The Use of Big Data Analytics by the IRS: Efficient Solutions or the End of Privacy as We Know It?

Kimberly A. Houser and Debra Sanders***

ABSTRACT

This Article examines the privacy issues resulting from the IRS's big data analytics program as well as the potential violations of federal law. Although historically, the IRS chose tax returns to audit based on internal mathematical mistakes or mismatches with third party reports (such as W-2s), the IRS is now engaging in data mining of public and commercial data pools (including social media) and creating highly detailed profiles of taxpayers upon which to run data analytics. This Article argues that current IRS practices, mostly unknown to the general public are violating fair information practices. This lack of transparency and accountability not only violates federal law regarding the government's data collection activities and use of predictive algorithms, but may also result in discrimination. While the potential efficiencies that big data analytics provides may appear to be

* Washington State University—Pullman, Clinical Associate Professor of Business Law, kim.houser@wsu.edu, 509-335-7385.

** Washington State University—Vancouver, Professor of Accounting. Authors thank the hosts of the 2016 Law and Ethics of Big Data Research Colloquium in Bloomington, Indiana; the Center for Business Intelligence and Analytics, Pamplin College of Business, Virginia Tech, and the Department of Legal Studies, Kelley College of Business, Indiana University. Authors particularly thank the participants in the colloquium for their helpful comments: John Bagby, Solon Barocas, Jody Blanke, Deven Desai, Janine Hiller, Dennis Hirsch, Peter Hook, Margaret Hu, Nancy King, David Nersessian, Nizan Geslevich Packin, Abbey Stemler, and Emma Young.

a panacea for the IRS's budget woes, unchecked, these activities are a significant threat to privacy. Other concerns regarding the IRS's entrée into big data are raised including the potential for political targeting, data breaches, and the misuse of such information. This Article intends to bring attention to these privacy concerns and contribute to the academic and policy discussions about the risks presented by the IRS's data collection, mining and analytics activities.

TABLE OF CONTENTS

I. INTRODUCTION

Although tax evasion cost the US government over $3 trillion during the first decade of the 2000s,[1] the Internal Revenue Service (IRS) budget was cut 17% and employees were reduced by 14% in 2010.[2] At the same time, there has been a 7% increase in tax returns filed as well as the passage of two statutes increasing the IRS's workload: the Implement Foreign Account Tax Compliance Act and the Patient Protection and Affordable Care Act.[3] In response, the Office of Compliance Analytics was created in 2011 as a new division of the IRS. The office is charged with developing an advanced analytics program, relying on the use of big data and predictive algorithms to reduce tax fraud.

According to Jeff Butler, the Associate Director of Data Management at the IRS Research, Analysis, and Statistics Organization:

> The IRS uses a wide range of analytic methods, tools, and technologies to address such problems as ID theft, refund fraud, inventory optimization, and other activities related to its statutory mandates. In an era of persistently reduced budgets, the use of data analytics has become more important than ever to drive innovation, risk management, and decision making across the agency.[4]

The IRS uses big data analytics to mine commercial and public data pools including social media sites (e.g., Facebook, Instagram, and Twitter).[5] This data is then added to its proprietary data bases, and

---

1. *Federal Revenue Lost to Tax Evasion*, DEMOS, http://www.demos.org/data-byte/federal-revenue-lost-tax-evasion [https://perma.cc/TMP9-GZB9] (last visited Apr. 9, 2017).

2. Chuck Marr & Cecile Murray, *IRS Funding Cuts Compromise Taxpayer Service and Weaken Enforcement*, CTR. ON BUDGET & POL'Y PRIORITIES (Apr. 4, 2016), http://www.cbpp.org/research/federal-tax/irs-funding-cuts-compromise-taxpayer-service-and-weaken-enforcement [https://perma.cc/W6GU-PBGB].

3. *Id.*

4. Sean Robinson, *Wise Practitioner – Predictive Analytics Interview Series: Jeff Butler at IRS Research, Analysis, and Statistics Organization,* PREDICTIVE ANALYTICS TIMES (Sept. 2, 2015), http://www.predictiveanalyticsworld.com/patimes/wise-practitioner-predictive-analytics-interview-series-jeff-butler-at-irs-research-analysis-and-statistics-organization09022015/6243/ [https://perma.cc/9KPH-94PB].

5. Dara Kerr, *Tax Dodgers Beware: IRS Could Be Watching Your Social Media,* CNET (Apr. 15, 2014), http://www.cnet.com/news/tax-dodgers-beware-irs-could-be-watching-your-social-media/ [https://perma.cc/2UFZ-GJTB]; *see also* Tim Sampson, *FYI, the IRS Is Looking at Your Online Activity for Signs of Tax Evasion,* DAILY DOT (Apr. 16, 2014), http://www.dailydot.com/news/irs-social-media-tax-evasion/ [https://perma.cc/F33W-M9FL]; *Report: IRS Data Mining Facebook, Twitter, Instagram and Other Social Media Sites*, CBSDC

pattern recognition algorithms are run to identify potential noncompliant taxpayers.[6] Data analytics has proven to be a useful tool in successfully identifying fraud victims, and, according to the IRS, computer identification of noncompliant taxpayers is less subjective than other methods.[7] However, the IRS is less forthcoming about its use of data analytics in deciding whom to audit; the decision is based on private, highly detailed profiles of each US taxpayer, created from sources other than the taxpayer's returns and third party reports.[8] Also, the question remains as to whether the data upon which algorithms rely is accurate and if the algorithms themselves may result in discrimination. Overall, the collection and use of this data without proper oversight and the increasing reliance on machine generated decisions may result in harm.

This Article will explore a number of potential issues pertaining to the IRS's use of big data and predictive algorithms. Part II explains data collection by the IRS, the history of improper audits, and how the IRS selects returns for audit. Part III outlines the legal issues raised by the IRS's data collection activities and their use of predictive analytics. Part IV discusses the potential for misuse of data and algorithms by the IRS. Part V provides the conclusion.

II. THE IRS

The IRS is the branch of the United States Department of Treasury that is responsible for administering the Internal Revenue Code and enforcing tax law.[9] Income taxes were introduced to the United States in 1913 when the Sixteenth Amendment was enacted.[10] While the Treasury Department collects the taxes, the IRS is responsible for examining the tax returns for accuracy and bringing criminal action against those who file incorrect returns.[11] Each tax return is checked internally for mathematical accuracy and consistency, regardless of whether it is submitted via mail or

(Apr. 16, 2014), http://washington.cbslocal.com/2014/04/16/report-irs-data-mining-facebook-twitter-instagram-and-other-social-media-sites/ [https://perma.cc/8G4W-GEZ3].

6. Kerr, *supra* note 5.

7. *See* Robinson, *supra* note 4.

8. The IRS has released very little information about the Office of Compliance Analytics. *Id.*

9. *Internal Revenue Service*, USA.GOV, https://www.usa.gov/federal-agencies/internal-revenue-service [https://perma.cc/ZEW8-8JEN] (last visited Mar. 5, 2017).

10. *A Brief History*, IRS, https://www.irs.gov/uac/brief-history-of-irs [https://perma.cc/SR5R-VAFF] (last visited Mar. 5, 2017).

11. *The Agency, Its Mission and Statutory Authority*, IRS, https://www.irs.gov/uac/the-agency-its-mission-and-statutory-authority [https://perma.cc/L4V9-W3NV] (last updated July 27, 2016).

electronically.[12] The IRS also compares the submitted returns to third-party materials that are required to be filed with the IRS, such as W-2s and 1099s.[13] Today the IRS is taking advantage of the large amount of data that can be purchased from data brokers as well as amassing its own data sets.[14]

A. IRS Data Collection

Prior to discussing the potential issues with the IRS's use of data analytics, it is important to understand what data it is collecting and from where it is collecting that data. While the IRS may request information from taxpayers to support the information provided on their tax returns,[15] individuals are having to consider the constitutionality of the IRS collecting and maintaining information on taxpayers from sources other than the taxpayer and *prior* to an audit. Even though a taxpayer is required to maintain the proof necessary to support any line item on a tax return, the taxpayer need not provide support along with her return, nor would she need to support an allowed deduction, such as the payment of mortgage interest, if she instead chose to take the standard deduction or simply not take the deduction at all.[16] While the burden is on the taxpayer to support their return, the IRS does not have unlimited power to obtain any data it desires regarding a taxpayer.

It is well known that the IRS is able to obtain information from third parties to verify line items on tax returns provided by taxpayers.[17] An example would be a W-2 from an employer. However, the right to third party information is not unlimited. Only recently have privacy scholars begun to examine these issues when it comes to electronic and phone communications.[18] Most of the rules permitting the IRS to obtain records from third parties were written prior to the existence of social media, and certainly prior to the current state of technology. "Modern technologies are creating 'minutely detailed records' of our existence, increasingly facilitating the 'persistent,

12. William J. Hunter & Michael A. Nelson, *An IRS Production Function*, 49 NAT'L TAX J. 105, 105–15 (1996).

13. *Id.*

14. Robinson, *supra* note 4; *see also National Research Program (NRP)*, IRS (Aug. 18, 2012), https://www.irs.gov/uac /national-research-program-nrp [https://perma.cc/X8HP-SGC2].

15. I.R.C. § 7602(a) (2016). The IRS may also investigate sources of income from those who fail to file any tax return. *See* I.R.C. § 6651 (2016).

16. *See* Beatty v. Comm'r, 40 T.C.M. (CCH) 438 (1980).

17. I.R.C. § 7602(a).

18. Jonathan P. West & James S. Bowman, *Electronic Surveillance at Work*, 48 ADMIN. & SOC'Y 628, 628 (2014).

continuous and indiscriminate monitoring of our daily lives.'"[19] The existence of data brokers and the ability to purchase information about pretty much anyone over the Internet has created a situation where users are losing control over who sees their once private information.[20] This is especially unsettling when that viewer is the IRS.

1. Phone Records

According to the American Civil Liberties Union (ACLU), the IRS is one of the agencies that purchased cell phone tracking technology in 2009 and 2012.[21] This phone tracking technology, known as Stingray, masks as a cell tower to trap metadata and content from cell phones that connect to them. This technology means the IRS has the ability to record phone conversations, text messages, and track the location of individuals using their cell phones without anyone being aware of this tracking.[22] Legal scholars believe that the IRS will increasingly rely on surveillance technology to reduce noncompliance.[23] A case is currently being heard in Maryland regarding the constitutionality of the government's use of the Stingray cell tracking device.[24] Although in 2015 the Department of Justice (DOJ) issued a guidance statement for the department's law enforcement constituents,[25] these guidelines do not apply to the IRS.[26]

---

19. Michael Hatfield, *Taxation and Surveillance: An Agenda,* 17 YALE J.L. & TECH. 319, 322 (2015) (quoting Danielle Keats Citron & David Gray, *Addressing the Harm of Total Surveillance: A Reply to Professor Neil Richards*, 126 HARV. L. REV. F. 262 (2013) and Neil M. Richards, *The Dangers of Surveillance*, 126 HARV. L. REV. 1934 (2013)).

20. *See* Mark Hachman, *The Price of Free: How Apple, Facebook, Microsoft and Google Sell You to Advertisers*, PC WORLD (Oct. 1, 2015), http://www.pcworld.com/article/2986988/privacy/the-price-of-free-how-apple-facebook-microsoft-and-google-sell-you-to-advertisers.html [https://perma.cc/2VX6-PKFF].

21. Steve Straehley & Danny Biederman, *Even the IRS Has Spied on American Citizens*, ALLGOV (Oct. 29, 2015), http://www.allgov.com/news/top-stories/even-the-irs-has-spied-on-american-citizens-151029?news=857740 [https://perma.cc/CT98-HX79].

22. Kay Bell, *IRS Using Cellphone Scrapers to Gather Data*, BANKRATE (Oct. 27, 2015), http://www.bankrate.com/financing/taxes/irs-using-cell-phone-scrapers-to-gather-data/#ixzz4JhlKj7p5 [https://perma.cc/9P83-BZQZ].

23. Hatfield, *supra* note 19, at 337.

24. Rebecca McRay, *A Lawsuit Could Rein in the Government's Use of Secret Surveillance Tools,* TAKEPART (Feb. 7, 2016), http://www.takepart.com/article/2016/02/07/stingray-lawsuits-maryland [https://perma.cc/ZE57-CVAV]; *see also EPIC v. FBI - Stingray / Cell Site Simulator*, ELECTRONIC PRIVACY INFO. CTR., http://epic.org/foia/fbi/stingray/ [https://perma.cc/M2BR-MLJ8] (last visited Mar. 22, 2017).

25. Kelly Phillips Erb, *IRS Joins FBI, DEA & Other Federal Agencies with Access to Cellphone Surveillance Technology*, FORBES (Oct. 26, 2015), http://www.forbes.com/sites/kellyphillipserb/2015/10/26/irs-joins-fbi-dea-other-federal-agencies-with-access-to-cellphone-surveillance-technology/#2da5c5b77377 [https://perma.cc/AEB2-KFNL].

2. Emails

Pursuant to a Freedom of Information Act (FOIA) request in 2013, the ACLU discovered that the IRS had been reading taxpayers' private emails without a warrant.[27] The 2011 IRS auditor's training manual indicated that investigators could obtain everything in an account using an Electronic Communications Privacy Act (ECPA) court order except for unopened email or voicemail stored with a provider for 180 days or less.[28] This policy is in direct contravention of the 2010 ruling in *United States v. Warshak*, which reaffirmed that citizens have a reasonable expectation of privacy in their emails and that the government needs a warrant to obtain them.[29] It should be noted that an ECPA court order can be issued fairly easily and does not require "probable cause" that a criminal statute has been violated.[30] In response to a Senate Finance Committee hearing, the IRS agreed to stop reading taxpayers' emails without a warrant[31] but was notably silent about its social media activities.

3. Social Media

According to a spokesperson for the UC-Berkeley Samuelson Clinic, the IRS confirmed in response to a FOIA request that it is collecting information from social media sites.[32] An IRS training document mentions Facebook, MySpace, and YouTube as possible

---

26. The IRS falls under the purview of the Treasury Department, not the Department of Justice. *The Agency, Its Mission and Statutory Authority*, *supra* note 11.

27. Nathan F. Wessler, *New Document Suggests IRS Reads Emails Without a Warrant*, ACLU (Apr. 10, 2013), https://www.aclu.org/blog/new-documents-suggest-irs-reads-emails-without-warrant?redirect=blog/technology-and-liberty-national-security/new-documents-suggest-irs-reads-emails-without-warrant [https://perma.cc/3SYX-CGQT].

28. INTERNAL REVENUE SERV., MANUAL TRANSMITTAL 9.4.9, § 9.4.9.5.3.4(1) (2011), https://www.aclu.org/legal-document/manual-transmittal-re-irm-949?redirect=national-security/manual-transmittal-re-irm-949 [https://perma.cc/4RXL-AG84]. For unopened email or voicemail stored with a provider for 180 days or less, the manual did indicate that a warrant was required. *Id.* § 9.4.9.5.3.4(1), (2).

29. United States v. Warshak, 631 F.3d 266, 274 (6th Cir. 2010).

30. I.R.C. § 2702(d) (2016) only requires facts that show the content is relevant to an ongoing investigation.

31. Tim Sampson, *IRS Reverses Course on Warrantless Email Snooping*, DAILY DOT (Apr. 17, 2013 4:40 PM), http://www.dailydot.com/news/irs-email-warrantless-snooping-reversal/ [https://perma.cc/S54F-YZ8K].

32. Jaikumar Vijayan, *IRS, DOJ Use Social Media Sites to Track Deadbeats, Criminal Activity*, COMPUTER WORLD (May 16, 2010), http://www.computerworld.com/article/2516372/web-apps/irs—doj-use-social-media-sites-to-track-deadbeats—criminal-activity.html [https://perma.cc/L9GA-JSKY].

sources for taxpayer information.[33] According to CNET, the IRS uses "online activity trackers to look though mass amounts of public Internet data for potentially incriminating information."[34] The IRS has also used evidence from Google Maps in a Tax Court case to revoke the 501(c)(4) tax exempt status of a homeowners' association.[35] There is, of course, a difference between locating publicly available information online about a taxpayer who is being audited and data mining for potential tax violators prior to the time the taxpayer has been selected for an audit. The IRS is reported to have used automated computer programs (sometimes known as spiders) to sort through social media sites.[36]

4. Data Mining

Data mining involves the analysis of large data sets, which have been collected for a purpose other than that for which they are being analyzed,[37] in order to search the data sets for previously unknown relationships in the data.[38] Data mining can be descriptive or predictive: descriptive data mining summarizes properties of the data set,[39] while predictive data mining performs analysis on a data set to build a model that makes predictions about data that is not available.[40] The IRS engages in data mining in order to develop analytics and algorithms to identify tax compliance issues.[41] According to an IRS report: "It is not possible to have compliance experts review every possible set of related tax returns. . . . Active learning can be used to refine targeting models. Common connections between possibly abusive transactions can be used to identify

33. *IRT-WBT Content 2009*, ELECTRONIC FRONTIER FOUND. 12 (2009), https://www.eff.org/files/filenode/social_network/training_course.pdf [https://perma.cc/ER9P-R6TM].

34. Sampson, *supra* note 5.

35. *Id.*

36. *Report: IRS Data Mining Facebook, Twitter, Instagram and Other Social Media Sites*, *supra* note 5.

37. DAVID HAND, HEIKKI MANNILA & PADHRAIC SMYTH, PRINCIPLES OF DATA MINING 1 (2001).

38. RAMESH SHARDA, DURSUN DELEN & EFRAIM TURBAN, DECISION SUPPORT AND BUSINESS INTELLIGENCE SYSTEMS 680 (9th ed. 2011).

39. JAIWEI HAN & MICHELINE M. KAMBER, DATA MINING: CONCEPTS AND TECHNIQUES 15 (2d ed. 2006).

40. Michael Wu, *Big Data Reduction 2: Understanding Predictive Analytics*, LITHIUM (Mar. 25, 2013), http://community.lithium.com/t5/Science-of-Social-blog/Big-Data-Reduction-2-Understanding-Predictive-Analytics/ba-p/79616 [https://perma.cc/MZV7-Q966].

41. DAVID DEBARR & MAURY HARWOOD, RELATIONAL MINING FOR COMPLIANCE RISK 175 (2004), https://www.irs.gov/pub/irs-soi/04debarr.pdf [https://perma.cc/6AP2-9PCP].

potential promoters of these transactions."[42] Sources have disclosed that the IRS is using data mining to create more detailed profiles of taxpayers.[43] "If Nike is analyzing my information, the worst consequence is that they market stuff to me that I don't want and it's annoying," stated Behnam Dayanim, co-chair of the privacy and data practice at Paul Hastings, "[i]f the government does it, the worst consequence is there could be legal ramifications, whether it's fines, penalties, or imprisonment."[44] Concerns about agency use of data mining were also discussed in the Senate hearings regarding the Federal Agency Data Mining Reporting Act of 2007.[45] The hearing report indicated that there were 199 different government data mining programs, including the IRS, and that there was very little control over these activities.[46]

B. History of Improper Audits

One of the concerns with the IRS's unprecedented access to private information is the IRS's history of misusing the audit function. An audit is intended to ensure tax compliance; the IRS audits returns to check for mathematical errors, document mismatching, and noncompliance.[47] However, since the creation of the IRS, government

---

42. *Id.* at 183.

43. Richard Satran, *IRS High-Tech Tools Track Your Digital Footprints*, U.S. NEWS & WORLD REP. (Apr. 4, 2013), http://money.usnews.com/money/personal-finance/mutual-funds/articles/2013/04/04/irs-high-tech-tools-track-your-digital-footprints [https://perma.cc/WUC5-86TJ]; *see also* Stacey Vanek Smith, *When the IRS 'Likes' Your Facebook Update*, MINN. PUB. RADIO: MARKETPLACE (Apr. 14, 2014), https://www.marketplace.org/2014/04/14/economy/when-irs-likes-your-facebook-update [https://perma.cc/MZV7-Q966].

44. Smith, *supra* note 43.

45. *Balancing Privacy and Security: The Privacy Implications of Government Data Mining Programs: Hearing Before the S. Comm. on the Judiciary*, 110th Cong. (2007), https://www.gpo.gov/fdsys/pkg/CHRG-110shrg33226/html/CHRG-110shrg33226.htm [https://perma.cc/WK2T-BURL].

46. *Id.* According to the 2015 Annual Privacy and Data Mining Report, the IRS indicates that it consolidates two reporting requirements to provide Congress and the public with a more comprehensive overview of the Treasury's privacy compliance and oversight activities: (1) The annual privacy report required by Section 522(a) of the Consolidated Appropriations Act of 2005; and (2) the Data Mining Reporting Act requirement contained in Section 803 of the Implementing Recommendations of the 9/11 Commission Act of 2007, 42 U.S.C. § 2000ee–3. DEP'T OF THE TREASURY, 2015 ANNUAL PRIVACY AND DATA MINING REPORT 5, 7 (2015), https://www.treasury.gov/privacy/annual-reports/Documents/Annual%20Privacy%20and%20Data%20Mining%20Report%20Fiscal%20Year%202015.pdf [https://perma.cc/WGJ4-EMT4] (alleging that everything they do complies with the law but not detailing the exact nature of their data mining activities, which would seem to be the purpose of requiring the report).

47. *IRS Audits*, IRS, https://www.irs.gov/Businesses/Small-Businesses-&-Self-Employed/IRS-Audits [https://perma.cc/5R48-W2KL] (last updated Mar. 23, 2017).

officials, particularly presidents, have been using the IRS for their own political agendas.[48]

President Franklin D. Roosevelt set the stage for presidents using the IRS as a weapon to investigate political rivals and business opponents.[49] Roosevelt's victims included Senator Huey Long, United Mine Workers leader John Lewis, Representative Hamilton Fish, Chicago Tribune publisher Robert "Colonel" McCormick, Philadelphia Inquirer publisher Moses Annenberg, William Randolph Hearst, Father Charles Coughlin, and, the former Treasury Secretary Andrew Mellon.[50] The latter is especially ironic because Andrew Mellon utilized the IRS to audit his rivals as Treasury Secretary under President Calvin Coolidge.[51]

Between 1956 and 1971, the FBI ran a counterintelligence program called COINTELPRO.[52] The brainchild of then-FBI Director J. Edgar Hoover, its purpose was initially to disrupt, discredit, and destroy Communist Party activities in the United States.[53] It later expanded to include other groups such as the Socialist Workers Party and the Black Panther Party.[54] Under COINTELPRO, the FBI was able to harass these individuals and organizations by having the IRS target them for tax audits.[55] Martin Luther King Jr. was a victim of this harassment, as was the National Association for the Advancement of Colored People and the National Council of Churches.[56]

When Robert Kennedy, chief counsel for the Senate Select Committee on Improper Activities in Labor and Management, investigated Teamsters leader Jimmy Hoffa in the late 1950s for illegal activities, he failed to obtain a conviction.[57] When he was the

---

48. Gail Chaddoci, *Playing the IRS Card: Six Presidents Who Used the IRS to Bash Political Foes*, CHRISTIAN SCI. MONITOR (May 17, 2013), http://www.csmonitor.com/USA/Politics/DC-Decoder/2013/0517/Playing-the-IRS-card-Six-presidents-who-used-the-IRS-to-bash-political-foes/ [https://perma.cc/F4KU-P5WL].

49. *Id.*

50. *Id.*; *see also The IRS's Long History of Scandal*, WEEK (June 8, 2013), http://theweek.com/articles/463448/irss-long-history-scandal [https://perma.cc/4G4B-TKJB].

51. Chaddoci, *supra* note 48.

52. *FBI Records: The Vault*, FBI, https://vault.fbi.gov/cointel-pro [https://perma.cc/WQE9-EH2X] (last visited Mar. 6, 2017).

53. *Id.*

54. *Id.*

55. OMAR V. GARRISON, *COINTELPRO Revisited—"A Rough, Tough Dirty Business"*, *in* PLAYING DIRTY: THE SECRET WAR AGAINST BELIEFS 53–73 (1980), http://www.whatreallyhappened.com/RANCHO/POLITICS/COINTELPRO/fbi_cofs.html [https://perma.cc/J3E4-DL6H].

56. *The IRS's Long History of Scandal*, *supra* note 50.

57. Chaddoci, *supra* note 48.

Attorney General under President John F. Kennedy, Robert again sought out Jimmy Hoffa.[58] One of the tactics used was requesting the IRS to repeatedly examine his returns and those of his associates for tax evasion.[59] This tactic was also employed for other alleged racketeers whom Kennedy had his eye on.[60] This targeting of those believed to be involved in criminal activity raised questions from legal experts who decried that tax laws are for revenue collection, not prosecuting criminals, and insisted that audits should be random.[61] Under the Kennedy administration, IRS investigations extended to groups with extreme conservative views such as the John Birch Society.[62] The IRS went so far as to establish the "Ideological Organizations Audit Project" to target these groups.[63]

The President who really excelled at wielding the IRS audit weapon against political enemies was Nixon. Besides targeting left-wing groups, Nixon sought out antiwar groups, churches and nonprofits supporting antiwar groups, civil rights groups, reporters, and prominent Democrats.[64] The White House tapes provide direct evidence of Nixon using the IRS to collect data on potential Democratic presidential candidates, including Senators Hubert Humphrey, Edward (Ted) Kennedy, and Edmund (Ed) Muskie.[65] Nixon had the IRS establish the Special Service Staff unit to utilize tax records to create dossiers on more than 11,000 individuals and groups, including supporters of Democrat Presidential nominee George McGovern for 1972.[66] In the House Judiciary Committee's 1974 Articles of Impeachment, one of the articles charged Nixon with trying to obtain confidential information contained in income tax returns for purposes not authorized by law, in violation of the taxpayer's constitutional rights, and causing the selection of audits in a discriminatory manner.[67]

---

58. *Id.*

59. James Kelly, *The Prince and the Pauper*, WALL STREET J. (Aug. 7, 2015), http://www.wsj.com/articles/the-prince-and-the-pauper-1438979666 [https://perma.cc/D9UG-UQK9].

60. Chaddoci, *supra* note 48.

61. *Id.*

62. *The IRS's Long History of Scandal*, *supra* note 50.

63. *Id.*

64. Chaddoci, *supra* note 48.

65. *Id.*

66. *The Nixon Administration and Watergate: Political Subordination of IRS*, HIST. COMMONS, http://www.historycommons.org/timeline.jsp?nixon_and_watergate_tmln_watergate_campaign_conspiracy=nixon_and_watergate_tmln_political_subordination_of_irs&timeline=nixon_and_watergate_tmln [https://perma.cc/3YYD-5ACS] (last visited Mar. 6, 2017).

67. Chaddoci, *supra* note 48; *see also The IRS's Long History of Scandal*, *supra* note 50.

There have also been IRS abuses by presidents subsequent to Nixon. Recently, the IRS singled out conservative organizations with "tea party" affiliations that were seeking tax-exempt nonprofit status and subjecting them to extra scrutiny.[68] While there may or may not have been a political motivation for the increased scrutiny, it does appear that at the very least gross mismanagement was involved.[69] Despite the Department of Justice's finding that no criminal conduct occurred, the House Ways and Means Committee has indicated that it will continue to investigate the targeting.[70]

C. Audit Selection History

Every year the IRS must shift through copious numbers of taxpayer returns and their related data. In order to ensure tax compliance, the IRS may audit a tax return to check for mathematical errors, document mismatching, and noncompliance.[71] The audit may be performed through the mail, at the taxpayer's home, or at an IRS office.[72] Historically, tax returns were selected randomly (based on a statistical formula), due to a mismatch with third party data, or when a return was linked to other taxpayers who were being audited themselves.[73] The majority of audits resulted from mismatches with third party data.[74] The Information Returns Processing (IRP) System was responsible for the data received from employers and other third parties reporting taxpayer income, pensions, interest, and dividends paid during the tax year.[75] The IRP would match income reported on information returns against income reported by taxpayers on their individual income tax returns based on Social Security numbers. When mathematical errors, inconsistencies, or a mismatch in the IRP system was identified, the taxpayer was contacted via mail and a bill or check was sent to the taxpayer.[76]

---

68. Andy Kroll, *The IRS Tea Party Scandal, Explained*, MOTHER JONES (Nov. 21, 2013), http://www.motherjones.com/politics/2013/05/irs-tea-party-scandal-congress-nonprofit-obama [https://perma.cc/TDK5-BLQB].

69. Evan Perez, *First on CNN: DOJ Closes IRS Investigation with No Charges*, CNN POL. (Oct. 23, 2015), http://edition.cnn.com/2015/10/23/politics/lois-lerner-no-charges-doj-tea-party/ [https://perma.cc/8HME-9M54].

70. *Id.*

71. *IRS Audits*, *supra* note 47.

72. *Id.*

73. *Id.*

74. *Id.*

75. *Information Returns Processing*, IRS (Nov. 16, 2016), https://www.irs.gov/uac/information-returns-processing [https://perma.cc/M6VG-ZQYB].

76. *Id.*

The rate of audits increased during the 1950s.[77] In the early 1960s over 5.5 percent of tax returns were chosen for audits.[78] The rate then began declining due to technological advancements in identifying potential tax returns to audit. The IRS first used computers for selecting tax returns in 1962 and created the Taxpayer Compliance Measurement Program (TCMP) two years later.[79] The TCMP randomly selected about 50,000 returns approximately every three years to perform detailed audits requiring substantiation of each line on the tax return.[80] This program initially reviewed delinquent returns to create a statistical summary, which evolved into an automated program known as the discriminant function analysis (DIF).[81] The DIF gives each tax return a score based on the probability of noncompliance. IRS personnel then manually screen the tax returns to ensure appropriate selection.[82] This process enhanced audit efficiency by allowing the IRS to manually review the machine scored returns and chose the tax returns with the highest likelihood of noncompliance while avoiding auditing compliant returns.[83] The first tax audits based on the DIF occurred in 1969, and refinements to the DIF were made during the 1970s and again in the 1980s with the addition of computerized third party document matching and mathematical accuracy.[84] Analysis of the most common errors by taxpayers led to policy changes. For example, in 1986 the TCMP identified a significant misreporting of dependency exemptions and wrongful claims of the earned income credit by individuals claiming children that did not qualify. The policy was then changed to require identification numbers for dependents. As a result, the number of dependents claimed in 1987 was 7 million fewer than

77. Hunter & Nelson, *supra* note 12, at 105–15.

78. *Id.*

79. *Id.*

80. JAMES ALM, DESIGNING RESPONSIBLE REGULATORY POLICIES TO ENCOURAGE TAX COMPLIANCE 8 (2013), http://murphy.tulane.edu/files/events/Alm-DesigningResponsible RegulatoryPolicies-MurphyInstitute-021113.pdf [https://perma.cc/5YSD-8DQH].

81. This is the first use of data analytics by the IRS, and it relied on data contained in the tax returns provided by the taxpayers. *Id.*

82. INTERNAL REVENUE SERV., IRS PUBLICATION 556, EXAMINATION OF RETURNS, APPEAL RIGHTS, AND CLAIMS FOR REFUND 2 (2013); *see also How Tax Returns Are Selected for Audit: Explaining DIF Scores and UI DIF Scores*, BROTMAN L., http://info.sambrotman.com/blog/how-tax-returns-are-selected-for-audit/ [https://perma.cc/SU89-2RSC] (last visited Mar. 7, 2017).

83. Personal Communication with Keith Nelson, Former Criminal Investigator, Internal Revenue Serv. (Mar. 16, 2016).

84. Hunter & Nelson, *supra* note 12, at 105–15.

claimed in 1986 when identification numbers were not required.[85] A similar decrease was found for those claiming an earned income credit.[86] In the 1980s, when third party reporting became required of income items such as wages, interest, and dividends, the accuracy of these amounts on tax returns substantially increased.[87]

Prior to the time the TCMP audit was used, only half of the audited returns found any errors.[88] During the time of the TCMP audits, the percentage of returns chosen for audits containing no errors (no-change) decreased from over 40 percent in 1968 to about 11 percent in the early 1990s.[89] However, TCMP audits were onerous because they required the taxpayer to support each line of their tax return with documentation.[90] In 1988, the TCMP was eventually phased out due to cuts in the IRS budget and criticisms by taxpayers, Congress, and the media.[91] The 1996 General Accounting Office (GAO) report on the IRS, suggested that the IRS find alternative methods (to the TCMP) for updating the DIF and develop a long-term strategy for obtaining compliance data with fewer resources.[92] Supporting the GAO's predictions of the detrimental effects of not updating the TCMP, the 1994 no-change rate for individual returns identified by the DIF was over 19 percent, and more than 24 percent

85. Jeffrey B. Liebman, *Who Are the Ineligible EITC Recipients?*, 53 NAT'L TAX J. 1165, 1171 (2000).

86. *Id.*

87. Jeffrey Dubin, Michael Graetz & Luis L. Wilde, *The Effect of Audit Rates on the Federal Individual Income Tax, 1977-1986*, 43 NAT'L TAX J. 395, 397 (1990).

88. Hunter & Nelson, *supra* note 12, at 105–15.

89. *Id.*

90. David Turner, *Taxpayers Beware of 'Audit from Hell'*, ORLANDO SENTINEL (Mar. 12, 1995), http://articles.orlandosentinel.com/1995-03-12/business/9503100011_1_irs-audit-taxpayer-compliance-measurement-regular-audit [https://perma.cc/3ZUH-L39W]. The IRS eventually concluded that the TCMP was too costly, burdensome, and time consuming. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-02-769, TAX ADMINISTRATION: NEW COMPLIANCE RESEARCH EFFORT IS ON TRACK, BUT IMPORTANT WORK REMAINS 4 (2002), http://www.gao.gov/assets/240/234955.pdf [https://perma.cc/MLF6-HPH4]. However, the General Accounting Office (GAO) determined that limiting the scope of the TCMP was unjustified and would undermine its benefits. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-95-39, TAX COMPLIANCE: STATUS OF THE TAX YEAR 1994 COMPLIANCE MEASUREMENT PROGRAM 1–2 (1994). The TCMP actually lessened the overall burden by decreasing the number of compliant taxpayers being audited. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-02-769, *supra*, at 5. The 1995 TCMP was planned to be the most comprehensive review, with over 150,000 returns audited covering individuals and small businesses and include more computerized analysis. *Id.* at 1, 4.

91. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-02-769, *supra* note 90, at 1, 4; *see also* U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-96-89, TAX ADMINISTRATION: ALTERNATIVE STRATEGIES TO OBTAIN COMPLIANCE (1996), http://www.gao.gov/assets/230/222435.pdf [https://perma.cc/G3SL-BFAA].

92. *Id.*

in 1998.[93] By not updating the data upon which the DIF was based, a greater percentage of compliant taxpayers had to suffer through the expense and stress of an audit.[94] Based on concern that the effectiveness of the DIF was deteriorating and reducing taxpayer confidence in the fairness of the tax system, in 2002 the IRS initiated the National Research Program (NRP) to replace the TCMP.[95] The idea was to increase the quality of the data and better predict which tax returns would result in a deficiency without the burdensome TCMP audits.[96]

The NRP gathers data learned from random audits to measure voluntary compliance with tax laws and improve DIF audit selection methods.[97] This data is used for analytical purposes such as identifying tax issues, reporting characteristics, and taxpayer segments that may lead to noncompliant behavior.[98] The rationale for performing this type of analytics is that efficiency in the audit process reduces unnecessary audits for compliant taxpayers. Today, the IRS uses big data analytics to target their audits on tax returns more likely to result in tax deficiencies.[99]

---

93. David Blattner & Robert Johnson, *IRS National Research Program*, 4 J. TAX PRAC. & PROC. 9, 9 (2002).

94. *Id.*

95. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-07-423, USING DATA FROM THE INTERNAL REVENUE SERVICE'S NATIONAL RESEARCH PROGRAM TO IDENTIFY POTENTIAL OPPORTUNITIES TO REDUCE THE TAX GAP 1–2 (2007); *see also* U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-03-614, TAX ADMINISTRATION: IRS IS IMPLEMENTING THE NATIONAL RESEARCH PROGRAM AS PLANNED, at i (2003). The NRP was developed to provide compliance data for updating the DIF to improve targeting noncompliant audits while minimizing the burden on taxpayers selected for the data collection audits. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-03-614, *supra*, at 1–2. It was also intended to identify potential methods for improving voluntary compliance. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-07-423, *supra*, at 1–2. About 46,000 tax returns were audited and of those, 8,000 were audited using information already in the possession of the IRS without contacting the taxpayer and another 9,000 were completed through letter correspondence. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-03-614, *supra*, at 19. The approximately 17,000 audits with minimal taxpayer contact were possible through a process called case-building, gathering IRS and third-party information to verify tax return data. *Id.* at 1, 19. The taxpayer was contacted for support for only those items that could be verified. *Id.* at 19.

96. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-07-423, *supra* note 95, at 1; *see also* U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-03-614, *supra* note 95, at 1.

97. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-07-423, *supra* note 95, at 1–2.

98. *Id.*

99. Historically, the IRS used data from taxpayers and third parties. *Id.* Today the IRS is using big data purchased from data brokers and the internet, as well as other questionable sources. Jeff Butler, Dir., Research Databases, IRS, Big Data and Analytics at the IRS, Address at the Techamerica Big Data Commission: Demystifying Big Data 5 (Nov. 14, 2012), https://www-01.ibm.com/events/wwe/grp/grp004.nsf/vLookupPDFs/Jeff%20Butler's%20Presentation/$file/Jeff%20Butler's%20Presentation.pdf [https://perma.cc/3NLN-NHHK].

Becoming more efficient has become increasingly important as the budget for the IRS continues to decrease. Since 2010, the budget has been cut by 17 percent and staff has decreased by 14 percent.[100] This has caused a decrease in compliance monitoring.[101] In fact, the number of audits dropped to an eleven-year low in 2015,[102] in which the IRS collected $3.3 trillion in revenue and processed about 243 million tax returns.[103] This corresponded to 35 cents spent for each hundred dollars it collected.[104] Although the IRS is one of the world's most efficient tax administrations,[105] IRS Commissioner John Koskinen stated that there comes a point when it is not possible to keep doing more with less without jeopardizing the mission of the IRS.[106] He also projected that with a larger budget, the IRS could increase the amount of taxes collected.[107] In his written statement, Koskinen added, "I don't know any organization in my 20 years of experience in the private sector that has said, 'I think I'll take my revenue operation and starve it for funds to see how it does.'"[108] It is estimated that for every dollar decrease in the IRS budget, there are five dollars owed that are not collected.[109]

Most of the changes the IRS has made to address their budget shortfall rely on the increased use of technology.[110] For the 2015 tax return filing season, around 90 percent of the returns were filed electronically, thus reducing the need for data entry employees.[111] In 2005 electronic filings comprised only 50 percent of the total.[112] The IRS uses the Automated Under-Reporter Program to match

---

100. Marr & Murray, *supra* note 2.

101. Daniel Bendtsen, *As IRS Budget Shrinks, So Does the Number of Audits*, DESERET NEWS (Dec. 15, 2015), http://national.deseretnews.com/article/6928/as-irs-budget-shrinks-so-does-the-number-of-audits.html [https://perma.cc/6D57-JEYF].

102. *Id.*

103. INTERNAL REVENUE SERV., INTERNAL REVENUE SERVICE DATA BOOK 2015, at iii (2015). The 243 million includes all tax returns such as income, employment taxes, excise, etc.

104. *The Agency, Its Mission and Statutory Authority*, *supra* note 11.

105. *Id.*

106. William Hoffman, *Koskinen Pledges Transparency and Accountability in Confirmation Hearing*, TAX ANALYSTS (Dec. 13, 2013), http://www.taxanalysts.org/content/koskinen-pledges-transparency-and-accountability-confirmation-hearing [https://perma.cc/E4SL-WLND].

107. *Id.*

108. *Id.*

109. *Id.*

110. Satran, *supra* note 43.

111. *U.S. Taxpayers E-filed More than 128 Million Returns in 2016*, EFILE.COM, http://www.efile.com/efile-tax-return-direct-deposit-statistics/ [https://perma.cc/VB3M-CJXJ] (last visited Mar. 6, 2017).

112. *Id.*

third-party information reports with tax returns and contact taxpayers via letters to resolve discrepancies.[113] In 2015, the IRS received 2.6 billion third-party information reports, of which over 87 percent were filed electronically. The ability to easily verify tax return information allowed the IRS to resolve more than 3.7 million tax return discrepancies in 2015, resulting in more than $6.3 billion.[114] All of this was accomplished with the equivalent of only 1,739 full-time employees—approximately a $3.6 million increase in post-audit collections per employee.[115]

Data analytics is being touted as the solution to the IRS's budget problems. Part of the IRS's data analytics program examines source data to identify noncompliant tax returns going beyond information provided by the taxpayer and the third party sources required to submit information (such as employers providing W-2s).[116] The IRS asserts that with analytics it can improve efficiencies and effectiveness of its investigations and avoid wasting taxpayers' time or creating unnecessary burdens on them.[117] Koskinen opined that without analytics, the future of the IRS would not be possible.[118] According to Dean Silverman, at that time the IRS's senior adviser to the commissioner for the Office of Compliance Analytics, the IRS is expanding its source data resources to include credit and debit card processors, PayPal, social media, and other Internet data.[119]

---

113. Gerard H. Schreiber Jr., *IRS Automated Underreporter Initiative*, TAX ADVISOR (Jan. 1, 2009), http://www.thetaxadviser.com/issues/2009/jan/irsautomatedunderreporter.html [https://perma.cc/2JQP-VAHY].

114. INTERNAL REVENUE SERV., *supra* note 103, at 37.

115. *Id.* at 38.

116. *Id.* at 37.

117. *Id.*

118. *Excerpt from Commissioner John Koskinen's Senate Finance Committee Testimony: Planning for the Future of the Taxpayer Experience*, IRS (Feb. 10, 2016), https://www.irs.gov/PUP/newsroom/FSTaxpayerInteraction.pdf [https://perma.cc/5DC8-PD5H]. For example, the IRS developed the data analytics program named Automated Substitute for Returns as a way to use third party information reports to identify non-filers, construct tax returns for them, and assess taxes, interest, and penalties. The IRS finalized more than 600,000 cases resulting in $2.7 billion in additional assessments. With ninety-three full-time equivalent employees for this program, the additional amount collected per employee after an audit is just over $29 million. INTERNAL REVENUE SERV., *supra* note 103, at 37–38.

119. Tam Habert, *IRS Implements Analytics for Compliance, Fraud Detection and Workforce Management*, DATA-INFORMED (Sept. 19, 2012), http://data-informed.com/irs-implements-analytics-for-compliance-fraud-detection-and-workforce-management/ [https://perma.cc/ACX8-NBVV].

III. POTENTIAL LEGAL ISSUES

This section explores the legal issues arising due to the IRS's data collection activities and analytics program. These include the failure to comply with fair information practices, the lack of transparency in the algorithm structure resulting in violations of the Administrative Procedure Act and potential discrimination, due process issues involving the collection of data without a warrant by the government, and other potential violations of federal statutes.

A. Fair Information Practices

According to Fred Cate, a privacy expert at Indiana University, the standard for data collection over the Internet is "notice and consent";[120] individuals should be informed that data is being collected about them and given the opportunity to correct such data.[121] In the beginning of the computer age, the US Department of Health, Education, and Welfare issued a report concerning the government's collection of data on individuals, which set standards known as the Fair Information Practices (FIPs).[122] These FIPs were revised by the Organization of Economic Cooperation and Development (OECD) and have been the basis of many federal, state and international privacy regulations.[123] The FIPs have been adopted by the Federal Trade Commission (FTC) as the "five core principles of privacy protection" and specifically name the notice-and-consent requirements as the basis of legal information privacy protection.[124] The main tenets of the FIPs are that (1) there should be no secret data collection systems; (2) there should be a way for data subjects to find out what information is in their records and how it is used; (3) data collected for one purpose should not be used for another without user permission; (4) the data subject should have the ability to correct inaccuracies; and (5) the data collector should keep reliable records and protect them.[125] The FTC continues to support this control by citizens over how their personal

120. VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA 153 (2014); *see also* Fred H. Cate & Viktor Mayer-Schönberger, *Notice and Consent in a World of Big Data*, 3 INT'L DATA PRIVACY L. 67, 67–73 (2013).

121. MAYER-SCHÖNBERGER & CUKIER, *supra* note 120, at 153; *see also* Cate & Mayer-Schönberger, *supra* note 120, at 67–73.

122. Paul M. Schwartz, *The EU-US Privacy Collision: A Turn to Institutions and Procedures*, 126 HARV. L. REV. 1966, 1969 (2013).

123. Susan Landau, *Control Use of Data to Protect Privacy*, 347 SCIENCE 504, 504 (2015).

124. *See generally* FED. TRADE COMM'N, PRIVACY ONLINE: A REPORT TO CONGRESS (1998).

125. Landau, *supra* note 123, at 504.

information is used, and the FIPs are specifically incorporated into the Privacy Act of 1974, discussed in Section III.D.1.[126]

1. No Notice

"It's well-known in the tax community, but not many people outside of it are aware of this big expansion of data and computer use [by the IRS]," says Edward Zelinsky, a tax law expert and professor at Benjamin N. Cardozo School of Law and Yale Law School.[127] "I am sure people will be concerned about the use of personal information on databases in government, and those concerns are well-taken. It's appropriate to watch it carefully. There should be safeguards."[128] Zelinsky went on to say that taxpayers should be made aware that what they say and do online could be used against them in IRS enforcement actions.[129] There have been instances of the IRS pointing to Facebook posts in defending their audit position that seem to support this statement.[130] Although the IRS website in no way reveals this to taxpayers, Dean Silverman, former Senior Advisor to the Commissioner in the Office of Compliance Analytics for the Internal Revenue Service, indicated that the IRS uses big data for the following[131]:

- Charting and analyzing social media such as Facebook
- Targeting audits by matching tax filings to social media or electronic payments
- Tracking individual Internet addresses and emailing patterns
- Sorting data in 32,000 categories of metadata and 1 million unique "attributes"

---

126. FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE, at i (2012).

127. Satran, *supra* note 43.

128. *Id.*

129. *Id.*

130. *IRS Has More Powers than Ever Before, Monitors Facebook, EBay*, NEWSMAX (May 13, 2013, 1:10 PM), http://www.newsmax.com/Newsfront/irs-expanded-powers-facebook/2013/05/13/id/504195/ [https://perma.cc/KLJ3-YFJF].

131. Satran, *supra* note 43.

> The IRS has brought in private industry experts to employ similar digital tracking—but with the added advantage of access to Social Security numbers, health records, credit card transactions and many other privileged forms of information that marketers don't see. 'Private industry would be envious if they knew what our models are,' boasted Dean Silverman, the agency's high-tech top gun who heads a group recruited from the private sector to update the IRS, in a comment reported in trade publications.

Id.

- Machine learning across "neural" networks
- Statistical and agent-based modeling
- Relationship analysis based on Social Security numbers and other personal identifiers.[132]

Nowhere in Facebook's terms of use,[133] or most likely on any social media site, is a provision indicating that users consent to the use of their information by the IRS. By making their posts private, Facebook users should be able to keep the IRS from accessing their information without a warrant; however, as previously noted, IRS agents were obtaining emails without a warrant as recently as 2013.[134] Although they agreed to stop this activity, they were silent with respect to accessing social media accounts,[135] and it seems pretty clear that the IRS has not provided adequate notice to tax payers of their data collection activities.

2. No Secret Data Collection Systems

There is little information available from the Treasury Department about the IRS's use of predictive analytics to conduct targeted audits.[136] A search of irs.gov comes up with only one hit for the name of the sub-agency responsible for these searches, the Office of Compliance Analytics, and that is on the back page of the 2014 Data Book.[137] When the IRS uses electronic information about taxpayers without their consent, the public does not have a way to check the information collected nor correct any mistakes in the information[138] that the IRS is using to determine whether they will be audited.[139] This lack of transparency also violates FIP requirements that there be no secret data collection activity.[140]

---

132. *Id.*

133. *See Data Policy*, FACEBOOK, https://www.facebook.com/about/privacy [https://perma.cc/SXU8-LJKJ] (last visited Mar. 6, 2016). Facebook does indicate that it will comply with a court order, subpoena, or search warrant and provide non-PII aggregated data to its analytics partners. *Id.*

134. Wessler, *supra* note 27.

135. *IRT-WBT Content 2009*, *supra* note 33, at 2–5.

136. *See generally* INTERNAL REVENUE SERV., BIG DATA ANALYTICS (2014), https://www.irs.gov/pub/irs-utl/BDA_pia.pdf [https://perma.cc/2KS7-NAFT].

137. INTERNAL REVENUE SERV., INTERNAL REVENUE SERVICE DATA BOOK 2014, at 74 (2014), https://www.irs.gov/pub/irs-pdf/p55b.pdf?_ga=1.199391210.597763954.1471625859 [https://perma.cc/JUP3-ZVVK].

138. See *infra* Section III.B.2 for how these mistakes can occur.

139. *See* Hatfield, *supra* note 19, at 349.

140. See *infra* Section III.D.1 regarding how this violates the Privacy Act of 1974.

In addition to the secrecy surrounding IRS data mining, they are also keeping the algorithms themselves secret. The reason is to prevent taxpayers from gaming the system by understanding the nature of the audit selection and working around it.[141] There are, however, enormous legal issues with respect to the failure of the IRS to disclose the algorithm structure.[142] Transparency is required by law with respect to predictive analytics because of the potential for violations of the Administrative Procedure Act (APA) and discriminatory decisions.[143] The Taxpayer Reform Act of 1998 also mandates IRS transparency.[144] The IRS bases its secrecy on the following language in the Taxpayer Reform Act:

> Such statement shall not include any information the disclosure of which would be detrimental to law enforcement, but shall specify the general procedures used by the Internal Revenue Service, including whether taxpayers are selected for examination on the basis of information available in the media or on the basis of information provided to the Internal Revenue Service by informants.[145]

However, the inability of individuals, entities, and even other branches of government to review the algorithms used by the IRS may be resulting in violations of law that are undiscoverable.

---

141. "Courts have concluded that the release of a taxpayer's DIF scores could reasonably be expected to risk circumvention of the law, as provided in 5 U.S.C. § 552(b)(7)(E), in that the release of such scores could enable taxpayers to determine how to lower DIF scores in order to avoid audits." Huene v. U.S. Dep't of the Treasury, No. 2:11–cv–02109 JAM KJN PS, 2012 WL 3730635, at *7 (E.D. Cal. Aug. 24, 2012), *quoted in* Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1512 n.54 (2013), https://www.illinoislawreview.org/wp-content/ilr-content/articles/2013/4/Zarsky.pdf [https://perma.cc/T6EE-VSLD].

142. *See* Richard Satran, *What Does the IRS Know About You?*, U.S. NEWS & WORLD REP. (Apr. 12, 2013, 9:00 AM), http://money.usnews.com/money/personal-finance/mutual-funds/articles/2013/04/12/what-does-the-irs-know-about-you [https://perma.cc/XX45-V9HE]. Accounting firms are also in the dark about these new practices, even though the ones that are aware admit they do not know how these algorithms work. *See id.*

143. There are numerous statutes addressing the need for transparency in government action, including but not limited to the Privacy Act of 1974, the Freedom of Information Act (FOIA), the Federal Agency Data Mining Reporting Act, and the E-Government Act, which also addresses the issues surrounding automated prediction processes. *See* Privacy Act of 1974, 5 U.S.C. §§ 552, 552a (2012); Federal Agency Data Mining Reporting Act of 2007, 42 U.S.C. § 2000ee-3 (2012); E-Government Act of 2002, 44 U.S.C. § 3501 (2012); Zarsky, *supra* note 141, at 1507 n.22.

144. Section 353 of the IRS Restructuring and Reform Act, Disclosure of Criteria for Examination Selection, requires the publication of general criteria for an audit trigger and informing the audited individual of the factors that triggered the audit. Pub. L. 105-206, § 3503, 112 Stat. 685, 771 (1998).

145. *Id.*; H.R. REP. NO. 105-599, at 295 (1998) (Conf. Rep.); S. REP. NO. 105-174, at 96 (1998); H.R. REP. NO. 105-364, at 74 (1997).

3. No Consent for Third Party Contact

According to Section 7602 of the Internal Revenue Code, the IRS is authorized to examine "any book, papers, records, or other data which may be relevant or material" to determining a taxpayer's tax liability.[146] However, the IRS may not contact a third party for the determination of a tax liability without providing reasonable notice to the taxpayer in advance.[147] The reason for this rule is that the IRS's inquiry regarding a taxpayer could have negative repercussions on that taxpayer's reputation.[148] The notice requirement allows the taxpayer to obtain the information for the IRS or otherwise resolve the issue in advance, making an IRS inquiry unnecessary.[149] Seeking information from holders of private electronic communications of a taxpayer without first providing notice to the tax payer would seem to violate this provision.[150]

4. Loss of Control over Use of Personal Information

While it has long been established that people have the right to determine when others may collect information about them and how such information may be used, the standards vary greatly from country to country.[151] The right to privacy was first documented in the United States in Brandeis and Warren's *Harvard Law Review* article "The Right to Privacy."[152] For hundreds of years, the United States firmly believed that this right not only created a tort action with respect to disclosures about private individuals but was also implied in the Constitution to prevent invasive government action.[153]

---

146. I.R.C. § 7602(a)(1) (2016). There are statutory exceptions to the requirement of notifying the taxpayer when a third party is contacted. *Id.* § 7602(c)(1). These include when providing notice would jeopardize the tax collection, the person being contacted fears reprisal from the taxpayer, and when the contact is made with respect to any criminal investigation. *Id.* § 7602(c)(3).

147. *Id.* § 7602(c)(1).

148. *See generally* Karen Schiller, *THIRD PARTY CONTACTS: IRS Third Party Contact Procedures Do Not Follow the Law and May Unnecessarily Damage Taxpayers' Businesses and Reputation*, 1 2015 ANNUAL REPORT TO CONGRESS 123 (2015), http://taxpayeradvocate.irs.gov/Media/Default/Documents/2015ARC/ARC15_Volume1_MSP_12_Third-Party-Contacts.pdf [https://perma.cc/9SEA-HRCP].

149. S. REP. NO. 105-174, at 77.

150. *See supra* Section III.A.1.

151. ALAN F. WESTIN, PRIVACY AND FREEDOM 7 (1967).

152. James H. Barron, *Warren and Brandeis, The Right to Privacy*, *4 HARV. L. REV. 193 (1890): Demystifying a Landmark Citation*, 13 SUFFOLK U. L. REV. 875, 875–76 (1979); Ben Bratman, *Brandeis & Warren's 'The Right to Privacy and the Birth of the Right to Privacy'*, 69 TENN. L. REV. 623, 624 (2002).

153. Bratman, *supra* note 152, at 624–26.

The right to privacy regarding health care and financial records has been long established; the government must be able to justify their need for such information.[154] Financial information is considered personal and disclosure tends to cause concern and anxiety in a reasonable person.[155] Individuals have the right to determine who can access such information.[156] Courts must weigh the government's interest in obtaining the information against an individual's right of privacy.[157]

When individuals provide information to a website, even if consent is given for the initial use of such data, a problem arises when that same data is being subjected to a secondary use.[158] This is because consent is not being given for these secondary uses, as such use is not envisioned at the time the consent is given.[159] There are cases where the US Tax Court has used information obtained by the IRS investigators from Facebook and eBay. In *Orellana v. Commissioner*, the taxpayer did not report the income she received from sales made on eBay.[160] The IRS subpoenaed various eBay and PayPal records to recreate the amount of unreported income.[161] In a different, much publicized 2014 case,[162] Rashia Wilson obtained tax refunds based on false information and was discovered because of her Facebook posts.[163] According to the CPA Practice Advisor, the IRS conducted searches of Wilson's public Facebook accounts to obtain the

---

154. James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1153, 1189–95 (2004).

155. *Id.* at 1193.

156. Ferdinand Schoeman, *Privacy and Intimate Information*, *in* PHILOSOPHICAL DIMENSIONS OF PRIVACY: AN ANTHOLOGY 403, 406 (Ferdinand David Schoeman ed., 1984) ("I think that what makes things private is in large part their importance to our conceptions of ourselves and to our relationships with others. . . . Selective self-disclosure provides the means through which people envalue personal experiences which are intrinsically or objectively valueless."), *cited in* Richard C. Turkington, *Legacy of the Warren and Brandeis Article: The Emerging Unencumbered Constitutional Right to Information Privacy*, 10 N. ILL. U. L. REV. 479, 507 n.92 (1990).

157. *See, e.g.*, Belle Bonfils Mem'l Blood Ctr. v. Dist. Court, 763 P.2d 1003, 1014 (Colo. 1988), *cited in* Turkington, *supra* note 156, at 513 n.108.

158. Joseph Jerome, *Big Data: Catalyst for a Privacy Conversation*, 48 IND. L. REV. 213, 236–39 (2014).

159. *See* Zarsky, *supra* note 141, at 1543; Edith Ramirez, Chairwoman, Fed. Trade Comm'n, Keynote Address at the Technology Policy Institute Aspen Forum: The Privacy Challenges of Big Data: A View From The Lifeguard's Chair 5 (Aug. 19, 2013).

160. Orellana v. Comm'r, No. 8950-08S, 2010 WL 1568447, at *2 (T.C. Apr. 20, 2010).

161. *Id.* at *6.

162. U.S. v. Wilson, 593 F. App'x 942 (11th Cir. 2014).

163. Elaine Silvestrini, *IRS Says Woman Bragged About Tax Fraud on Facebook*, TAMPA TRIB. (Mar 10, 2013), http://www.cpapracticeadvisor.com/news/10876997/irs-says-woman-bragged-about-tax-fraud-on-facebook [https://perma.cc/LL2A-YXKP].

damning information.[164] While the IRS may subpoena records in connection with an audit, if the IRS is using data mining on Facebook or other Internet sites to locate *potential* noncompliant activity, this would violate, at a minimum, the consent requirement of the FIPs.[165]

In addition, even when such information is given anonymously, the IRS most likely would be able to tie it back to an individual.[166] While website and data brokers may claim the information collected has been cleaned and anonymized, this does not protect an Internet user's privacy.[167] Today, search terms entered into a search engine for a research paper are collected as part of big data.[168] Even data that does not seem private can be used negatively and can be traced back to the individual.[169] Re-identification of allegedly anonymous data is easily accomplished. When data is first anonymized, personal information such as names, date of birth, etc. are removed from the data set.[170] While this works with small data sets, large data sets can easily result in re-identification.[171] A pair of scholars at the University of Texas were able to identify Netflix users based on de-

---

164. *Id.*

165. Richard Satran, *The IRS Has More Data About You than Ever Before*, U.S. NEWS & WORLD REP. (May 13, 2013, 11:48 AM), http://www.businessinsider.com/the-irs-ramps-up-online-tracking-2013-5 [https://perma.cc/A6JZ-GBYX].

> What the IRS does with all the information it can now access is not clear even to the agency's oversight boards and congressional overseers. The IRS's Information Reporting Program Advisory Committee, made up of tax professionals and advisers, in its annual report, raised "many questions" and numerous concerns over how the agency will use and manage data and said there was "a strong need for guidelines." The agency's mission statement says it will give "America's taxpayers top-quality service by helping them understand and meet their tax responsibilities and enforce the law with integrity and fairness to all." But the agency would make no comment for a story by U.S. News & World Report in early April that documented the growing array of new technology the agency has in its arsenal, including a $350 million investment in data mining tools. The agency declined numerous requests to detail any portion of its online policies. It did make a statement later to refute a charge not made in the story, that the IRS targets taxpayers for audit based on their online information.

Satran, *supra* note 142.

166. Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset)*, 2008 IEEE SYMP. ON SECURITY & PRIVACY 111, 111 (2008), http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4531148 [https://perma.cc/ND5X-ASEL].

167. Nate Anderson, *'Anonymized' Data Really Isn't—and Here's Why Not*, ARS TECHNICA (Sept. 8, 2009, 6:25 AM), http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/ [https://perma.cc/SV69-2YMT].

168. MAYER-SCHÖNBERGER & CUKIER, *supra* note 120, at 98–122.

169. *Id.* at 152.

170. *Id.* at 154.

171. *Id.*

identified data provided by Netflix, who had a contest to determine if a better movie recommendation system could be developed.[172] University of Colorado Law Professor, Paul Ohm, an expert on the harm done by de-anonymization, indicates that perfect anonymization is not possible.[173]

"Anonymized" data was long thought of as safe in terms of individual privacy, but has proved possible of re-identification.[174] Re-identification of anonymous data is possible in some instances with as little information as a name and a birthdate.[175] By aggregating the data of individuals from sites such as Netflix, Twitter, and Facebook, re-identification can be accomplished through a process of elimination.[176] This process has been proved multiple times, but notably by a graduate student, who combined hospital records with voter data to re-identify the Governor of Massachusetts's[177] hospital information.[178] Google, for example, collects and sells data sets including "your name, email address, telephone number, credit card (if you enter it), details on how you use Google's services, how you interact with other websites that use AdWords and other Google technologies, your device, [and] search queries. . . ."[179]

The IRS is training auditors to search Internet addresses, Facebook postings and other social media to back audit enforcements.[180] While the one posting on social media sites or providing information to websites is not contemplating that the IRS may view the material, it may very well be doing so. Because it has access to highly personal information about individuals, including Social Security numbers, income, and expenditure information, the IRS likely can recreate profiles from anonymized data.[181] However, a larger issue is that use of predictive analytics based on data gained from the Internet may be faulty because individuals often do not post reliable information on Facebook and other online platforms.[182]

---

172. Narayanan & Shmatikov, *supra* note 166, at 1–2, 10–12.

173. MAYER-SCHÖNBERGER & CUKIER, *supra* note 120, at 1.

174. *Id.* at 154–56.

175. Anderson, *supra* note 167.

176. *Id.*

177. William Weld was the governor at the time. *Id.*

178. *Id.*

179. Hachman, *supra* note 20.

180. Vince Polley & KnowConnect PLLC, *IRS Tracks Your Digital Footprint*, A.B.A. (Apr. 10, 2013), http://apps.americanbar.org/buslaw/blt/content/2013/05/mirln_1606.shtml #sthash.W4fetqiK.dpuf [https://perma.cc/B7TA-WSTY].

181. Satran, *supra* note 43.

182. Minas Michikyan, Jessica Dennis & Kaveri Subrahmanyam, *Can You Guess Who I Am? Real, Ideal, and False Self-Presentation on Facebook Among Emerging Adults*, 3 EMERGING

B. Lack of Transparency in Algorithm

The IRS specifies that tax returns are selected for audit through a variety of methods: random selection, computer screening or scoring, document matching, and statistical formula.[183] Algorithms are self-contained formulas for solving recurring problems, a series of steps that can be applied to data sets.[184] According to the White House Report (2014):

> In simple terms, an algorithm is defined by a sequence of steps and instructions that can be applied to data. Algorithms generate categories for filtering information, operate on data, look for patterns and relationships, or generally assist in the analysis of information. The steps taken by an algorithm are informed by the author's knowledge, motives, biases, and desired outcomes. The output of an algorithm may not reveal any of those elements, nor may it reveal the probability of a mistaken outcome, arbitrary choice, or the degree of uncertainty in the judgment it produces. So-called "learning algorithms" which underpin everything from recommendation engines to content filters evolve with the datasets that run through them, assigning different weights to each variable. The final computer-generated product or decision—used for everything from predicting behavior to denying opportunity—can mask prejudices while maintaining a patina of scientific objectivity.[185]

While the IRS provides general information regarding the selection of returns for audits, it does not reveal how the DIF algorithm, big data, or predictive analytics algorithms are utilized to select returns for audits.[186] There are numerous statutes addressing the need for transparency in government action, including, but not limited to, the Privacy Act of 1974, the Freedom of Information Act (FOIA), the Federal Agency Data Mining Reporting Act, and the E-Government Act, which also address issues surrounding automated prediction processes.[187] If the government is making decisions based completely on a computer model, the mechanism must be reviewed for procedural due process and potential discriminatory results.[188] Unfortunately,

---

ADULTHOOD 55, 60 (2015), http://journals.sagepub.com/doi/pdf/10.1177/2167696814532442 [https://perma.cc/E2PR-JR5C].

183. INTERNAL REVENUE SERV., PUBLICATION 556, EXAMINATION OF RETURNS, APPEAL RIGHTS, AND CLAIMS FOR REFUND 2 (2013), https://www.irs.gov/publications/p556/ [https://perma.cc/B8EG-P2TW]; *IRS Audits*, *supra* note 47.

184. EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 46 (2014).

185. *Id.*

186. *See* Hatfield, *supra* note 19, at 337–42; Satran, *supra* note 165.

187. *See* Privacy Act of 1974, 5 U.S.C. §§ 552, 552a (2012); Federal Agency Data Mining Reporting Act of 2007, 42 U.S.C. § 2000ee-3 (2012); E-Government Act of 2002, 44 U.S.C. §§ 3501, 3601 (2012); Zarsky, *supra* note 141, at 1507 n.22.

188. *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1271 (2008).

the courts have consistently denied FOIA requests by taxpayers to obtain access to these automated systems by citing I.R.C. § 6103(b)(2)[189] and 5 U.S.C. § 552(b)(3), which exempt disclosure when such disclosure might undermine law enforcement.[190]

1. Violations of Administrative Procedure Act

Federal agencies such as the IRS are subject to the APA.[191] The APA sets forth requirements for procedural due process in rulemaking.[192] Section 551 of the APA defines a "rule" as an "agency statement of general or particular applicability and future effect designed to implement, interpret, or prescribe law or policy . . . ."[193] It can be argued that those who create algorithms that make decisions impacting people's rights are engaging in rulemaking.[194] By failing to engage in a notice-and-comment period prior to the creation and adoption of such algorithms, the APA may be violated.[195] Danielle Keats Citron cites numerous examples of court cases where automated decisions systems failed to pass constitutional scrutiny.[196] Because the IRS relies on

189. 26 U.S.C. § 6103(b)(2) (2012).

190. Zarsky, *supra* note 141, at 1510–12, 1512 n.54.

191. 5 U.S.C. §§ 551–559 (2012).

192. Martin H. Redish & Lawrence C. Marshall, *Adjudicatory Independence and the Values of Procedural Due Process,* 95 YALE L.J. 455, 471 (1986).

193. 5 U.S.C. § 551(4).

194. *See* Citron, *supra* note 188, at 1288.

195. *See id.* at 1288–91.

196. *See id.* at 1264 n.97 (citing Petition to Determine Invalidity of Proposed Rule 65A-1.400 and ESS Online Benefits Application Form at 6, Tamara Clark v. Dep't of Children & Family Servs., No. 05-2105RP (Fla. Div. Adm. Hrgs. June 10, 2005) [hereinafter Clark Petition] as "arguing that relative caregivers could not apply for Temporary Assistance to Needy Families due to the design of the online application in violation of Florida law"); *see also id.* at 1290 n.275 (citing Clark Petition, *supra*, at 7–8 as "arguing that Florida's Department of Children and Families failed to follow applicable rulemaking procedures for change in rule embedded in design of Florida ACCESS online application that precluded relative caregivers from applying for TANF benefits in violation of state law"). "Florida's Department of Children and Family Services settled the litigation, agreeing to fix the system to allow relative caregivers to apply for benefits on behalf of children as required by federal law." *Id.* at 1264 n.97 (citing Telephone Interview with Valory Greenfield, staff attorney for Florida Legal Services, in Miami, Fla. (June 1, 2007)). "New York's automated public benefits system similarly failed to offer 'battered qualifying alien' as a choice in its drop-down menu for food stamp eligibility, thus precluding such individuals from applying for food stamps." *Id.* (citing M.K.B. v. Eggleston, 445 F. Supp. 2d 400, 418 (S.D.N.Y. 2006) as "granting preliminary injunction ordering New York City agencies to fix automated system to comply with established policy").

computers to make the decision on whether an individual will be audited or not, important procedural safeguards are being ignored.[197]

The secret nature of the algorithm[198] used by the IRS in targeting audits also would seem to violate open-government laws and regulations that are intended to provide the public access to basic information about the conduct of agencies.[199] The notice-and-comment rules are meant to allow the public to have input into changes in policy that could impact their rights.[200] In addition, without a record of the policy behind the algorithm, judicial review of such agency decision making is impaired.[201] Private corporations such as IBM, SAS, and EMC are behind providing big data sets as well as

---

197. *See id.* at 1281. According to Citron, automated decisions made by the government based on computerized algorithms often "deprive individuals of their liberty and property" in contravention "of the Due Process Clauses of the Fifth and Fourteenth Amendments." *Id.*

198. *See* Satran, *supra* note 165.

199. *See* 5 U.S.C. § 552(a)–(b) (2012).

200. Cass R. Sunstein, *"Practically Binding": General Policy Statements and Notice-and-Comment Rulemaking*, 68 ADMIN. L. REV. 491, 513 (2016).

201. *See* Citron, *supra* note 188, at 1298 n.327 ("[Henry H. Perritt, Jr., *The Electronic Agency and the Traditional Paradigms of Administrative Law*, 44 ADMIN. L. REV. 79, 89 (1992)]('Judicial review necessitates a 'record' of agency decisionmaking.') . . . [*s*]*ee also* Gordon G. Young, *Judicial Review of Informal Agency Action on the Fiftieth Anniversary of the APA: The Alleged Demise and Actual Status of Overton Park's Requirement of Judicial Review "On the Record*,*"* 10 ADMIN. L.J. AM. U. 179 (1996)."); *id.* at 1293–94 n.302.

> Computer programmers also arguably comprise advisory committees subject to the transparency requirements of the Federal Advisory Committee Act (FACA). FACA requires advisory committees—those "established or utilized" by the President or an agency for advisory purposes—to open their meetings, minutes, reports, and records to the public. 5 U.S.C. app. §§ 3(2), 10(a), 10(b) (2000). Courts exempt government contractors from FACA's mandates because procurement regulations impose transparency requirements on contractors in order to prevent the misuse of government resources. *Food Chem. News v. Young*, 900 F.2d 328, 331 (D.C. Cir. 1990) (citing H.R. REP. NO. 1403-92, at 2 (1972) (Conf. Rep.)). An argument can be made that the contractors here—computer programmers—should not fall within that exemption. Unlike the transparency provided by the contracting process that the FACA exemption addresses, here, the key issue is the opaque nature of the advice that software engineers provide in embedding new rules into an automated system's code. Such programmers do not solely execute policy. Instead, they effectively provide advice to the agency by changing established policy in the course of translating it into computer language and encoding it. That advice is, in turn, adopted by the agency through its automated decision system. Because FACA aims to secure transparency in the policy advice given to agencies, the spirit of the statute counsels its applicability to the consultants that design automated systems like the Federal Parent Locator Service. *See* A. Michael Froomkin, *Wrong Turn in Cyberspace: Using ICANN to Route Around the APA and the Constitution*, 50 DUKE L.J. 17, 139 (2000) (questioning whether private company running ICANN on behalf of Department of Commerce should be covered by FACA's mandates).

Id.

developing algorithms used by the IRS.[202] The creation of these algorithms by private companies creates the additional issue of no government oversight during the development of these algorithms.[203] This is problematic because courts are unable to determine whether the policy behind the algorithm is an abuse of discretion[204] or if an agency's decision is arbitrary and capricious.[205]

2. Lack of Accuracy of Big Data

One of the curious aspects of predictive algorithms is that they enable the creation of detailed individual profiles manufactured based on aggregated data which may not even be applicable to the individual selected for an audit.[206] These new inaccurate profiles result from the initial data set being combined with the profiles of those with similar characteristics.[207] The problem occurs when working backwards to

---

202. *Analytics and the IRS: A New Way to Find Cheaters*, FORBES (Jan. 28, 2016, 11:45 PM), https://www.forbes.com/sites/metabrown/2016/01/28/analytics-and-the-irs-a-new-way-to-find-cheaters/#2f16bb483187 [https://perma.cc/G2SB-QWQD] ("SAS founder Jim Goodnight points out that the IRS uses SAS analytics products for fraud detection, as do the Medicaid programs of every state."); Ian Armas Foster, *IRS to Utilize Big Data to Improve Returns*, DATANAMI (Apr. 15, 2013), http://www.datanami.com/2013/04/15/irs_to_utilize_big_data_to_improve_returns/ [https://perma.cc/S4PR-9ZLV]; David Huber, *SAS Arms IRS with New Fraud Detection Tools*, WASH. TECH. (Dec. 9, 2011), https://washingtontechnology.com/articles/2011/12/09/sas-assists-irs.aspx [https://perma.cc/H8HQ-KSNU].

> The Internal Revenue Service is getting a new weapon in its ceaseless battle to detect, prevent and resolve criminal and civil noncompliance with tax law. SAS, a business analytics software and services company, has won a $6.25 million contract to support IRS's new electronic Return Review Program system, which is designed to help reduce the $345 billion tax gap.

Id.; Jason Miller, *IRS' Approach to Big Data Focuses on Business Outcomes*, FED. NEWS RADIO (Sept. 17, 2014), http://federalnewsradio.com/management/2014/09/irs-approach-to-big-data-focuses-on-business-outcomes/ [https://perma.cc/84T5-2C8S].

203. However, it is questionable whether the current IRS rules and regulations are sufficient to manage the possibilities of big data analysis as they were written before access to electronic data such as this existed. Foster, *supra* note 202.

204. 5 U.S.C. § 706 (2012); KENNETH CULP DAVIS & RICHARD J. PIERCE, ADMINISTRATIVE LAW TREATISE § 11.5, 204 (3rd ed. 1994) ("An agency action that constitutes an unexplained departure from precedent must be reversed as arbitrary and capricious . . . ."), *cited in* Citron, *supra* note 188, at 1298 n.328.

205. Motor Vehicles Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 46–48 (1983) (holding the agency's rescission of the rule arbitrary and capricious because the agency did not provide factual or evidentiary support for rule), *cited in* Citron, *supra* note 188, at 1298 n.329; *cf.* Citizens to Preserve Overton Park, Inc. v. Volpe, 401 U.S. 402, 419–20 (1971) (holding the agency's decision to build a highway through a park arbitrary and capricious after a hearing on the grounds that the agency failed to explain its decision), *cited in* Citron, *supra* note 188, at 1298 n.329.

206. MAYER-SCHÖNBERGER & CUKIER, *supra* note 120, at 160.

207. *Id.* at 160–61.

individualize the information from the large set because these other characteristics may stick with the re-identification even though they were not present in that individual's initial profile.[208] Barocas and Nissenbaum (2014) have indicated in connection with such re-identification issues that "[t]he willingness of a few individuals to disclose information about themselves may implicate others who happen to share the more easily observable traits that correlate with the traits disclosed."[209]

Big data is able to create new profiles by using multiple data sets that effectively re-create an individual's information based on information obtained about others in the group that the individual is lumped in, or on faulty data associated with the individual in the first place.[210] Individuals misrepresent themselves on commercial websites for a variety of reasons, not only to embellish themselves[211] but also to meet the requirement for obtaining what is of interest to them from the website.[212] These misrepresentations enter the data pool and are not identified as such when using big data in predictive analysis.[213] If a pattern recognition algorithm is used on this tainted data to develop a profile of noncompliant taxpayer behaviors, misidentifications are likely to occur. It is easy to be seduced into treating these algorithmic patterns as predictions of real behavior because they appear to be objective. It is difficult to challenge inaccuracies of an algorithmic-produced data profile because the inaccuracies are not about the individual's actual behavior, but rather reported behavior that may have been intentionally misrepresented by the individual for reasons completely unrelated to the pattern being predicted by the algorithm itself. Thus, verification of the accuracy of the database is critical for developing useful algorithms predicting behavior.[214] This is why the

208. Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 117 (2014).

209. Solon Barocas & Helen Nissenbaum, *Big Data's End Run Around Procedural Privacy Protections*, 57 COMM. ACM 31, 32 (2014), *cited in* Landau, *supra* note 123.

210. Crawford & Schultz, *supra* note 208, at 125–26.

211. Susmita Baral, *Lying on Facebook Results in 'Digital Amnesia' Which Rewrites Your Memories*, IDIGITALTIMES (Dec. 31, 2014, 4:21 PM), http://www.idigitaltimes.com/lying-facebook-results-digital-amnesia-which-rewrites-your-memories-403976 [https://perma.cc/7ASS-VC2C].

212. Sandra Braman, *Tactical Memory: The Politics of Openness in the Construction of Memory*, FIRST MONDAY (July 3, 2006), http://firstmonday.org/ojs/index.php/fm/article/view/1363/1282 [https://perma.cc/B6QL-TWL7]; *see also Facebook & Your Privacy: Who Sees the Data You Share on the Biggest Social Network?*, CONSUMER REP. (June 2012), http://www.consumerreports.org/cro/magazine/2012/06/facebook-your-privacy/index.htm [https://perma.cc/4H4Z-MYS2].

213. MAYER-SCHÖNBERGER & CUKIER, *supra* note 120, at 32–49.

214. Crawford & Schultz, *supra* note 208, at 119.

IRS conducted the NRP audits to verify all information used to update the DIF.[215]

Both the Internet and data sets from data brokers contain information posted online by either the individual or a third party. Self-reported information online is notoriously suspect as it is often designed to enhance one's self-image by strategically selecting how and what to disclose.[216] Research on deception finds that in many online social websites, exaggeration regarding oneself, also known as airbrushing, has become the norm.[217] Anyone who has ever been on a dating website knows this very well. Facebook presents a slightly different issue because others may choose to post on your wall, which means you do not fully control the information that may be associated with your page. A wildly conservative friend may post a video on your wall with which you do not agree, but you feel no need to remove it. If the government is collecting this information about you, it may present an inaccurate picture of your politics.[218] Information posted about you on a website you do not visit can also contain falsehoods. The problem is those false data points could also increase your chances of being unfairly targeted for an audit. An audit itself can be viewed as a punishing experience due to the stress and potential defense costs involved even if a "no change" order occurs.[219]

By keeping the data collected proprietary, the IRS is effectively preventing people from viewing and correcting information about themselves that the IRS may be using in its predictive analytics.[220] This not only violates the FIPs but also is in direct contravention of the Privacy Act of 1974, which incorporates the FIPs.[221] Since individuals do not even know that there has been incorrect information collected about them, they have no ability to correct this

---

215. BRIAN ERARD, COMPLIANCE MEASUREMENT AND WORKLOAD SELECTION WITH OPERATIONAL AUDIT DATA 1–2, 4 (2002), https://www.irs.gov/pub/irs-soi/compmewo.pdf [https://perma.cc/K4LB-AJML].

216. Andrew T. Fiore & Judith S. Donath, *Online Personals: An Overview*, *in* ACM, EXTENDED ABSTRACTS ON HUMAN FACTORS IN COMPUTING SYSTEMS 1395 (2004).

217. *Id.* at 1398.

218. Joseph P. Mazer, Richard E. Murphy & Cheri J. Simonds, *I'll See You on "Facebook": The Effects of Computer-Mediated Teacher Self-Disclosure on Student Motivation, Affective Learning, and Classroom Climate*, 56 COMM. EDUC. 1, 3, 11 (2007).

219. Kelly Spors, *Audit Nightmares: How to Cope When the IRS Comes Knocking*, OPENFORUM (2014), https://www.americanexpress.com/us/small-business/openforum/articles/audit-nightmares-how-to-cope-when-the-irs-comes-knocking/ [https://perma.cc/YXB4-XH8N]; *see also Taxpayer Stories of Painful IRS Audit and Tax Experiences*, EFILE.COM, http://www.efile.com/painful-taxpayer-irs-audit-experiences-tax-stories/ [https://perma.cc/ZY9R-LUCE] (last visited Mar. 6, 2017).

220. Landau, *supra* note 123, at 504; *see also* Zarsky, *supra* note 141, at 1510–13.

221. Zarsky, *supra* note 141, at 1541–42.

data that is being entered into the IRS database. Thus, they cannot avail themselves of the opportunity to disprove the data.[222] Previous data analytics relied on sampling and the need for accurate data.[223] Today, big data varies in quality, [224] and computers have made the ability to analyze large data sets possible in a way that could not occur prior to their existence.[225] Accuracy is sacrificed for volume.

3. Potential Discrimination

Another interesting aspect of predictive analytics is that algorithms can be programmed to "learn" over time.[226] While this may be advantageous for companies looking to narrowly define their target market, it creates a dangerous situation when the result is a targeted audit. Thus, the programmer's initial purpose for developing the pattern-recognition algorithm can change as the algorithm evolves free from human intervention.[227] It can create its own identification function based on the patterns it recognizes within the big data sets, and these functions may not be consistent with the original function.[228] Thus, the algorithm itself, while not initially set up to use factors such as race or religion, may result in targeting certain groups based on the associations created as the algorithm learns.[229] Because the IRS collects and maintains highly personal information about taxpayers, they can easily identify someone's race, gender, ethnicity, and religion.[230]

The New York Police Department came under fire for its use of predictive analytics to focus its policing on certain communities.[231] Many of the areas targeted were primarily composed of minorities.[232] Because the use of predictive analytics relies on correlation rather than causation, it is unable to explain why things occur, but rather

---

222. MAYER-SCHÖNBERGER & CUKIER, *supra* note 120, at 176.

223. *Id.* at 32.

224. *Id.* at 32, 176.

225. *Id.* at 8–11.

226. Melissa De Zwart, Sal Humphreys & Beatrix Van Dissel, *Surveillance, Big Data and Democracy: Lessons for Australia from the US and UK*, 37 U. NEW S. WALES L.J. 713, 718 (2014).

227. *Id.*

228. *Id.*

229. FED. TRADE COMM'N, BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION?: UNDERSTANDING THE ISSUES 28 (2016), https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf [https://perma.cc/K8YS-WAJ8].

230. Hatfield, *supra* note 19, at 321–22.

231. 1 Bennett Capers, *Rethinking the Fourth Amendment: Race, Citizenship, and the Equality Principle*, 46 HARV. C.R.-C.L. L. REV. 1, 17, 17 n.120 (2011).

232. *Id.* at 16, 17, 17 n.120.

what is predicted to occur based on the data set.[233] It is possible that two things may behave similarly based on coincidence.[234] With correlation there is only probability, not certainty.[235] Spurious correlations are more frequent using big data sets.[236] No longer do individuals have to come up with search terms to test; instead, the proxies reveal themselves when big data is analyzed. Society is moving from a hypothesis-driven approach to a data driven one.[237]

Because predictive analytics does not ask why, it does not reveal why people may have higher than normal expenses on their tax returns. For example, many bankruptcies result from exceedingly high medical bills.[238] This may be a reason for the unusual deductions on a tax return, but not a reason for an audit. If it is found that minorities have higher than normal medical expenses, this correlation could result in minorities being targeted for audits in violation of equal protections laws. The problem, of course, is that with a secretive IRS and no access to the algorithm itself, claims of discrimination would have little chance of success.

While current law prohibits discrimination against protected classes, big data is not the panacea that many proponents allege. Because of the potential correlation of characteristics that could trigger an audit with those of a certain protected class, the mathematical model could result in unfair treatment of such class.[239] In addition, data mining could "inherit the prejudices of prior decision-makers."[240] As discussed by other legal analysts, access to an automated program's source code would allow an individual to challenge an agency's actions and show that the system may in fact be biased.[241]

In a recent FTC Report regarding big data, the potential violations of law and risks of using predictive analytics on low-income

---

233. *Id.* at 40, 40 n.246.

234. MAYER-SCHÖNBERGER & CUKIER, *supra* note 120, at 52–53.

235. Capers, *supra* note 231, at 16.

236. MAYER-SCHÖNBERGER & CUKIER, *supra* note 120, at 160–62.

237. *Id.* at 55.

238. David U. Himmelstein et al., *Medical Bankruptcy in the United States, 2007: Results of a National Study*, 122 AM. J. MED. 741, 741, 743 (2009).

239. Solon Barocas & Andrew Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671, 688, 720, 726 (2016), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899 [https://perma.cc/CK5E-PLUH].

240. *Id.* at 674, 682.

241. Christopher W. Clifton, Deirdre K. Mulligan & Raghu Ramakrishnan, *Data Mining and Privacy: An Overview*, *in* PRIVACY AND TECHNOLOGIES OF IDENTITY 191, 203 (Katherine J. Strandburg & Daniela Stan Raicu eds., 2006) (explaining that, without access to the underlying data and logic of the "No Fly" program, individual's ability to challenge inclusion on list is impaired), *cited in* Citron, *supra* note 188, at 1284 n.240.

and underserved populations was explored.[242] This report was prepared with input from a public workshop with big data stakeholders.[243] The report reiterates that results from hidden biases can manifest at either the collection or analytics stage,[244] or from the algorithm's ability to learn.[245] This presents an enormous issue because there is no way to verify that the IRS's use of analytics is not resulting in discrimination.

4. Arbitrary and Capricious Agency Action

In general, agency determinations may not be overturned by a federal court unless the agency action is found to be arbitrary and capricious.[246] As previously mentioned, the use of predictive analytics may violate APA procedural due process in rulemaking requirements, but it may also serve to provide a way to force the disclosure of the analytics program being used by the IRS today.[247] According to Citron, government-automated decision-making systems which are kept secret have an impact on people's rights in a way that may go unchecked.[248] She suggests the addition of "technological due process" to ensure accuracy and fairness.[249] Citron points to the FTC's auditing of CompuServe's scoring process for giving credit, which was found to have used unfair criteria.[250] She suggests the FTC could be charged with running

242. FED. TRADE COMM'N, *supra* note 229, at i–ii.

243. *Id.* at i.

244. *Id.* at iv.

245. Zarsky, *supra* note 141, at 1508–12.

246. 5 U.S.C. § 706(2)(A) (2012); United States v. Bean, 537 U.S. 71, 77 (2002); High Sierra Hikers Ass'n v. Blackwell, 390 F.3d 630, 638 (9th Cir. 2016); Gardner v. U.S. Bureau of Land Mgmt., 638 F.3d 1217, 1224 (9th Cir. 2011); Latino Issues Forum v. Envtl. Prot. Agency, 558 F.3d 936, 941 (9th Cir. 2009); Pub. Util. Dist. No. 1 of Snohomish Cty. v. Fed. Emergency Mgmt. Agency, 371 F.3d 701, 706 (9th Cir. 2004).

247. *See* Margaret Hu, *Big Data Blacklisting*, 67 FLA. L. REV. 1775, 1799 (2015) ("Core liberties may be obstructed in a way that is rapidly evolving and systemic, however, nearly impossible to detect because of the opacity and complexity of big data technologies, and the administrative systems that support them.").

248. Danielle Citron, *Big Data Should Be Regulated by 'Technological Due Process'*, N.Y. TIMES (July 29, 2016, 4:34 PM), http://www.nytimes.com/roomfordebate/2014/08/06/is-big-data-spreading-inequality/big-data-should-be-regulated-by-technological-due-process [https://perma.cc/KR76-7ZGS].

249. *Id.*

250. *Id.*

> hypothetical scenarios to assess whether algorithmic predictions are statistical proxies for race, gender, religion and disability—thereby cutting down the possibility that the algorithms infringe on civil rights. The ever-present threat of an audit would encourage the adoption of precautions and, perhaps, encourage entities that are building scoring systems to be more mindful of concerns about discrimination and inaccurate predictions based on polluted data.[251]

While companies who use big data to perform targeted marketing can profit even with low accuracy rates,[252] the government is subject to a different set of rules.[253] A taxpayer's rights could be affected if the algorithm results in the auditing of a suspect due to payments that are correlated to a certain ideology or religion. For example, high charitable contributions may be associated with certain religions that observe tithing, or new home ownership with no associated mortgage may identify religions with prohibitions against usury charges. However, the IRS's analysis program could potentially flag these types of line items on a tax return, resulting in unfair targeting of those from a certain religion. If the FTC or a newly created oversight board was able to review both the data for accuracy and the algorithm for potential discrimination, this could help repair some of the many problems with the IRS's use of big data analytics.

C. Data Collection

There are also potential issues with the IRS's methods of data collection. The ability of the government to obtain information about people is limited by the Constitution as well as federal and state law. In 1967, the Supreme Court issued two decisions regarding the Fourth Amendment with respect to private communications.[254] In *Katz v. United States*, the Court held that the Fourth Amendment's prohibition against "unreasonable searches and seizures" entitled an individual to a reasonable expectation of privacy in his or her private communications, thus precluding unwarranted government intrusion.[255] In *Berger v. New York*, the Court struck down a New York wiretap law as violating the Fourth Amendment because of its failure to provide adequate safeguards for the privacy interests of

---

251. *Id.*

252. Hatfield, *supra* note 19, at 343.

253. *See* Citron, *supra* note 188, at 1281–82.

254. Katz v. United States, 389 U.S. 347, 359 (1967); Berger v. New York, 388 U.S. 41, 63–64 (1967).

255. *Katz*, 389 U.S. at 359.

those whose communications were being wire tapped.[256] The ECPA was enacted in 1986 to extend these same privacy protections.[257]

1. Electronic Communications Privacy Act

The ECPA was intended to protect private electronic communications in furtherance of the two aforementioned Supreme Court cases.[258] It described the limited circumstances under which the government may obtain information stored by electronic communications services and remote computing services from providers.[259] It required the federal government to obtain either a warrant, if the communication sought had been in storage for less than 180 days,[260] or a subpoena, with notice to the customer, if more than 180 days.[261] Section 2703(d) required the government to show "specific and articulable facts, showing that there are reasonable grounds to believe that the contents of a[n] . . . electronic communication . . . are relevant and material to an ongoing criminal investigation."[262] The ECPA has been interpreted to apply not only to emails but also to text messages[263] and social media.[264] In any case, the ECPA requires the government to provide notice to the individual whose communications are being requested.[265] The ACLU has obtained information indicating that not only has the IRS viewed private electronic communications without first obtaining a warrant,[266] it is likely not disclosing requests to taxpayers as required by law.[267]

256. *Berger*, 388 U.S. at 63–64.

257. Electronic Communications Privacy Act of 1986, Pub L. No. 99-508, 100 Stat. 1848 (1986).

258. Jim Harper, *Reforming Fourth Amendment Privacy Doctrine*, 57 AM. U. L. REV. 1381, 1383–88 (2007).

259. 18 U.S.C. § 2703 (2012).

260. § 2703(a).

261. § 2703(d).

262. *Id.* A Section 2703(d) order is similar to the *Terry* rule applied to law enforcement stop-and-frisks, which requires less than probable cause to believe a crime has been committed, but more than a mere hunch. *See* Terry v. Ohio, 392 U.S. 1, 21–22 (1968).

263. Quon v. Arch Wireless Operating Co., 529 F.3d 892, 901 (9th Cir. 2008).

264. Crispin v. Christian Audigier, Inc., 717 F. Supp. 2d 965, 980, 989 (C.D. Cal. 2010).

265. 18 U.S.C. §§ 2703(a), 2705.

266. Nathan Freed Wessler, *IRS Says It Will Respect 4th Amendment with Regard to Email, but Questions Remain*, ACLU (Apr. 16, 2013, 4:07 PM), https://www.aclu.org/blog/irs-says-it-will-respect-4th-amendment-regard-email-questions-remain?redirect=blog/technology-and-liberty-national-security/irs-says-it-will-respect-4th-amendment-regard-email [https://perma.cc/6VUH-LZ5W].

267. *Id.*

In *United States v. Warshak*, the government obtained 27,000 private electronic communications with a 2703(b) subpoena and 2703(d) order under the ECPA.[268] The *Warshak* court ruled that the obtaining of such private emails violated the Fourth Amendment's protection against unreasonable searches and seizures and that to the extent that it permitted such retrieval without a warrant the ECPA violated the Constitution.[269] The court stated that the Fourth Amendment applies to electronic communications and that "The Fourth Amendment must keep pace with the inexorable march of technological progress."[270]

Congress recently attempted to amend the ECPA to subject governmental retrieval of electronic communications to a warrant, but to date none of the amendments have passed.[271] As mentioned above, despite the *Warshak* opinion, the IRS was still obtaining taxpayer emails without a warrant until 2013, when they had to answer for the practice in a Senate hearing.[272] An ACLU statement issued after the hearing remarked:

> Although Miller stated that the IRS Criminal Investigation unit obtains warrants for all emails, he did not discuss other forms of electronic communication such as text messages, instant messages, and direct messages on social media . . . . Under the Fourth Amendment, a warrant should be required for those private communications as well.[273]

---

268. United States v. Warshak, 631 F.3d 266, 283 (6th Cir. 2010).

269. *Id.* at 274.

270. *Id.* at 285; *see also* Orin S. Kerr, *Applying the Fourth Amendment to the Internet: A General Approach*, 62 STAN. L. REV. 1005, 1007 (2010).

271. *See* RICHARD M. THOMPSON, II & JARED P. COLE, CONG. RESEARCH SERV., R44036, STORED COMMUNICATIONS ACT: REFORM OF THE ELECTRONIC COMMUNICATIONS PRIVACY ACT (ECPA) 8 (2015).

> The ECPA Amendments Act of 2015 (S. 356, H.R. 283) and the Email Privacy Act (H.R. 699) were re-introduced in the 114th Congress. Similar to the past Congress, the Email Privacy Act has obtained a majority of the Members of the House as co-sponsors (261). The Online Communication and Geolocation Protection Act, which would make similar amendments to ECPA, was introduced in the 113th (H.R. 983) and 114th (H.R. 656) Congresses. A competing bill, the Law Enforcement Access to Data Stored Abroad (LEADS) Act (S. 512, H.R. 1174), which covers, among other things, the extraterritorial reach of ECPA orders, was first introduced in the 113th Congress and has been re-introduced in the 114th Congress.

Id.

272. Sampson, *supra* note 31.

273. Wessler, *supra* note 266.

2. Warrantless Search

The Fourth Amendment provides:

> The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things . . . .[274]

As previously discussed, until 2013, the IRS maintained that it was permitted to view private email messages on servers that were over 180 days old without first obtaining a warrant.[275] The *Warshak* court was very clear that electronic communications—specifically, email—should be treated the same as all other private communications and are subject to Fourth Amendment protections.[276]

---

274. U.S. CONST. amend. IV.

275. Audrey Henderson, *The IRS May Be Spying on You Through Social Media*, OPTIMA TAX RELIEF (Apr. 29, 2014), http://optimataxrelief.com/irs-may-be-spying-social-media [https://perma.cc/2MAX-ZUE9].

276. United States v. Warshak, 631 F.3d 266, 285–86 (6th Cir. 2010).

> Given the fundamental similarities between email and traditional forms of communication, it would defy common sense to afford emails lesser Fourth Amendment protection. . . . Email is the technological scion of tangible mail, and it plays an indispensable part in the Information Age. Over the last decade, email has become "so pervasive that some persons may consider [it] to be [an] essential means or necessary instrument[ ] for self-expression, even self-identification." . . . It follows that email requires strong protection under the Fourth Amendment; otherwise, the Fourth Amendment would prove an ineffective guardian of private communication, an essential purpose it has long been recognized to serve. . . . As some forms of communication begin to diminish, the Fourth Amendment must recognize and protect nascent ones that arise. . . . If we accept that an email is analogous to a letter or a phone call, it is manifest that agents of the government cannot compel a commercial ISP to turn over the contents of an email without triggering the Fourth Amendment. An ISP is the intermediary that makes email communication possible. Emails must pass through an ISP's servers to reach their intended recipient. Thus, the ISP is the functional equivalent of a post office or a telephone company. As we have discussed above, the police may not storm the post office and intercept a letter, and they are likewise forbidden from using the phone system to make a clandestine recording of a telephone call—unless they get a warrant, that is. . . . It only stands to reason that, if government agents compel an ISP to surrender the contents of a subscriber's emails, those agents have thereby conducted a Fourth Amendment search, which necessitates compliance with the warrant requirement absent some exception.

Id. (quoting City of Ont. v. Quon, 560 U.S. 746, 760 (2010)) (relying on the following authorities for the following propositions: Patricia L. Bellia & Susan Freiwalk, *Fourth Amendment Protection for Stored E-Mail*, 2008 U. CHI. LEGAL F. 121, 135 (2008) as "recognizing the need to 'eliminate the strangely disparate treatment of mailed and telephonic communications on the one hand and electronic communications on the other'"; *Quon*, 560 U.S. at 762–63 as "implying that 'a search of [an individual's] personal e-mail account" would be just as intrusive as "a wiretap on his home phone line'"; United States v. Forrester, 512 F.3d 500, 511 (9th Cir. 2008) as "holding that '[t]he privacy interests in [mail and email] are identical'"; United States v. U.S. Dist. Court, 407 U.S. 297, 313 (1972); United States v. Waller, 581 F.2d 585, 587 (6th Cir. 1978)

Although the Supreme Court has not yet had the opportunity to rule on whether predictive policing is constitutional, parallels may be drawn to the IRS's use of big data to predict which taxpayers may be noncompliant. Andrew Ferguson concludes that "because predictive policing does not provide personal knowledge about an ongoing crime, or particularized identification of the suspect involved, it cannot support the weight of reasonable suspicion."[277] In general, reasonable suspicion requires corroboration of individual actions.[278] Because it is unknown how the IRS's algorithm is choosing returns to audit, it is unknown whether they are using predictive analytics to target individuals and businesses without proper constitutional protections. If the IRS is not predicting future behavior, but rather examining prior actions of a particular person or viewing tax returns that have already been filed, it is possible that the prediction is based on that individual's own behavior self-reported to the IRS. When a return is filed, the taxpayer agrees with the following statement: "Under penalties of perjury, I declare that I have examined this return and accompanying schedules and statements, and to the best of my knowledge and belief they are true, correct and complete."[279]

In *United States v. Jones*, the Supreme Court held that attaching a GPS device to a car and tracking its movements without a search warrant violated the owner's reasonable expectation of privacy and constituted a search under the Fourth Amendment.[280] An argument can be made that just as a vehicle's movements allow for an expectation of privacy, so too should an individual's online activities from a home computer.[281] A government agency running data analytics on the Internet and searching for tags that fit their profile of

---

as "noting the Fourth Amendment's role in protecting 'private communications'"; Warshak v. United States, 490 F.3d 455, 473 (6th Cir. 2007) as claiming that "[i]t goes without saying that like the telephone earlier in our history, e-mail is an ever-increasing mode of private communication, and protecting shared communications through this medium is as important to Fourth Amendment principles today as protecting telephone conversations has been in the past"; United States v. Jacobsen, 466 U.S. 109, 114 (1984); Katz v. United States, 389 U.S. 347, 353 (1967)).

277. Andrew Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 EMORY L.J. 259, 305–06 (2012).

278. *Id.* (citing Illinois v. Gates, 462 U.S. 213, 241 (1983)).

279. INTERNAL REVENUE SERV. & DEP'T OF TREASURY, OMB No. 1545-0074, FORM 1040: U.S. INDIVIDUAL INCOME TAX RETURN 2 (2016).

280. United States v. Jones, 565 U.S. 400, 404 (2012).

281. Some have argued that all online activity is subject to the third party doctrine because the ISP would have knowledge of such activities. However, the third party doctrine was established long before the Internet. While the third party doctrine may well apply to *public* postings, such as a photo on Facebook where the user has not set her settings to private, the same reasoning cannot hold for private online activities (such as a Google search).

a noncompliant taxpayer could arguably constitute a search under the Fourth Amendment.[282] In *Kyllo v. United States*, the Supreme Court held that using a thermal imaging device without a search warrant constitutes a search under the Fourth Amendment.[283] The use of advanced technology to essentially spy on a US citizen without a warrant reached beyond the reasonable expectation of privacy.[284] Again, the analogy can be made to a citizen's use of the Internet from a home computer.[285]

In *Riley v. California*, the Supreme Court held that cell phones may not be searched without a warrant.[286] According to Orin S. Kerr, a law professor at George Washington University, "[t]his is a bold opinion" because "it is the first computer-search case, and it says we are in a new digital age. You can't apply the old rules anymore."[287] Chief Justice John G. Roberts Jr. indicated that cell phones contain private and personal information and that

> [o]ne of the driving forces behind the American Revolution was revulsion against "general warrants," which allowed British officers to rummage through homes in an unrestrained search for evidence of criminal activity. The fact that technology now allows an individual to carry such information in his hand does not make the information any less worthy of the protection for which the founders fought.[288]

The IRS's use of cell phone tracking technology would seem to be a warrantless search as well.

3. Due Process

There are both procedural and substantive issues that arise from the government's use of big data analytics to categorize individuals into groups such as the No Fly List and the No Citizenship

---

282. Unreasonable searches by the government are prohibited under the Fourth Amendment. Essentially, individuals have an expectation of privacy with respect to their persons and homes. However, government searches pursuant to a warrant or when criminal activity is being committed in plain view would not be considered unreasonable. Muna Busailah & Stephen P. Chulak, *Fourth Amendment Search and Seizure, Qualified Immunity and the Technological Age*, 2012 (6) AELE MONTHLY L.J., 501, 501, 505.

283. Kyllo v. United States, 533 U.S. 27, 34–35, 37, 40 (2001).

284. David A. Sklansky, *Back to the Future: Kyllo, Katz, and Common Law*, 72 MISS. L.J. 143, 208 (2002).

285. It is important to distinguish for the purposes of this argument an individual's use of a home computer from a computer at a library or another public place where there is not the same expectation of privacy.

286. Riley v. California, 134 S. Ct. 2473, 2485 (2014).

287. Adam Liptak, *Major Ruling Shields Privacy of Cellphones: Supreme Court Says Phones Can't Be Searched Without a Warrant*, N.Y. TIMES (June 25, 2014), http://www.nytimes.com/2014/06/26/us/supreme-court-cellphones-search-privacy.html?_r=0 [https://perma.cc/2TFV-NSZV].

288. *Id.*

List.[289] There is a procedural problem because of the potential for wrongful inclusion due to errors that can have large consequences on the individual classified.[290] These mistakes can result in serious legal issues for the individual, such as the inability to travel or work while the issues are resolved.[291] Furthermore, large amounts of data that help the government identify "suspicious" people ignores the principle of "innocent until proven guilty," which is a substantive issue and an inalienable right of the people.[292] These issues are concerning until the government can show, using accurate algorithms and data, that big data is the least intrusive way that the government can go about identifying these people.[293]

The IRS's data collection activities potentially employ inadequate privacy safeguards, a likely violation of the aforementioned Supreme Court cases.[294] Jennifer Lynch, a senior staff attorney with the Electronic Frontier Foundation, a San Francisco-based privacy-rights group, told Bloomberg:

> Especially with the IRS, I don't know why these agencies are getting access to this kind of information. These systems treat every person in the area as if they're under investigation for a crime—that is not the way our criminal justice system was set up or the way things work in a democratic society.[295]

4. Self-Incrimination

When an individual's life, liberty, and property are affected due to a governmental decision, there is a fairness requirement.[296] The Fifth Amendment provides:

> No person shall be held to answer for a capital, or otherwise infamous crime, unless on a presentment or indictment of a grand jury, . . . nor shall be compelled in any criminal case to be a witness against himself, nor be deprived of life, liberty, or property, without due process of law . . . .[297]

In general, individuals cannot be compelled to testify against themselves in a criminal case. This includes the right not to produce

---

289. Hu, *supra* note 247, at 1775–77.

290. *Id.*

291. *Id.*

292. *Id.*

293. *Id.* at 1796.

294. Those cases are *Riley v. California*, 134 S. Ct. 2473 (2014), *United States v. Jones*, 565 U.S. 400 (2012), and *Kyllo v. United States*, 533 U.S. 27 (2001).

295. Kay Bell, *IRS Getting Sneakier in Tracking Tax Cheats*, DON'T MESS WITH TAXES (May 12, 2014), http://dontmesswithtaxes.typepad.com/dont_mess_with_taxes/2014/05/irs-sneaky-tax-cheat-tracking-via-facebook.html [https://perma.cc/L4NL-3RZM].

296. Citron, *supra* note 248.

297. U.S. CONST. amend. V.

private papers.[298] When the IRS views private emails and private social media communications and postings, this could be deemed to be a violation of the Fifth Amendment's right against self-incrimination.[299] Although people feel free to speak frankly with those they connect with on social media, the fact that these conversations are placed in fixed form presents problems that verbal communications do not. While wiretapping by the government must be done pursuant to subpoena, the IRS's ability to collect data from social media and online electronic communications that are meant to be private can result in taxpayers being forced to testify against themselves. If privacy settings are being ignored by the IRS and private communications are being accessed by the IRS without a warrant, this is akin to reading private emails without a warrant, which is prohibited under *Warshak*.[300]

D. Other Federal Violations

1. Privacy Act of 1974

The Privacy Act of 1974 limits the federal government's use of private data about US citizens and provides a mechanism for individuals to obtain information about themselves maintained by the government.[301] This Act was passed at the dawn of the computer age at a time when the public became aware of domestic spying by the US government on its own citizens[302] and was intended to curb the government collection of private information about its citizens.[303]

Specifically, the Privacy Act gives citizens the right to access private information about themselves maintained by government agencies, as well as a right to correct inaccurate information maintained by the government.[304] The Act also placed restrictions on the sharing of personally identifiable information (PII) with other government agencies and other entities.[305] Because the Privacy Act

298. Akhil Reed Amar & Renée B. Lettow, *Fifth Amendment First Principles: The Self-Incrimination Clause*, 93 MICH. L. REV. 857, 870 (1995).

299. Citron, *supra* note 248; *see also* Hatfield, *supra* note 19, at 332 n.77.

300. United States v. Warshak, 631 F.3d 266, 288 (6th Cir. 2010).

301. 5 U.S.C. § 552a (2012).

302. Evan Stone, *The Invasion of Privacy Act: The Disclosure of My Information in Your Government File*, 19 WIDENER L. REV. 345, 347–48 (2013).

303. S. REP. NO. 93-1183, at 1–2 (1974).

304. 5 U.S.C. § 552a(d)(1)–(2).

305. *The Privacy Act of 1974*, ELECTRONIC PRIVACY INFO. CTR., https://epic.org/privacy/1974act/ [https://perma.cc/RE4K-DFRE] (last visited Mar. 2, 2017) ("There are several exceptions to the Privacy Act for law enforcement. In addition, agencies have also gotten around the restriction on information sharing using the 'routine use' exemption.").

requires agencies to have a mechanism to allow individuals to correct mistakes in the information kept by the government, it appears that the IRS is violating this right by maintaining secrecy over the data being kept and created about individuals.[306]

It is generally understood in legal circles that certain information—medical, financial, location, etc.—is meant to be kept confidential.[307] Emails and private communications are undeniably meant to be protected against government searches and seizures.[308] It is expected that credit card information and banking information are to be kept confidential as well.[309] Government intrusion is not just a violation of law; it can result in a permanent loss of confidentiality, as well as significant problems for affected individuals if PII is publicly released.[310] US News and World Report reported:

> While the [IRS] has declined to give details about what third-party personal data it will use in robo-audits and data mining, it has told government and industry groups that its computers are capable of scanning multiple networks at the same time to collect "matching" comprehensive profiles for every taxpayer in America. Such profiles will likely include shopping records, travel, social interactions and information not available to the public, such as health records and files from other government investigators, according to IRS documents.[311]

Privacy experts, including the IRS National Taxpayer Advocate, have asked the IRS to make public the information it examines in audits in order to facilitate compliance.[312] However, this is not being done by the IRS. IRS Commissioner Steven T. Miller testified that the

---

306. *See* 26 U.S.C. § 8752(e) (2012) (limiting a taxpayer's ability to correct tax returns).

307. Mike Hatch, *The Privatization of Big Brother: Protecting Sensitive Personal Information from Commercial Interests in the 21st Century*, 27 WM. MITCHELL L. REV. 1457, 1463 (2000).

308. United States v. Warshak, 631 F.3d 266, 284–86, 291–92 (6th Cir. 2010).

309. *See* Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338 (codified as amended in scattered sections of 12 and 15 U.S.C.); Health Insurance Portability and Accountability Act, Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered sections of 29 and 42 U.S.C.).

310. In the oft-referenced Target mishap, a teenager's pregnancy was predicted based on her vitamin purchases, which caused much grief and stress for her family. *See* Landau, *supra* note 123, at 504 (citing Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES MAG. (Feb. 16, 2012), http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html [https://perma.cc/74L7-MHSY]). Uber has also claimed that it is able to identify customers engaging in one-night stands based on the data it collects about its riders. *Id.* (citing Voytek, *Rides of Glory*, UBER (Mar. 26, 2012), https://web.archive.org/web/20140827195715/http://blog.uber.com/ridesofglory [https://perma.cc/Q7SK-GZHN]).

311. Richard Satran, *IRS Data Web Snares Mostly Low- and Middle-Income Taxpayers*, U.S. NEWS & WORLD REP. (May 1, 2013, 12:35 PM), http://money.usnews.com/money/personal-finance/mutual-funds/articles/2013/05/01/irs-data-web-snares-mostly-low—and-middle-income-taxpayers [https://perma.cc/JYN9-263L].

312. *Id.*

"stealth approach" is "less intrusive,"[313] but Senator Charles Grassley (R-IA) said the IRS is not doing enough to stop its "abusive intrusion of privacy."[314] Section 11.3.14.12 of the IRS Manual referring to the Privacy Act indicates that the IRS may only collect information relevant and necessary to accomplish the purposes of the agency.[315] By collecting and amassing detailed data files on individuals, the IRS is violating its own requirements as well as the Privacy Act.

2. Computer Matching and Privacy Protection Act

The Computer Matching and Privacy Protection Act (CMPPA) is a 1988 amendment to the Privacy Act of 1974.[316] The purpose of the amendment, according to the IRS Manual 11.3.39, is to add:

> Certain protections for the subjects of Privacy Act records whose records are used in automated matching programs. These protections have been mandated to ensure:
>
> - Procedural uniformity in carrying out matching programs
> - Due process for subjects in order to protect their rights
> - Oversight of matching programs through the establishment of Data Integrity Boards at each agency engaging in matching to monitor the agency's matching activity.[317]

The CMPPA allows the matching of computer data when legal authority exists and it is appropriate for achieving the desired action.[318] CMPPA is intended to ensure privacy, integrity, and verification of any data disclosed for computer matching by the government.[319] There are four factors that must exist for CMPPA to apply: computerized comparison, categories of subjects, federal benefit program, and a matching purpose.[320] Although certain matching programs, such as the tax administration, are exempt from CMPPA,[321]

---

313. *Id.*

314. *Id.*

315. *11.3.14.12 Privacy Act General Provisions*, IRS § 11.3.14.12 (Sept. 12, 2013), https://www.irs.gov/irm/part11/irm_11-003-014.html#d0e426 [https://perma.cc/ZGY3-EDD2].

316. U.S. DEP'T OF JUSTICE, *Overview of the Privacy Act of 1974*, *in* DEPARTMENT OF JUSTICE MANUAL COMMENTARY § 3-17.000, at 3 (2015 ed.).

317. *11.3.39 Computer Matching and Privacy Protection Act*, IRS § 11.3.39.2 (Sept. 17, 2013), https://www.irs.gov/irm/part11/irm_11-003-039.html#d0e65 [https://perma.cc/J835-XV94].

318. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-14-44, COMPUTER MATCHING ACT: OMB AND SELECTED AGENCIES NEED TO ENSURE CONSISTENT IMPLEMENTATION 13 (2014).

319. Privacy Act of 1974: Final Guidance Interpreting the Provisions of Public Law 100-503, The Computer Matching and Privacy Protection Act of 1988, 54 Fed. Reg. 25,818-01, 25,818 (June 19, 1989).

320. *11.3.39 Computer Matching and Privacy Protection Act*, *supra* note 317, § 11.3.39.7.

321. 5 U.S.C. § 552a(a)(8)(B)(iv) (2012).

the Office of Management and Budget intended the law to cover the tax system.[322] The IRS is required to:

- Develop, execute and obtain approval of a written agreement, prepared in conformance with 5 USC § 552a(o), with the other agency or the other IRS function,
- Provide notice of the matching program to record subjects,
- Prepare a report to Congress on the new matching program[, and]
- Prepare any Federal Register notice and report required (unless prepared by the recipient agency).[323]

FOIA requires that each agency "maintain in its records only such information about an individual as is relevant and necessary to accomplish a purpose of the agency required to be accomplished by statute or by executive order of the President."[324] Maintaining these records violates the Privacy Act as well. In *Clarkson v. Internal Revenue Service*, the Eleventh Circuit held that the IRS improperly maintained records regarding the exercise of Clarkson's First Amendment rights.[325] The plaintiff, a tax protester, was followed and investigated by the IRS, who kept a file on him containing surveillance reports, newsletters, and press releases.[326] The court found that the collection and maintenance of these materials was in violation of the Privacy Act, even though the IRS contended that the records were not kept in a "system of records," since they were kept in a general "Tax Protest File" from which the IRS said it could not retrieve individual records by name.[327]

FOIA also provides that each agency shall "collect information to the greatest extent practicable directly from the subject individual when the information may result in adverse determinations about an individual's rights, benefits, and privileges under Federal programs." The IRS is required to prepare a notice in accordance with 5 U.S.C. § 552a(e) to notify the subject individuals that their records may be part of a matching program prior to the actual conduct of the matching. These are to be published in the Federal Register.[328]

---

322. *11.3.39 Computer Matching and Privacy Protection Act*, *supra* note 317, § 11.3.39.7.1.

323. *Id.* § 11.3.39.8.

324. § 552a(e)(1).

325. Clarkson v. Internal Revenue Serv., 678 F.2d 1368, 1374–77 (11th Cir. 1982).

326. *Id.* at 1369–70.

327. *Id.* at 1373.

328. § 552a(e)(4).

In conformance with the CMPPA, the IRS conducts Privacy Impact Assessments (PIA) on its collection of PII.[329] The PIAs ensure the following:

- The public is informed regarding the information that is collected;
- Any impact the collection may have on personal privacy is adequately addressed;
- The IRS collects sufficient personal information to administer its programs, and no more;
- The information collected is used only for the purpose intended;
- The information is maintained to be timely and accurate;
- The information is protected while the IRS has custody and the IRS has custody only for as long as is necessary;
- Information is withheld if its release might harm IRS systems, compromise law enforcement efforts or, jeopardize competitive businesses.[330]

The clear target of the Privacy Act was federal government agencies. The Privacy Act empowered citizens with a right of access to their federal government agency files along with a civil remedy to enforce that right.[331] The Privacy Act also required an agency to publicly announce its record systems that were meant to store information about citizens.[332] Further, the government agency could only maintain relevant and necessary information,[333] and government agencies were restricted from disclosing information to third parties without consent or specific exceptions.[334] By compiling and

329. *Privacy Impact Assessments—PIA*, IRS (Mar. 3, 2017), https://www.irs.gov/uac/Privacy-Impact-Assessments-PIA [https://perma.cc/NUK8-LLKQ].

330. *Id.*

331. § 552a(d)(1) ("Each agency that maintains a system of records shall upon request by any individual to gain access to his record or to any information pertaining to him which is contained in the system, permit him . . . to review the record and have a copy made of all or any portion thereof . . . ."); § 552a(g)(1)(B) ("Whenever any agency . . . refuses to comply with an individual request under subsection (d)(1) of this section . . . the individual may bring a civil action against the agency . . . .").

332. *Id.* § 552a(e)(4) ("Each agency that maintains a system of records shall . . . publish in the Federal Register upon establishment or revision a notice of the existence and character of the system of records . . . .").

333. *Id.* § 552a(e)(1) ("Each agency that maintains a system of records shall maintain in its records only such information about an individual as is relevant and necessary to accomplish a purpose of the agency . . . .").

334. *Id.* § 552a(b) ("No agency shall disclose any record which is contained in a system of records by any means of communication to any person, or to another agency, except pursuant to a written request by, or with the prior written consent of, the individual to whom the record pertains . . . .").

maintaining comprehensive data profiles of taxpayers, the IRS is violating the Privacy Act.

3. Internal Revenue Code Section 6013

Internal Revenue Code Section 6103 requires that tax returns and return information[335] be held confidential and not disclosed in any manner.[336] Where disclosure is permitted, Section 6103 generally imposes strict technical, administrative, and physical safeguarding requirements to prevent IRS employees from using or disclosing the returns and return information in an unauthorized manner.[337] It also requires the IRS to monitor and enforce compliance with those requirements.[338] This includes keeping records that detail inspections and disclosures of return information.[339] There are criminal penalties for IRS employees that willfully inspect data without authorization or disclosure of information, and taxpayers have the right to civil action for the wrongful inspection or disclosure of their return information.[340] However, there are numerous authorized exceptions allowing information to be shared with individuals or agencies having a material interest in the tax information.[341] For example, federal tax return information is available to any state agency responsible for state taxation to the extent necessary for the agency to fulfill its mandate.[342]

---

335. Under I.R.C. § 6103(b)(1) (2016), the term "return" means any tax or information return, declaration of estimated tax, or claim for refund required by, or provided for or permitted under, the provisions of this title which is filed with the Secretary by, on behalf of, or with respect to any person, and any amendment or supplement thereto, including supporting schedules, attachments, or lists which are supplemental to, or part of, the return so filed. Under Section 6103(b)(2)(A), the term "return information" means a taxpayer's identity, the nature, source, or amount of his income, payments, receipts, deductions, exemptions, credits, assets, liabilities, net worth, tax liability, tax withheld, deficiencies, over assessments, or tax payments, whether the taxpayer's return was, is being, or will be examined or subject to other investigation or processing, or any other data, received by, recorded by, prepared by, furnished to, or collected by the Secretary with respect to a return or with respect to the determination of the existence, or possible existence, of liability (or the amount thereof) of any person under this title for any tax, penalty, interest, fine, forfeiture, or other imposition, or offense.

336. *Id.* § 6103(a).

337. *Id.* § 6103(b)(5)(B).

338. *Id.* § 6103(p)(3).

339. *Id.*

340. 5 U.S.C. § 552a(i) (2012).

341. I.R.C. § 6103(d), (e), (i), (k), (l).

342. *Id.* § 6103(a), (d).

> (a) General rule Returns and return information shall be confidential, and except as authorized by this title—(1) no officer or employee of the United States, (2) no officer or employee of any State, any local law enforcement agency receiving information under subsection (i)(7)(A), any local child support enforcement agency, or any local

Section 6103(k) specifically covers disclosure of return information for tax administration purposes.[343] The investigation of a taxpayer through Internet searches may involve disclosures of tax information because a taxpayer's name and address is return information.[344] This disclosure is only permitted if it is in order to obtain information not otherwise reasonably available.[345] Thus, IRS Internet searches on taxpayers may in and of themselves violate I.R.C. § 6103(k).

4. Data Quality Act

The Data Quality Act requires federal agencies to take steps to ensure the quality of their data.[346] In response, the Office of Management and Budget (OMB) issued guidelines for federal agencies in order to ensure the "quality, objectivity, utility, and integrity" of

---

> agency administering a program listed in subsection (l)(7)(D) who has or had access to returns or return information under this section or section 6104(c), and (3) no other person (or officer or employee thereof) who has or had access to returns or return information under subsection (e)(1)(D)(iii), subsection (k)(10), paragraph (6), (10), (12), (16), (19), (20), or (21) of subsection (l), paragraph (2) or (4)(B) of subsection (m), or subsection (n), shall disclose any return or return information obtained by him in any manner in connection with his service as such an officer or an employee or otherwise or under the provisions of this section. For purposes of this subsection, the term "officer or employee" includes a former officer or employee. . . . (d) Disclosure to State tax officials and State and local law enforcement agencies (1) In general Returns and return information with respect to taxes imposed by chapters 1, 2, 6, 11, 12, 21, 23, 24, 31, 32, 44, 51, and 52 and subchapter D of chapter 36 shall be open to inspection by, or disclosure to, any State agency, body, or commission, or its legal representative, which is charged under the laws of such State with responsibility for the administration of State tax laws for the purpose of, and only to the extent necessary in, the administration of such laws, including any procedures with respect to locating any person who may be entitled to a refund. Such inspection shall be permitted, or such disclosure made, only upon written request by the head of such agency, body, or commission, and only to the representatives of such agency, body, or commission designated in such written request as the individuals who are to inspect or to receive the returns or return information on behalf of such agency, body, or commission. Such representatives shall not include any individual who is the chief executive officer of such State or who is neither an employee or legal representative of such agency, body, or commission nor a person described in subsection (n). However, such return information shall not be disclosed to the extent that the Secretary determines that such disclosure would identify a confidential informant or seriously impair any civil or criminal tax investigation.

Id.

343. *Id.* § 6103(k).

344. *Id.* § 6103(b)(2).

345. *11.3.21 Investigative Disclosure*, IRS (Apr. 27, 2016), https://www.irs.gov/irm/part11/irm_11-003-021.html [https://perma.cc/3WTV-GRC2].

346. 44 U.S.C. §§ 3504(d)(1), 3516 (2012).

information disseminated to the public.[347] The guidelines also address the sharing of information between federal agencies and require each agency to develop its own data quality assurance guidelines.[348] This includes the requirement that each agency develop a mechanism for individuals to correct information contained in that agency's records.[349] The IRS is instructed to provide taxpayers access to tax returns, tax return transcripts, and open-case-file work papers and records.[350]

The IRS has developed its own Information Quality Guidelines pursuant to the Data Quality Act to comply with the requisite data quality assurance.[351] All information and methodologies used by the IRS are to be consistent with professional standards.[352] The IRS is charged with ensuring the accuracy of the data contained on individuals in its Customer Account Data Engine 2 (CADE2).[353] The Treasury Inspector General for Tax Administration (TIGTA), who audits the IRS for compliance with the Data Quality Act, noted that inaccurate data could make the CADE2 database ineffective.[354] In 2014, the TIGTA audited the IRS's validation testing process to ensure that the databases upon which CADE2 is built were accurate and complete and found that the automated validation comparison tools and data-sampling methodology were sound, but the supporting documentation was seriously lacking.[355] Further, the data coverage and data defect reporting required improvement.[356] Some of the tools used to compare and trace data back to CADE2 were insufficient for validation; thus, the accuracy or completeness of the data is

347. John D. Graham, *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*, WHITE HOUSE (Sept. 24, 2001), https://obamawhitehouse.archives.gov/omb/fedreg_final_information_quality_guidelines/ [https://perma.cc/C66Z-T8B2].

348. CURTIS W. COPELAND & MICHAEL SIMPSON, CONG. RESEARCH SERV., RL32532, THE INFORMATION QUALITY ACT: OMB'S GUIDANCE AND INITIAL IMPLEMENTATION 2 (2004).

349. *Id.*

350. *Routine Access to IRS Records*, IRS (Dec. 9, 2016), https://www.irs.gov/uac/routine-access-to-irs-records [https://perma.cc/SME7-P7DN].

351. INTERNAL REVENUE SERV., INFORMATION QUALITY 1 (2002), https://www.irs.gov/pub/irs-utl/infoqualityguidelines.pdf [https://perma.cc/F3XG-2E2E].

352. *Id.*

353. *See generally* TREASURY INSPECTOR GEN. FOR TAX ADMIN. OFFICE OF AUDIT, CUSTOMER ACCOUNT DATA ENGINE 2 DATABASE VALIDATION IS PROGRESSING; HOWEVER, DATA COVERAGE, DATA DEFECT REPORTING, AND DOCUMENTATION NEED IMPROVEMENT (2014), https://www.treasury.gov/tigta/auditreports/2014reports/201420063_oa_highlights.pdf [https://perma.cc/G84S-KQZD].

354. *Id.*

355. *Id.*

356. *Id.*

questionable.[357] While the IRS agreed with the TIGTA's recommendation to develop or improve its documentation, including a manual for data validation, it disagreed with the detail necessary for traceability of defects to unique data fields, claiming it was not compatible with maintaining consistency across systems.[358] This is most likely due to the large data sets the IRS is purchasing and discovering through data mining.[359]

IV. POTENTIAL MISUSE OF DATA AND ALGORITHM BY IRS

Although the IRS must be able to verify information provided on tax returns, new technology has created a situation where current law may not sufficiently protect US citizens from government abuse and negligence. However, the IRS has very broad powers to identify and investigate potential tax evaders, and due to both their reduced budget and fewer employees, the IRS is turning to computers to identify and investigate violators.[360] The IRS is collecting vast amounts of data on US citizens, combining it with private information found on individual tax returns, and compiling an incredibly detailed dossier on all US citizens.[361] This is problematic because these activities not only violate current law, but the IRS's history suggests that continuing down this path could be very dangerous for US citizens. The following Sections explain the potential harms that could result.

A. Data Breach

The US government has had several major data breaches in recent years.[362] The IRS's collection of personal data is creating a very desirable target for identity thieves. In addition, the TIGTA recently reported that 21% of FOIA/Privacy Act information requests answered

357. *Id.*

358. *Id.*

359. Kerr, *supra* note 5; *see also* Sampson, *supra* note 5; *Report: IRS Data Mining Facebook, Twitter, Instagram and Other Social Media Sites*, *supra* note 5.

360. Robinson, *supra* note 4.

361. *Id.*; *see also* Kerr, *supra* note 5; Sampson, *supra* note 5; *Report: IRS Data Mining Facebook, Twitter, Instagram and Other Social Media Sites*, *supra* note 5.

362. *Massive IRS Data Breach Much Bigger than First Thought*, CBS THIS MORNING (Feb. 29, 2016, 7:00 AM), http://www.cbsnews.com/news/irs-identity-theft-online-hackers-social-security-number-get-transcript/ [https://perma.cc/KEJ7-XV3T]; *see also IRS Computer Center*, CTR. LAND USE INTERPRETATION (Dec. 16, 2011, 12:50 PM), http://clui.org/ludb/site/irs-computer-center [https://perma.cc/TQ34-PKDD] (showing the location of IRS data center).

by the IRS wrongly disclosed "sensitive taxpayer information."[363] The IRS has a horrible record in keeping the American public's private information private. Recently, the IRS reported that more than 700,000 Social Security numbers had been stolen from the "Get Transcript" function on its website.[364] A 2015 audit of security procedures at the IRS performed by the Government Accountability Office found that the IRS had ignored previous audit recommendations and was failing to keep taxpayer data secure.[365] The report listed forty-three deficiencies, including the failure to encrypt its data.[366]

Given the highly sensitive information kept by the IRS, it is risky to allow it to track and maintain large data sets about US citizens.[367] The IRS's failure to comply with the Privacy Act's instructions—to only use relevant private information, not share that

---

363. Zoe Crain, *New TIGTA Report Reveals More IRS Incompetence*, AMERICANS TAX REFORM (Oct. 1, 2014, 4:23 PM), https://www.atr.org/new-tigta-report-reveals-more-irs-incompetence [https://perma.cc/5NE2-457X].

364. *Massive IRS Data Breach Much Bigger than First Thought*, *supra* note 362.

365. *See* Stephen W. Mazza, *Taxpayer Privacy and Tax Compliance*, 51 U. KAN. L. REV. 1065, 1102–03 (2003). *See generally* U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-15-337, INFORMATION SECURITY: IRS NEEDS TO CONTINUE IMPROVING CONTROLS OVER FINANCIAL AND TAXPAYER DATA (2015), http://www.gao.gov/products/GAO-15-337 [https://perma.cc/E9UG-LGNX].

> The GAO has been warning about problems with IRS security since it started writing these reports in 2007. In each report, the GAO has issued recommendations for the IRS to improve security. After each report, the IRS did a few of those things, but ignored most of the recommendations. In this year's report, for example, the GAO complained that the IRS ignored 47 of its 70 recommendations from 2015. In its 2015 report, it complained that the IRS only mitigated 14 of the 69 weaknesses it identified in 2013. The 2012 report didn't paint IRS security in any better light.

Id.

Bruce Schneier, *Can You Trust IRS to Keep Your Tax Data Secure?*, CNN (Apr. 13, 2016, 8:42 AM), http://www.cnn.com/2016/04/13/opinions/is-data-you-send-to-irs-secure-opinion-schneier/ [https://perma.cc/QP3Y-B8LR].

366. Schneier, *supra* note 365.

> Every year, the GAO—Government Accountability Office—reviews IRS security and issues a report. The title of this year's report kind of says it all: "IRS Needs to Further Improve Controls Over Financial and Taxpayer Data." The details are ugly: failures in identification and authentication of network users, failures to encrypt data, failures in audit and monitoring and failures to patch vulnerabilities and update software.

Id.

367. Paul Caron, *The IRS Scandal, Day 1037*, TAXPROF BLOG (Mar. 11, 2016), http://taxprof.typepad.com/taxprof_blog/2016/03/the-irs-scandal-day-1037.html [https://perma.cc/5RVQ-L28D].

information with others, and discard it after use—puts everyone at risk, especially since they are not disclosing what is being kept.[368]

B. Misuse of Information and Targeting by Government

As detailed in Part II.B., the government has a long history of misusing the audit function.[369] This problem is exacerbated by the extensive information now being amassed on taxpayers.[370] During the 2012 presidential election, the IRS started flagging conservative political groups for additional reviews to see if they were violating their tax-exempt status.[371] According to Lois Lerner, head of the IRS division that oversees tax-exempt groups, organizations with the words "tea party" or "patriot" in their applications were targeted.[372] "In almost every administration since the IRS's inception," wrote David Burnham, author of *A Law Unto Itself: Power, Politics and the IRS*, "the information and power of the tax agency have been mobilized for explicitly political purposes."[373]

In 1942, the US Census Bureau began supplying data regarding the whereabouts of Japanese-Americans to facilitate their removal to internment camps.[374] Over 100,000 names were eventually provided to the military,[375] and these individuals were then held in internment camps until the end of World War II.[376] The records

368. Stone, *supra* note 302, at 345–48; *see also* S. REP. NO. 93-1183 (1974), *as reprinted in* COMM. ON GOV'T OPERATIONS, 94TH CONG., LEGISLATIVE HISTORY OF THE PRIVACY ACT OF 1974, at 154–55 (Comm. Print 1976) (expressing the Senate's purposes of its privacy bill as: respecting privacy, accountability, responsibility, oversight, open government, prevention of illegal and secret information gathering); *id.* at 295–97 (expressing similar concerns as those expressed by the Senate).

369. *See* Caron, *supra* note 367; *see also supra* Section II.B. This is an ongoing problem, as illustrated by the IRS's recent targeting of conservative 501(c)(4) tax-exempt status applicants.

370. *See supra* Section II.B.

371. Stephen Ohlemacher, *IRS Apologizes for Targeting Conservative Groups*, YAHOO! NEWS (May 10, 2013), http://news.yahoo.com/irs-apologizes-targeting-conservative-groups-144349480.html [https://perma.cc/2LJN-47HR].

372. *Id.* This statement was made at a 2013 American Bar Association conference where Lerner was a speaker. *Id.*

373. *The IRS's Long History of Scandal*, *supra* note 50.

374. MAYER-SCHÖNBERGER & CUKIER, *supra* note 120, at 151.

375. Haya El Nasser, *Papers Show Census Role in WWII Camps*, USA TODAY (Mar. 30, 2007, 1:33 AM), http://usatoday30.usatoday.com/news/nation/2007-03-30-census-role_N.htm [https://perma.cc/VN8R-ULT3].

376. *Id.*

included name, address, age, sex, citizenship status, and occupation of Japanese Americans in these block areas.[377]

Many of the IRS's problems have been detailed publicly, but even members of the Senate have criticized the continuous mistakes made by this agency.[378] Senator Thune (R-SD) posted the following on his website:

> A look under the IRS'[s] hood exposes systemic troubles that continue to throttle quality taxpayer services. Even 16 years after Congress passed sweeping taxpayer rights laws, a culture of mismanagement continues to steer the IRS away from sorely needed public redemption. Instead, misguided decisions and more violations of taxpayer privacy clog its core mission to serve the taxpaying public with integrity. . . . What's worse, the Government Accountability Office found that the IRS sent out $5.8 billion in fraudulent tax refunds in 2013. Considering the recent massive data breaches at the IRS and Office of Personnel Management, the federal government is clearly facing a steep curve to thwart cyber crimes that put sensitive personal information at risk of piracy.[379]

C. Surveillance by Government (Big Brother)

Government is increasing surveillance and, despite laws prohibiting the sharing of data among federal agencies, such data sharing may become possible through combined data centers.[380] NSA data centers are currently collecting and storing information, making it easier for authorities to search for information already stored in its databases rather than having to start from scratch when a suspect is identified.[381] These cost saving measures could result in agencies having access to the IRS database without citizen consent, in direct contravention of data protection laws. This potential for sharing data would seem to violate the Privacy Act of 1974, the CMPPA, and I.R.C. § 6103 because of the reduced costs in locating information from multiple agencies on shared servers and the fact that they are warehoused in the same locations. The intent of many of these laws is to make sure the public is aware of what information the government is collecting on them, as well as the ability to correct information about them being used by the government to make decisions

---

377. *Id.* In Europe, the Netherlands governmental records were used by Nazis to round up and persecute the Jews (the numbers imprinted on their arms came from the IBM Hollerith punch-cards numbers used by the data processing facilities at the time). MAYER-SCHÖNBERGER & CUKIER, *supra* note 120, at 152.

378. John Thune & Chuck Grassley, *Accelerate Customer Service at the IRS*, JOHN THUNE (June 19, 2015), http://www.thune.senate.gov/public/index.cfm/op-eds?ID=aab87cf4-4117-4ab6-b8f0-ec54ab1acedb [https://perma.cc/3S7M-PMLK].

379. *Id.*

380. MAYER-SCHÖNBERGER & CUKIER, *supra* note 120, at 157.

381. *Id.*

concerning their rights.[382] The secrecy surrounding the data analytics program and the type of information already held by the IRS is creating an environment where unchecked surveillance can not only create dire consequences for the public but also will preclude determining whether someone is targeted for an audit by a legitimate machine decision or political motivation.

V. CONCLUSION

To cope with the ever-increasing tax gap between what taxpayers owe and what they pay, as well as the steady decline in its budget, the IRS has turned to big data, data mining, and predictive analytics.[383] For the IRS, data analytics is not trying to predict the future behavior of taxpayers, but predicting data that it does not have; that is, predicting whether tax returns are compliant with the tax law.[384] There are serious issues with their collection of data, mining of data, and use of data.

The IRS is working on validating its databases, but the TIGTA found that there are problems with these data sets and improvements are necessary to ensure the accuracy of information collected on taxpayers.[385] The IRS has been hesitant to trace defects in its data to particular data fields, meaning that inaccurate data for a particular individual may not be discovered, disclosed or corrected.[386]

While individuals may consider their social media posts to be private communications, when they make them available to the public the IRS may view them. However, as the IRS collects this public information and adds it to its private information, confidentiality and privacy concerns become apparent. The IRS's databases are targets for identity theft, as seen by the massive breaches in recent years.[387] This is in addition to the IRS itself divulging private information to inappropriate parties[388] and continually failing to protect taxpayer information.[389]

---

382. Jerome, *supra* note 158, at 229.

383. Hatfield, *supra* note 19, at 322–23.

384. *Excerpt from Commissioner John Koskinen's Senate Finance Committee Testimony*, *supra* note 118.

385. *See generally* TREASURY INSPECTOR GEN. FOR TAX ADMIN., REF. NO. 2016-20-094, ANNUAL ASSESSMENT OF THE INTERNAL REVENUE SERVICE INFORMATION TECHNOLOGY PROGRAM (2016), https://www.treasury.gov/tigta/auditreports/2016reports/201620094fr.pdf [https://perma.cc/7BWG-SR6T].

386. *Id.*

387. *Massive IRS Data Breach Much Bigger than First Thought*, *supra* note 362.

388. *Id.*

389. *See supra* Section IV.A.

With the budget reductions and loss of 14% of its staff over the past several years, the IRS has been forced to do more with less.[390] In turn, the IRS has chosen to use machines (rather than employees) to make decisions.[391] This entails the obvious benefit of efficiencies in data collection and the ability to locate tax evaders.[392] There is another legitimate concern that if the algorithms were made public, taxpayers could find a way to game the system.[393] However, this concern should not preclude consideration of the other concerns raised in this article.

One of these other concerns is the fact that audits are both extremely stressful and costly to defend.[394] They have also been used as a political weapon by presidents and the government in the recent past.[395] Furthermore, big data results are based on correlation, not causation, and it is inappropriate to judge people based on correlation; just because people share characteristics or interests does not mean that they will have similar tax compliance behavior.[396] If, for example, people with dachshunds are associated with overstating medical expenses, is it appropriate to audit the medical expenses of everyone with a dachshund? While this may seem like an unlikely example, imagine if the commonality was race or religion. If audit targeting is based on correlation, rather than causation, this can easily lead to profiling and discrimination.[397]

There is an enormous difference between selecting returns for audits based on a comparison between a taxpayer's own return and required third party filings (such as W-2s), and those based on an unverified computer algorithm using data mined from the Internet.[398] The secrecy surrounding the use of big data and predictive analytics by the IRS makes it difficult to flesh out how the audit function is influenced by the use of big data, and the extent to which the IRS audit-selection process is violating the law.[399] It does seem clear, however, that because of the IRS's budget woes, it is turning more and

390. Marr & Murray, *supra* note 2.

391. Robinson, *supra* note 4.

392. INTERNAL REVENUE SERV., *supra* note 103, at 37.

393. *4.1.3 Sources of Returns-Priority Programs-DIF and Ordering*, IRS (Aug. 10, 2012), https://www.irs.gov/irm/part4/irm_04-001-003.html [https://perma.cc/G8QH-4F95].

394. Spors, *supra* note 219; *How to Ease the Stress of a Tax Audit*, TOP TAX DEFENDERS (May 1, 2012, 6:00 AM), http://www.toptaxdefenders.com/blog/bid/137112/How-to-Ease-the-Stress-of-a-Tax-Audit [https://perma.cc/5EVA-SN6D].

395. EXEC. OFFICE OF THE PRESIDENT, *supra* note 184, at 66.

396. *See supra* Section III.B.3.

397. *See id.*

398. *See supra* Section III.B.

399. *See id.*

more to data analytics.[400] More transparency by the IRS regarding its data collection, data mining, and predictive algorithms would help to ensure compliance with the Constitution and laws regarding due process and privacy.[401] What is being sold as an efficient fraud detection system may actually be the end of privacy as we know it.

400. Robinson, *supra* note 4.

401. *See* Citron, *supra* note 188, at 1249–313, 1293 n.302.