Aidan M. Levinson

Professor Joshua Harms

WRTG-100-064

9 December 2019

Data Science In Politics and Privacy Concerns

As a relatively new field, data science has evolved as a pivotal part of most every organization's planning and operations. The UC Berkeley School of Information defines the role of a data scientist as someone "able to identify relevant questions, collect data from a multitude of different data sources, organize the information, translate results into solutions, and communicate their findings in a way that positively affects business decisions" (UC Berkeley School of Information 2019). Not only can data science be used to positively affect business decisions, but it can and has been used more recently as an important tool at the disposal of political campaigns and party organizations. Today, campaigns utilize data science as an important and necessary part of their electoral plans and general operation. The former Democratic National Committee (DNC) Director of Data Science during the majority of the 2016 presidential campaign, Andrew Therriault, offers quite a unique perspective on this new way to utilize data science and big data in the changing technological landscape of society. In Therriault's article *Finding a Place in Political Data Science* in the American Political Science Association's <u>Profession Symposium</u>, he examines the uses of applying data science to political operations:

> To make things more efficient, we typically employ some form of targeting: choosing the most persuadable voters to contact, finding the best TV channels to run ads on, or identifying the best potential donors to send fundraising mail, for example. And to make things more effective, we use testing and analysis to determine which messages, tactics, or strategies will have the greatest impact (Therriault 531).

Therriault's examination of the use of data science as a means to target voters is a warning sign considering the accessibility of information and the potential to manipulate for a political operative's own purposes. As its use for political campaigns and party organizations has increased, data science and its methods have been improved and sharpened over this time. Although the field of data science and its use of data mining and big data is a remarkable and innovative trend in political campaigning that is remaking the political system from the ground up, the American public needs to be on guard and keep in mind the potential for the violation of data privacy, a threat that is continuing to grow not only at home but also abroad.

The crux of data science as applied to political activities stems from the use of big data, massive sources of data that are being constantly updated. Most common in politics when it comes to big data is the voter file, a data repository containing a wealth of information on the voting population. The voter file is used to target voters for many of the available outreach methods at the disposal of political campaigns and party organizations. Data science is utilized in the operations as a means to interpret the information found within the voter file, a source of big data, utilized by political operatives to draw meaningful conclusions on which voters to target to achieve their electoral or other goals. It is a central duty of these political data science professionals to use big data sources such as the voter file to target voters based on their ideologies, voting history, etc. all found within the voter file. This information and the interpretations of it made by data scientists can be then used to drive the social media/digital campaigns, as well as targeting voters for mailings, canvassing (door-to-door), phone calls, and donors for fundraising campaigns. David Nickerson and Todd Rogers' paper *Political*

Campaigns and Big Data in the Harvard Kennedy School Faculty Research Working Paper

Series delves deeper into this idea of targeting voters based on their ideologies and behaviors:

> Contemporary political campaigns amass enormous databases on individual
> citizens and hire campaign data analysts to create models predicting citizens' behaviors,
> dispositions, and responses to campaign contact. This new technology allows campaigns
> to simultaneously target campaign outreach tactically at particular individuals … to
> inform large-scale strategic decisions (Nickerson and Rogers 2).

Essentially, these massive data sources have become invaluable to the modern day political

campaign. The wealth of information stored within these data sources and the data scientists who

interpret the information that has afforded campaigns the opportunity to effectively target their

message to voters that are pivotal to their electoral efforts and goals has led to a potential threat.

As time goes on, political campaigns and parties not only continue to utilize more data

sources but also they continue to improve the methods used to interpret them furthering the

concern of privacy. In Helen Margetts' article *The Data Science of Politics* in Political Studies

Review, she highlights the continued development of methods used to interpret big data on

voters and the like. Margetts recognizes that as technologies continue to change and the general

public's reliance on them continues to increase, political campaigns and parties are moving to

attempt to siphon data on voters from sources like social media, online digital interactions with

political content, etc. to add to their big data repositories. Marrgetts also reports that these data

professionals enlisted for political efforts need to learn new and creative ways to adapt to these

new ways to gather data on voters to utilize for electoral outreach efforts (Margetts 204). Modern

campaigns have recognized this need to adapt to this new method of gathering data from digital

sources and seized on the opportunity to improve their targeting and knowledge with it.

However, the massive network of data sources used today has intensified the peril of voter data privacy.

The first real pioneering attempt of utilizing data science and big data for campaign operations was during the 2008 and 2012 presidential campaigns of President Barack Obama. President Obama's presidential campaigns truly set the stage for modern-day developments in the field of political data science and the use of big data for electoral and campaign operations. In Andrew Thrriault's book *Data and Democracy*, veteran political data scientist and current political digital marketer Daniel Scarvalone recognizes in his essay *Digital Advertising in the Post-Obama Era* the contributions the Obama campaigns made to the political data science field. Scarvalone claims "President Obama's presidential campaigns revolutionized the way technology and data could be used together to identify and speak to voters' interests, reflecting the increasingly sophisticated application of digital marketing to politics" (Scarvalone 31). Heading into the 2016 presidential elections, the consultants and staff of the campaigns during that time built on the progress made by the Obama campaigns to hone their data efforts.

This honing of skills for using data was clearly evident in social media and digital advertising. Scarvalone investigates this point writing:

> As voters spend more time than ever online, digital advertising combines the targeting, persuasion, and measurement capabilities that move the most votes at the least cost … This shift is critical to what will make or break the future of political campaigns, as well as how corporations talk to consumers or advocacy organizations talk to supporters (Scarvalone 32).

As campaigns, especially those of the 2016 election cycle, recognized the increasing shift of voters' daily lives to the world of digital and social media, they adapted, advertising directly to voters where they spend most of their time: online. Data siphoned from online sources are

driving political campaigns and parties' efforts to move their advertising online as to be more effective and drive their message to their targeted voters and supporters. Scarvalone noted this is an important and necessary move seeing as it is less expensive than other advertising options, and at the same time more effective. That being said, voters, political activists, and those that are tuned in to political happenings can expect to see more of these digital advertisements. As previously stated, these methods are becoming better and better at doing what they do, and because of that more and more political campaigns and party organizations will use these methods as a pivotal and important part of their electoral plans and outreach strategies. With data science being the driving force behind these moves and operations, we can expect to see an increasing reliance on the field and the professionals that practice it in order to drive the expected expanse of these digital and data-driven operations in political campaigning and general political persuasion.

With the development and expansion of the political data science field and use of big data for political purposes albeit guaranteed, the question begs to be asked, what are the drawbacks? The fact of the matter is that data privacy plays a major role in the concerns voters should have over the data that can be mined: ideology, voting history, locality information, and demographics so readily available, and at the disposal of hundreds of campaigns and party organizations across the nation. The large amount of big data at the disposal of these campaigns and organizations presents a data privacy concern that should be at the forefront of every voter's mind when considering their informational vulnerability.

Big data and data science's improvements in their methods and techniques present a major concern in terms of personal data privacy. Big data collected on voters does not just apply

to political activities here in the United States of America, but it also applies to political

operations in other democracies as well. Turning focus to the United States' allies across the

Atlantic, Laszlo Dellei of Information Systems Audit and Control Association (ISACA)

published a blog post on their news site entitled *Data Security and Access to Voters' Personal*

Data by Political Parties: An EU Case Study. Dellei highlights in his post that election data

integrity concerns do not just apply to United States presidential elections alone, but they also

pose serious implications for the elections in the European Union as well. Dellei explains:

> During microtargeting, information may be used to deliver political messages to the
> recipients … Candidates and nominating organizations (mostly political parties) may
> request the names and addresses of voters in the voter register from the relevant electoral
> office for campaign purposes. The information may be provided by age, gender, or
> address of the data subjects (Dellei).

Dellei recognizes the continuing danger to the privacy of voter information in political

operatives' efforts to microtarget their advertising and other communications operations to voters

that are persuadable or of a certain ideology in order to be successful on election day and

beyond. The privacy aspect of this development comes into play where information on voters is

readily accessible for purposes voters may not want it used for in political campaigning. There is

also a perception that it is up to those voters to what level they want their data used for political

purposes. However, the average voter is likely to be completely unaware of the data available to

others concerning their personal information. No matter the case, as these sources of big data on

voters become more and more vital to campaigning and political activities in general, it is highly

likely that the databases will continue to grow.

     With these expected developments in expanding big data, it is likely that the privacy

concerns over voter data will have a greater potential for being compromised. New forms of data

are bound to be created that can bring even greater accessibility to voter's personal information only deepening privacy fears of voters at home and abroad. Chris Evans' law journal article, *It's the Autonomy Stupid: Poltical Data-Mining and Voter Privacy in the Information Age* in the *Minnesota Journal of Law, Science, and Techonolgy*, draws the clear connection between data science, big data, and data-mining developments to the overarching privacy concern that implicates political operations. Evans embodies this point charging, "Information Age political tactics are unraveling the anonymity afforded by the secret ballot. To more effectively target voters, campaigns have become voracious collectors of personal data" (Evans 867). What Evans correctly identifies in his journal article is that the anonymity that most every voter has enjoyed in recent history until the mass data sources on voters came about is in jeopardy. Now, unlike ever before, voting history, political ideologies, and personal as well as demographic information is at the disposal of political data operatives who run political campaigns and political organizations on a daily basis.

Data mining is another aspect of the political data science field that Evans explored, explaining data mining as the process of siphoning data from consumers' online habits to learn of their shopping habits, locality, income, and even medical conditions (Evans 881). So essentially, the data that is accessible by political campaigns and organizations is not limited to just ideological and demographic information, but also may include online and some offline activities. This data can be utilized to predict not just consumer behavior but political behavior, as well, based on the actions of others who share similar habits. This threatens a whole new domain of political data privacy. The point that encircles this entire debate concerning political data privacy is this: now more than ever due to recent developments in the field of data science,

acquiring data on voters has become easier and more frequent, effectively shifting the control

concerning voters' privacy over their information to political operatives.

Evans utilized the example of the innovative campaign of President Barack Obama in

both the 2008 and 2012 presidential elections. He recognizes the pioneering of the use of

methods such as data mining for political purposes writing:

> New methods of politicking yield more data. President Obama's campaign Facebook
> page has twenty-three million "likes" and an app to gather data on all those twenty-three
> million users. President Obama's campaign shares data from field-level organizers up the
> hierarchy through its own social networking tool (Evans 886).

In practice, the Obama campaign of both 2008 and 2012 utilized these innovative data

acquisition and utilization methods in order to carry out their electoral plans. The campaign

collected information on their biggest supporters through their own platform as well as other

platforms (social media platforms, shopping websites, advertising agencies, etc.) using data

mining and big data that exist on those individuals. The two campaigns of President Barack

Obama are prime examples of how data can be exploited for political purposes without the

consent of the voters to use their own personal information for political operations.

Political campaigns, parties, and organizations have become increasingly reliant on data

science, big data, and data mining and as a result are getting better at targeting potential voters.

While the developments in the field of political data science and political data in general are

revolutionary and of interest to many, the invasion of privacy that comes with it must be of

concern to voters across our nation and the globe in other democracies. At a time when personal

and other data are constantly exploited for corporate and organizational gain, success, and

strategy, it can not be forgotten that it is being exploited for political use as well. Millions of

Americans each and every day are barraged with political content they were targeted for based

on personal, ideological, and other information accessed by political operatives. The question that voters need to ask themselves is: are they are comfortable with this invasion of privacy, and if so, to what extent? Those concerned about data privacy often sweep the political aspect under the rug. Therefore, exposing this threat is vital to keeping a democracy strong and Americans' privacy respected. Personal data is the individual's alone, and until citizens start and continue these conversations concerning these issues, data privacy can not be fully ensured.

Concurrently, the question must be asked over how much freedom these political campaigns, parties, and organizations should have when obtaining and utilizing these data sets for their own purposes. Currently, there is next to no regulation over the use of data for these purposes. Voters and the American public must ask themselves: What should be the role of big data in political campaigning and how much control should political campaigns and organizations have over the acquisition and distribution of this data? In his journal article, Chris Evans recognizes the possibility of implementing laws that cover this topic and concern in the United States which can be modeled globally. Evans outlines the priorities that a data privacy law would have to set: "A successful voter privacy policy must satisfy three general privacy constraints: (1) balancing free speech with ownership, (2) providing meaningful notice and consent, and (3) holding data-miners accountable to individuals and society" (Evans 894). Furthermore, Evans warns of the consequences were a data privacy law not enacted. Specifically individual privacy would be eroded: "Absent new restrictions on political data-mining, the low cost and easy availability of political information such as voter registration, voting history, and campaign contributions, may actually slow (but not stop) the erosion of privacy" (Evans 892).

Regulating political data can and should be considered by democracies across the globe, while balancing the needs of campaigns and political organizations. The question must be asked by voters and lawmakers together what level of data use and access should be in political campaigning, and what restrictions are necessary to make it fair while respecting the privacy that voters in the country and the nations of the world so deserve because it is their private data being used without their consent.

Works Cited

Dellei, Laszlo. "Data Security and Access to Voters' Personal Data by Political Parties: An EU

Case Study." *ISACA*, 2019, https://tinyurl.com/yxxry6qp. Accessed 25 Nov. 2019.

Evans, Chris. "It's the Autonomy, Stupid: Political Data-Mining and Voter Privacy in the

Information Age." *Minnesota Journal of Law, Science and Technology,* vol. 13, no. 2,

2012, p. 867-898. *HeinOnline*, https://tinyurl.com/tdondd3.

Margetts, Helen. "The Data Science of Politics." Political Studies Review, vol. 15, no. 2, SAGE

Publications, May 2017, pp. 201–09, https://tinyurl.com/ugw49p2.

Nickerson, David W., and Todd Rogers. "Political Campaigns and Big Data." HKS Faculty

Research Working Paper Series RWP13-045, Revised February 2014.

Therriault, Andrew. Finding a Place in Political Data Science. Vol. 49, no. 3, Cambridge

University Press, July 2016, pp. 531–34, https://tinyurl.com/vuvphyf.

Therriault, Andrew. *Data and Democracy*. 1st ed., O'Reilly Media, 2016.

UC Berkeley School of Information. "What is Data Science?" *datascience@berkeley*, Berkeley

School of Information, 2019, https://tinyurl.com/y6htykdb. Accessed 16 Nov. 2019.