Rettig’s response was: “Yes, all data work for this project involving confidential information was done at IRS facilities, including telework locations approved as part of COVID-19 mitigation, on IRS computers, and by IRS employees. At no time was confidential taxpayer data outside of the IRS computing environment.”

Unfortunately, the response is a mere restatement of the information provided in the question posed to the Commissioner and does not provide the substantiation requested. We do not know what protocols and safeguards are in place, or how they are monitored and documented, to ensure that leaks of troves of confidential taxpayer information do not occur. We do not know

² NBER Working Paper 28542, available at https://www.nber.org/system/files/working_papers/w28542/w28542.pdf.

whether private, legally-protected taxpayer data were in some form shared outside of IRS facilities with authors not employed in some manner by the IRS. TIGTA has examined the RAAS division before and identified issues of concern. A 2018 TIGTA audit report on uses of taxpayer resources in the RAAS division identified that RAAS had “not instituted project management controls” to track research projects.³

Efforts to Anonymize Data May Still Leave Taxpayers Vulnerable

Even when anonymized for research purposes, taxpayer data can be unmasked to detect individual taxpayers. Note that researchers and advocates have been at work, sometimes in collaboration with researchers at the IRS, to refine and expand “synthetic public use files” of confidential administrative tax data.⁴

The notion behind creation of synthetic public-use files is to somehow purge administrative data of anything that could allow a researcher, analyst, advocate, or journalist to associate certain tax-filing characteristics with a particular individual or business. However, as evidenced in a paper posted on the IRS website by researchers affiliated with the Urban-Brookings Tax Policy Center (TPC), the advent and development of “big data,” machine learning, and similar advancements make it increasingly likely that information intended to be anonymized in synthetic files can be unmasked and associated with specific individuals or businesses.⁵ From that, privacy violations and increased potential for political targeting can easily ensue. The TPC-affiliated paper, for example, identifies that:

“Although specific characteristics of individuals may not uniquely distinguish them, combinations of characteristics might. Over time, the scope of publicly available information on individuals has grown, especially through the Internet, and the power of computers and software to link information has also grown. A specific risk is the rising threat of identity theft and data breaches that target and steal individuals’ sensitive data, including the kind of information that might appear on tax returns. These trends significantly increase the likelihood that an individual represented on any microdata file like the PUF [public-use file] might be identifiable, and put all the items in their record at risk of disclosure.”⁶

In fact, ProPublica acknowledges using information available from other sources to supplement the confidential taxpayer information that forms the basis of their reporting. According to ProPublica, “We then verified the information by comparing elements of it with dozens of already public tax details (in court documents, politicians’ financial disclosures and news stories) as well as by vetting it with individuals whose tax information is contained in the trove.”⁷

³ “The Research, Applied Analytics, and Statistics Organization Project Management Practices Need Improvement,” TIGTA, May 4, 2018, Reference Number 2018-10-026, at [The Research, Applied Analytics, and Statistics Organization Project Management Practices Need Improvement \(treasury.gov\)](https://www.tigta.gov/~/media/Files/2018/05/2018-10-026-The-Research-Applied-Analytics-and-Statistics-Organization-Project-Management-Practices-Need-Improvement-treasury.gov.pdf) .

⁴ See, for example, “Creating Homogeneous Synthetic Individual Tax Files for Distributional Analysis,” E. Saez and G. Zucman, August 2018, IRS-SOI at <https://www.irs.gov/pub/irs-soi/18rpsynthindtaxfiles.pdf> , and “Safely Expanding Research Access to Administrative Tax Data: Creating a Synthetic Public Use File and a Validation Server,” L. Burman et al, at <https://www.irs.gov/pub/irs-soi/18resconburman.pdf> .

⁵ “Safely Expanding Research Access to Administrative Tax Data: Creating a Synthetic Public Use File and a Validation Server,” L. Burman et al, at <https://www.irs.gov/pub/irs-soi/18resconburman.pdf> .

⁶ See page 228 of <https://www.irs.gov/pub/irs-soi/18resconburman.pdf> .

⁷ “The Secret IRS Files: Trove of Never-Before-Seen Records Reveal How the Wealthiest Avoid Income Tax,” ProPublica, June 8, 2021, at <https://www.propublica.org/article/the-secret-irs-files-trove-of-never-before-seen-records-reveal-how-the-wealthiest-avoid-income-tax> .

Scrutiny of IRS Research Functions is Merited by the Nature of the ProPublica Articles

Several of the stories published by ProPublica using confidential taxpayer information analyze data from multiple taxpayers. For example, in a story meant to criticize the deduction for pass-through businesses enacted as part of the Tax Cuts and Jobs Act, ProPublica notes, “In the first year after Trump signed the legislation, just 82 ultra-wealthy households collectively walked away with more than \$1 billion in total savings, an analysis of confidential tax records shows.”⁸ A different story suggesting wealthy individuals re-characterized income in order to decrease tax liability notes that, “Secret IRS data shows multiple instances in which salaries for top executives and owners suddenly and inexplicably dropped in the first year after the Trump tax cut, reducing their tax bills even as their companies appeared to thrive. The mysterious pay cuts played out across industries, from logistics companies to real estate firms to makers of bathtubs, and among executives of varying degrees of prominence.”⁹

This level of analysis of complicated taxpayer information from many individuals and across many industries raises the question as to whether the analysis was done strictly by ProPublica or the analysis was performed by someone with legitimate access to the data, and then handed over as a complete package to ProPublica.

Expressing concern for research utilizing IRS data while hinting that such access to IRS data is one avenue that may have facilitated the ProPublica release, researcher and TPC affiliate Dr. Eugene Steuerle writes on the TPC’s TaxVox website that:

“My fear is the *Pro Publica* leak will make gathering, analyzing, and releasing tax information by Treasury and IRS even more difficult. Access to data by outside researchers, including those of us at the Tax Policy Center, could also be further restricted. Some in Congress and at the IRS may limit access both inside and outside the agency because it risks additional leaks. After all, this illegal disclosure couldn’t have happened if someone somewhere hadn’t had access to the data.”¹⁰

A TIGTA Audit is Necessary to Prove that Taxpayer Information is Adequately Protected, and to Demonstrate the Integrity of IRS Research Activities

The vast majority of research conducted by RAAS, and the SOI division of IRS, is scholarly, nonpartisan, and designed to provide improved tax administration and taxpayer service. However, given advancements in technology and data analysis, it is more important than ever that this research doesn’t compromise the basic privacy protections afforded to the taxpayer. In light of the information still being published by ProPublica in a continuing series of articles, it is important the Congress exercise oversight over any avenue by which taxpayer information may

⁸ “Secret IRS Files Reveal How Much the Ultrawealthy Gained by Shaping Trump’s “Big, Beautiful Tax Cut,” ProPublica, August 11, 2021, at <https://www.propublica.org/article/secret-irs-files-reveal-how-much-the-ultrawealthy-gained-by-shaping-trumps-big-beautiful-tax-cut> .

⁹ “How the Trump Tax Law Created a Loophole That Lets Top Executives Net Millions by Slashing Their Own Salaries,” ProPublica, August 19, 2021, at <https://www.propublica.org/article/how-the-trump-tax-law-created-a-loophole-that-lets-top-executives-net-millions-by-slashing-their-own-salaries> .

¹⁰ “ProPublica’s Billionaires: Stories behind the Story, June 15, 2021, TaxVox, at <https://www.taxpolicycenter.org/taxvox/pro-publicas-billionaires-stories-behind-story>.

be put at risk. The current inability of the IRS or the Treasury Department to explain the ProPublica leak or hack and convincingly demonstrate that taxpayer information is safe threatens bipartisan support for IRS research activities.

Given TIGTA's prior identification that RAAS had not shown adequate project management controls, and that the source of the private confidential information that ProPublica continues to publish remains unknown, we ask that TIGTA institute an audit of IRS research activities, and security protocols surrounding those activities, including and particularly regarding the activities of the Research, Applied Analytics, and Statistics division of the IRS and the use of contractors. We additionally ask for responses to the following questions by October 18.

1. Provide a description and documentation of all protocols regarding access to, and use of, private, legally-protected taxpayer data available to researchers specially employed by the IRS under the Joint Statistical Research Program of the Statistics of Income Division of the IRS, under agreements made possible by the Intragovernmental Personnel Act of 1970 (5 U.S.C. 3371-3376). Please include substantiation that research performed for the NBER working paper identified above involving confidential information:
 - a. Was done at IRS facilities, including telework locations approved as part of COVID-19 mitigation, on IRS computers;
 - b. Was done only by IRS employees;
 - c. Was done in ways in which at no time confidential taxpayer data was outside of the IRS computing environment.
2. Provide a detailed description of how IRS monitors usage of confidential data containing personal identifying information on taxpayers by specially employed researchers, such as by the authors of the NBER paper cited in this letter.
3. Identify whether IRS can be confident that any confidential information that may have been anonymized (made "synthetic") cannot be unmasked to identify individuals by big-data experts working within and in conjunction with the IRS.
4. Identify criteria used by IRS to select researchers from outside the IRS to perform research using confidential taxpayer information as special IRS employees, and the extent to which selection of researchers able to work at the IRS using administrative, legally-protected tax data is or is not "some combination of skill and luck."¹¹
5. Identify security measures in place, if any, to ensure confidentiality of legally-protected administrative tax data once those data, in whatever manipulated form, leave secure IRS facilities.
6. Has TIGTA ever identified any misuse or inappropriate disclosure of confidential taxpayer data as a result of any use by the RAAS SOI Division, or any individual with access to IRS data for research purposes?
7. Are there any researchers or research projects with authorized access to taxpayer information that are exploring the same or similar topics to the subjects of any of the articles published by ProPublica? Have researchers conducting any of this research been examined by TIGTA?
8. Have all of the TIGTA recommendations made in TIGTA report 2018-10-026 been successfully resolved? Please identify any other TIGTA recommendations made that apply to RAAS within the last 10 years and note the status of those recommendations.

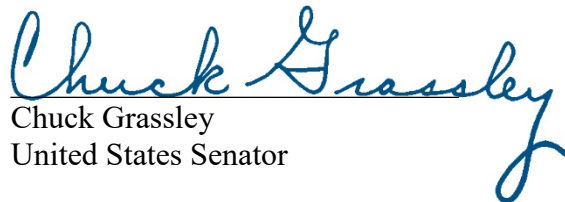
¹¹ See "How Two Economists Got Direct Access to IRS Tax Records," by Jeffrey Mervis, Science, May 22, 2014, at <https://www.sciencemag.org/news/2014/05/how-two-economists-got-direct-access-irs-tax-records>.

9. This letter cites responses from Commissioner Rettig to written questions for the record that note that IRS information was accessed at “telework locations approved as part of COVID-19 mitigation.” Has TIGTA ensured that any telework locations used to access or analyze taxpayer information were secure and that information was not put at risk due to any Covid mitigation effort?
10. Provide a list of all contract researchers hired by RAAS and/or SOI over the past 15 years, including any researchers specially employed by the IRS under the Joint Statistical Research Program of the Statistics of Income Division of the IRS and agreements made possible by the Intragovernmental Personnel Act of 1970 (5 U.S.C. 3371-3376).
11. From 1992 to 2014, SOI compiled and released a separate report on the top 400 taxpayers in the U.S. by adjusted gross income (the Top 400 Report).¹² The Top 400 Report anonymously identified details about their tax situations on average. The Top 400 Report was discontinued after 2014. After 2014, however, similar information was released in a report titled, Individual Income Tax Rates and Tax Shares.¹³ That report includes data on the top 0.001 percent of taxpayers, which, according to the latest version (2018), amounted to 1,443 taxpayers. Because of the similarity between the data reported in the Top 400 Report, and its later iteration, and the data leaked to ProPublica, describe the program that creates Individual Income Tax Rates and Tax Shares, including its history, the security in place to protect the data from internal and external threats, and a general description of who compiled this information, including if there were part-time employees or contractors.

Sincerely,



Mike Crapo
United States Senator



Chuck Grassley
United States Senator

¹² Data on the 400 Individual Income Tax Returns Reporting the Largest Adjusted Gross Incomes, <https://www.irs.gov/statistics/soi-tax-stats-top-400-individual-income-tax-returns-with-the-largest-adjusted-gross-incomes> .

¹³ SOI Tax Stats - Individual Income Tax Rates and Tax Shares, <https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-rates-and-tax-shares> .