International Journal of Communication 4 (2010), 1032–1050 1932–8036/20101032

New Challenges to Political Privacy: Lessons from the First U.S. Presidential Race in the Web 2.0 Era

DANIEL KREISS
Yale Law School

PHILIP N. HOWARD
University of Washington

Pundits and scholars laud online campaigning for its potential to democratize politics and praise the 2008 Barack Obama campaign for using new information technologies to mobilize voters. Underneath these extraordinary forms of technologically-enabled political participation, however, is an infrastructure and industry for political data that has received far less attention. To help fill this gap in scholarly understanding, we provide an overview of the data practices of political campaigns over the last decade and take a particularly close look at many of the new tools used by the Obama campaign. As a call for further research, we then outline a set of potential normative concerns about this use of data. We suggest that the data practices of campaigns and other political organizations may undermine important democratic norms. Campaigns erode privacy and narrow political debate by using data on citizens and social networks to tailor messages and communicate with narrowly-defined segments of voters. The lack of policy oversight erodes institutional transparency and leaves citizens vulnerable to breeches in personal data.

Introduction

One month before the 2008 presidential election, Barack Obama's campaign sent an e-mail to its earliest supporters, purportedly from Campaign Manager David Plouffe. Plouffe reminded them that they were "one of the first million people to own a piece of this campaign" and "helped build this movement when the odds were long and Election Day was far in the future" (D. Plouffe, personal communication, September 28, 2009). Over the course of the long primary and general election, 13 million people signed up to receive Obama e-mails. They too received communication from the campaign a month before the election, and many other days besides, tailored to their own involvement. Indeed, an author of this article received 313 e-mails from the Obama campaign in a little over a year.

In language that echoes Plouffe's message to supporters, many pundits and academics argue that millions of citizens organizing from the bottom-up were the driving force behind the Obama campaign. In this telling, a major technological shift from static Web pages to what in the marketing

language of Silicon Valley is known as the social technologies of "Web 2.0" (Scholz, 2008) culminated in a radically participatory "movement" to elect Obama. Citizens planned fund-raising events and knocked on the doors of their neighbors identified by the campaign, all while using an unprecedented array of digital campaign tools. These tools included the campaign's own social networking platform and campaign site, MyBarackObama.com (MyBO), as well as an array of commercial technologies, such as Facebook and MySpace. All of which led many to agree with media theorist Henry Jenkins when he stated that the Obama campaign was "as much structured around connections between voters as it is around connections between the candidate and the electorate" (Jenkins, 2008).

This account of the Obama campaign, however, mentions little about political data. Yet, as Plouffe (2009) makes clear in his firsthand account, data was central to the campaign's ability to take on a well-funded front-runner in the primary and trounce the Republican candidate during the general election. Millions of volunteer phone calls to voters, click-throughs on the campaign's Web site, and small donations generated data that the campaign then acted upon. Targeted door-to-door voter contacts updated databases so the campaign quickly knew the strength of its support in key districts and where to deploy more resources. The campaign continually analyzed user behavior on its Web site and made minor adjustments in design and content that increased the percentage of citizens who signed up for its e-mail list and donated money. Together, the electoral databases maintained by the Democratic National Committee and by the private firm Catalist, in conjunction with a vast array of online behavioral and relational data collected from use of the campaign's Web site and third-party social media sites such as Facebook, provided the Obama campaign with data on more than 250 million Americans.

The Obama campaign was only the fullest realization of trends in the political field toward crafting better means of collecting, storing, analyzing, and acting upon data about citizens, their online behavior, and their social relationships. At various points over the last two decades, both parties embarked on extensive projects to build comprehensive voter files and model the electorate so campaigns would have a good sense of who their supporters were, which individuals were open to persuasion, and which citizens were likely to turn out. While cable and broadcast television advertising outlays still dominate campaign spending (Miller, 2008), the profusion of technologies since the 1990s facilitating more targeted communication to a fragmented electorate—from e-mail and banner advertising on the Internet to advertising in video games—offer new opportunities for campaigns to "narrowcast" persuasive political communication based on media consumption habits (Howard, 2006). While direct mail attests to the fact that narrowcasting has been around for over 20 years, when combined with new social networking technologies, campaigns now have expanded opportunities to use supporters and their social networks as the vehicles of strategic communication. Building from the innovations of Howard Dean's primary run (Kreiss, 2009, 2010), campaigns deploy event planning tools, build Facebook applications, and create online calling tools that help make social networks both visible in the form of data and productive for fund-raising, mobilization, and voter identification efforts.

Despite two decades of research into the new media practices of campaigns, the workings of the data and analysis "backend" (Hindman, 2007) of electoral politics and the industry that supports it are still largely hidden from view, particularly given the press focus on social media. That said, scholars have made some important inquires into these practices. As Gandy (2000) noted at the beginning of the

decade, political organizations were segmenting and directing narrowly targeted persuasive communication to the electorate based on detailed data profiles. As a consequence, scholars have suggested not only how the range of democratic debate will become more narrow (Bennett, 2008; Bennett & Manheim, 2006) but that certain people and ideas are likely to be excluded from political communication (Howard, Carr, & Milstein, 2005). These data analyses and communication practices mirror changes in other domains of social activity. Turrow (2006), for instance, powerfully documents marketing to niche audiences based on consumer data in commercial contexts (see also Andrejevic, 2003; Elmer, 2003).

The uptake of digital media in campaigning is primarily about voter identification, persuasion, and mobilization in the service of very narrow electoral ends. This distinguishes these practices from the far more troubling consequences of surveillance by a repressive state (Baruh, 2007) and the collection of data by governing bodies, which the Supreme Court has found threatens associational freedom (Hunter, 2002). and from the selling of commercial wares. Yet, this does not mean that there are not troubling downsides to the way that modern campaigns are conducted. While strategic and tailored political communication can increase turnout and electoral participation, particularly when compared with broadcast politics (Kang, 2004), as a body of literature discussed throughout this article suggests, there are ample reasons for scholars to be concerned about the data and analysis practices of campaigns. For example, there is a pervasive lack of transparency around what data political campaigns and parties collect, purchase, and generate, as well as how it used in practice. Ironically, this opacity regarding the practices of campaigning often comes from the very candidates that are at the forefront of calls for open government. Citizens can legally access their FBI files, for instance, but have no legal right to see what the Obama campaign knew about them. At the same time, there are widespread possibilities for data breaches, a concern that grows with each election cycle, given the increasingly sensitive information stored on citizens and the ways it is traded on an enormous, open and unregulated market.

The general lack of regulation and transparency around this market of personal information also raises concerns regarding who has access to this data and how it is used in the context of democratic debate. Data moves fluidly across commercial and political contexts. Quasi-state actors, such as parties and campaigns, purchase data from commercial providers, just as corporate interests build their databases from a host of publicly available information. All of which means an increasing constellation of interests have the ability to track the social and intellectual associations of citizens. And, while campaigns have very transactional ends, the lack of transparency means that it is often hard to know where data is housed, who has access to it, and where it migrates once a candidate is elected. Finally, data can be used to truncate public debate and narrow political representation. Scholars have long feared a "democratic deficit" when campaigns can tailor their communication to particular voters, while ignoring sizable portions of the electorate. This means that candidates vie to represent segmented and small slices of the public. Furthermore, the ability to compete in modern, data-driven campaigns comes at enormous fiscal cost, meaning that only the well-funded candidates can effectively compete.

The sophisticated data gathering and analysis techniques of contemporary campaigns both catalyze political participation and undermine important democratic norms. Given that much of the scholarly account of the 2008 elections, particularly that of Obama's campaign, has centered on digital technologies as a boon to electoral participation, this article seeks to spark a critical discussion of the data

practices that underpin contemporary campaigning. This article draws from an older, established body of literature to provide an initial inquiry into a set of questions rarely considered in much scholarship on campaigning in the Facebook and Web 2.0 era, where networked technologies are celebrated for facilitating new modes of social relationships. We ask relatively simple questions: What political data is collected about citizens? What is it used for? Who owns it? How is it stored, and with what consequences for democratic practice?

In doing so, we hope to provide the initial groundwork for a larger discussion about the centrality of data gathering and analysis in contemporary campaigns. In keeping with our aim to begin a scholarly conversation, we do not provide the definitive theory of political data or a clear set of policy recommendations. Simply raising the issue of political privacy is overdue and timely, particularly given the pervasive celebration of the impact of digital media, especially Web 2.0 technologies, on democratic life. We begin by providing a brief history of political data practices, focusing on innovations that took shape over the last decade. We then detail the regulatory context in the United States around political data. Finally, we issue a call for critical scholarly work on the generation, use, and storage of political data, given the potential implications of data for democratic practice. Evidence for this article is drawn from a comprehensive review of primary and secondary sources, including journalistic reports and academic texts on new media campaigns, as well as the authors' research work in other related contexts.

A Brief History of Political Data in the United States

A series of institutional changes beginning in the 1960s are at the origins of the modern data practices of campaigns. One change was the move to direct primaries in the wake of the 1968 Chicago Democratic Convention, given the crisis in democratic legitimacy (Polsby, 1983). A second shift was the rise of political consultants as a distinct class of professionals independent of parties. This professional field helped develop new polling techniques, more sophisticated broadcast advertising and ways of working with television news, and new modes of fundraising, given that campaign finance reform in the 1970s capped the amount of money citizens and corporations could give to campaigns. Together, these changes helped produce a new electoral environment with independent candidates, autonomous campaign organizations, and a thriving class of political consultants operating independently of, yet servicing, political parties. Meanwhile, powerful advances in computing during the 1970s made data management skills and raw datasets stored on mainframes newly bankable assets for consultants and parties (Howard, 2006; Sabato, 1981; Whitman & Perkins, 2003). During this time, political professionals began thinking more seriously about their knowledge of the citizenry and the ways that it could be leveraged in the service of electoral ends.

The important turning point towards more portable, national, and comprehensive data, however, came with "prospect direct mail," developed in the 1980s. Driven by Republican innovations, direct mail enabled consultants and party operatives to micro-target political communication to discrete individuals, based on sophisticated knowledge of the electorate. Consultants used small desktop machines to gather public and commercial information, such as magazine subscription lists, and to identify prospective donors and voters. By the 1990s, direct mail was a potent fund-raising and communications tool, and political marketing of this sort became a professional undertaking distinct from political communications and

polling work. With elected leaders perpetually on the campaign trail, a constant flow of data about fundraising prospects and voter intent was crucial for dynamic image management (Lees-Marshment, 2002; Scammell, 1996). Republican Party databases of the electorate during this time became much more national in scope. The Democrats, however, lacked many of these systems. Throughout the 1990s and well into the 2000s, the Democratic Party's collection, storage, and analysis of data on the electorate was largely a haphazard affair, coordinated mostly by state parties or the stewards of local precincts.

During this time, the Internet emerged as a new technology for political campaigning. Its use by campaigns was halting throughout much of the 1990s, as communications staffers and consultants largely appropriated it as a broadcast medium to push information out to undecided voters (Foot & Schneider, 2006). By the 2000 presidential election, however, it was increasingly clear to campaigns that many visitors to campaign Web sites were already sympathetic to, if not supporting, the candidate (Bimber & Davis, 2003). In response, candidates began using the Internet as a tool for mobilization, providing supporters with opportunities to contribute to finance and communication efforts.

It was during this time that the twin features of Internet campaigning emerged that defined the online efforts of the 2008 presidential cycle. Both features were pioneered by Democrats. The first came with respect to campaigns creating expanded opportunities for supporters to participate in the electoral process as active fund-raisers, canvassers, and communication agents. Al Gore's online campaign, for instance, enabled supporters to create and circulate their own personalized Web pages to their friends and family. Howard Dean's campaign explicitly made supporter participation in fundraising and electioneering a central part of the campaign's strategy (Kreiss, 2009). The campaign used commercial sites such as MeetUp and developed their own event planning tools to help supporters host fund-raisers and other events. Dean's Internet staffers also developed and deployed tools that enabled supporters to create their own affinity group sites for Dean, as well as an application that enabled supporters to create personalized fund-raising goals and reach out to their friends and family for contributions.

The second feature was the development of the data infrastructure that helped campaigns coordinate this supporter involvement and reach the electorate. In this sense, the Internet has largely "amplified" (Agre, 2002) institutionalized campaign practices extending back over 40 years. After the 2004 election cycle, the veterans of presidential bids, especially those of Dean and Wesley Clark, became political professionals specializing in digital media (Kreiss, 2010). With lessons learned during 2004 in hand, they set out to develop better systems to manage interactions and communicate with supporters. After 2004, the firms founded by these professionals developed dedicated "customer relations management" (CRM) systems for political purposes that enabled campaigns to keep better records on supporter involvement, including tracking donations and volunteer actions. These systems also provided campaigns with the capacity to send targeted e-mails to supporters based on any number of individual characteristics, from their ZIP codes and demographics to their level of involvement with the campaign.

These online systems powerfully came together during the 2008 campaign cycle, especially on the Obama campaign. Whereas John Kerry, the Democratic nominee in 2004, lacked the basic capacity to e-mail all of his supporters, the Obama campaign built an e-mail list of 13 million members that was continually segmented, depending on the action the campaign wanted supporters to perform. This e-mail

list, in turn, was synched with both the log-in data on the campaign's main BarackObama.com site and with the 2 million profiles that supporters set up on the campaign's MyBarackObama.com platform. Supporters used MyBO to set up personal fund-raising pages and to plan visibility and other events for Obama. The campaign then tracked this supporter involvement, telemarketing and e-mailing standout volunteers, and first-time donors to increase their involvement.

The affordances of the Internet offered new opportunities for the Obama campaign to know its supporters and craft effective appeals. The analytics team within the new media division of the campaign—many staffers of which were engineers and analysts with extensive commercial experience—continually ran statistical tests on use of the site and on communications with voters to determine optimal content and design. For example, "cookies" installed on Web browsers helped the Obama campaign deliver content on BarackObama.com tailored to each individual. Analytics lay behind everything, from the shape, color, and content of the buttons asking individuals to donate to the content that first-time, repeat, and active users encountered on the site. The campaign continually tested the design and content of its e-mails by sending different versions to small samples of supporters. They then tracked the "click-through" rates of these e-mails, enabling the campaign to see which message resulted in more donations.

All of these online data practices, in turn, were backed up by extensive databases on the electorate, which was behind Obama's extraordinarily successful field campaign. Democrats finally caught up to the Republicans in voter data during the 2008 cycle. The Republican Party had developed a national voter database in the mid-1990s called Voter Vault that integrated various direct mail databases. It went online for the 2002 midterm elections, providing a common interface for Republican field organizers throughout the country to access and update the database based on field efforts. By 2004, it was reputed to have information on 168 million citizens.

Republicans were the widely acknowledged leader in voter identification and field turnout efforts throughout much of the 1990s and onward until Dean's tenure as the chair of the Democratic Party began in 2005. After evaluating a number of firms, Dean hired Voter Activation Network to develop VoteBuilder, an advanced interface launched in 2007 for field organizers to use in canvassing, event organizing, and volunteer management. The DNC subsequently provided this technology to state parties, while retaining their data. This arrangement ensured that one central national voter database, stewarded by the Party, was being continuously updated with the on-the-ground data gained across election cycles. During the 2008 presidential cycle, candidates supplemented this Party-supplied data with other voter databases of public, commercial, and generated data. The Obama, Clinton, and Edwards campaigns, for instance, utilized the services of Catalist, a private firm founded by prominent Democratic political consultants and backed by George Soros (Edsall, 2006). The Catalist database is reputed to encompass 450 points of data compiled from public and commercial sources on more than 250 million people in the United States. To a lesser extent, these campaigns used the services of Aristotle, the most well-established and largest bipartisan provider of data, which is also often the subject of investigative journalists (Verini, 2007; Zetter, 2003).

Campaigns use these vast troves of data to generate profiles of their likely supporters and then build their voter contact operations around this modeling. These voter databases capture an extraordinary

range of public, commercial, and generated information. Public information includes voter registration, party identification, turnout records, vehicle registration, and ZIP code-based census data. Commercial data entails everything from magazine subscriptions to credit card and grocery store "club-card" purchases. Firms such as Strategic Telemetry poll groups of likely voters and then work backwards to figure out the demographic and lifestyle characteristics of likely supporters and undecided voters. For example, the Obama campaign utilized Strategic Telemetry to develop complicated correlational data models that assigned each member of the electorate an overall score of 1–100 that indicated the voter's predicted preference for Obama. Based on these scores, the firm generated a universe of targets that campaign volunteers then followed up with door-to-door or via phone. The data generated from these contacts were then used to test and update the models.

These data and analysis practices in turn shaped everything from the campaign's marketing and communications efforts to candidate visits to swing states. They were also behind a big innovation of the Obama campaign: the merging of field efforts and new media. The campaign took the first, fraught steps toward integrating the national Democratic Party voter database with the online tools created for the campaign by the consulting firm Blue State Digital. In doing so, the campaign was better able to direct its distributed network of volunteers toward field efforts. For example, during the general election, the campaign opened up its voter files to volunteer canvassers across the country through calling tools such as "Neighbor to Neighbor," which the Democratic Party had commissioned Blue State Digital to build for the nominee. This online tool enabled volunteers in uncontested states to call voters the campaign had targeted.

Campaign volunteers using "Neighbor to Neighbor" could access voter profiles through Web-based systems. Volunteers typically could see the address, age, gender, size of household, and primary language of a prospective voter. They could also see the results of previous canvasses. The volunteer was instructed to speak only to the targeted voter in question and was provided with an online script soliciting the voter's preferences. The volunteer then recorded the outcome of the call instantly on MyBO, according to predetermined response categories. The door-to-door canvas of voters to ascertain their likely poll behavior was thus supplemented with an online tool that enabled supporters to collect data on the national electorate from the convenience and comfort of their own homes. This not only increased the number of potential volunteers available for Obama but also provided data on voter contacts to the campaign nearly instantaneously. Field organizers, for instance, could dispatch volunteers based on the results of these calls.

The integration of online, Web-based tools, and voter databases is expected to become more sophisticated in the years to come. A number of early forays in this direction were pioneered by the Obama campaign and other advocacy organizations. For example, the campaign gathered data from individuals who "friended" the campaign on commercial social networking sites such as Facebook. This data was then synched with voter databases and used to organize youth volunteers in early primary states. The campaign's Facebook Connect application, meanwhile, enabled supporters to urge their friends in early primary states to vote. This merging of voter registration files with social network data in such a way that leverages the data about links between people, not just variables about people (Ambinder, 2008), is also taking root outside of electoral campaigns. MoveOn.org, a progressive advocacy

organization, built an application called "Vote Poke" in partnership with Catalist that allowed citizens to find out if friends in their online social networks were registered to vote and then reach out to them if they had not (Stoller, 2008).

Finally, campaigns took initial forays into online advertising during this electoral cycle. Spending on traditional cable and broadcast ads still dwarfed the online advertising budget, but the Obama campaign used new media to deliver messages to targeted demographic and geographic groups. Through online advertising networks, his campaign also targeted advertising based on browsing behavior. This targeting, in turn, was integrated across sites. A profile of the online advertising practices of the 2008 presidential campaigns describes it this way:

> For example, when people visit the volunteer section of the Obama Web site but click away without signing up, the campaign puts a cookie on their Web browser. Then, as surfers move around the Web, the campaign looks for opportunities to bring them back. If they go to a parenting blog, Obama can deliver an ad about education policy. If they read a story on a tech news site, the campaign can serve up something about technology policy. (Green, 2008)

These data practices can spur turnout in elections, and they amplify citizens' ability to participate in institutionalized electoral activities. Yet, as the next section reveals, the data practices of campaigns are generally unregulated and proceed almost entirely without public attention.

The U.S. Regulatory Context

Despite the massive amount of political information generated by political campaigns and parties in the United States, these entities face almost no regulation with respect to the collection, use, storage, and dissemination of data on citizens. In contrast to the greater restrictions on the use of data for expressly commercial purposes (Turow, 2006), provided they remain non-state actors, candidates and parties enjoy broad latitude with respect to their data practices, generally on the grounds that their ability to speak to citizens is protected by the First Amendment. This section provides an overview of the patchwork of regulations that govern political data in the United States, then makes the case that scholars and citizens should be concerned about this lack of oversight, given advances in gathering, storing, analyzing, and acting upon data.

There have long been distinctions between the use of data by official state agencies and representatives in the course of governance versus quasi-state agencies, such as parties and campaigns for the purposes of electioneering. There is, for instance, an extensive body of privacy law that governs the executive branch. The 1974 Privacy Act and the Electronic Communications Privacy Act, for example, restrict the types of data that federal agencies can collect while mandating how it is to be used, stored, and disclosed.[1]

[1] That said, there are a number of holes in this legislation. In the last few years, scholars have expressed concern over the practice of public agencies contracting out to private databases to circumvent these Acts.

Outside of the executive branch, there are rules that govern other state actors. The Franking commissions of the House and Senate regulate congressional representatives by distinguishing between and setting separate rules for official governance and electioneering activities. For example, the rules issued by these commissions place restrictions on the types of content permitted in publicly funded messages to constituents, on the databases that can be used for e-mail communications, and on the information that can be posted on official Web sites.

The fear of potential abuse of data and of office is the logic undergirding these rules for state bodies and representatives concerning the use of data on the electorate. Congress passed the Privacy Act in the wake of Watergate, which revealed the pressing need to protect the citizenry from governmental surveillance. Franking rules, which date from the earliest days of the republic, balance the need for representatives to communicate with their constituents on matters of state affairs with the value of competitive elections. In restricting some of the ways that Congressional members may communicate with constituents, Franking helps ensure that incumbents do not have an insurmountable electoral advantage.

While there are defined data and communication rules for official state actors, these do not extend to the actions of candidates. The logic here is that candidates for higher office, as well as political parties, have a general right of free political expression. Along these lines, lawmakers and courts have generally protected candidates and parties' use of databases to know who to speak to and what to say. Indeed, public information collected by state agencies serves as the foundation of the voter databases used by campaigns and parties. As noted, these databases include information on voter identification and party affiliation, voting history, division of motor vehicle records, and campaign donations.

While the availability of much of this information is determined by each state in lieu of a federal standard for accessing and purchasing public data, there is a general lack of regulation and broad protection for its political use. The California Voter Foundation, for instance, produced a comprehensive report detailing the public information collected by each state and subsequently made available to third parties (Alexander & Mills, 2004). They found that most states do not restrict the use of this data by even commercial entities.[2]

In the meantime, expressly political organizations, such as campaigns, advocacy organizations, and parties enjoy broad exemptions from any restrictions under these state laws—as do the commercial companies that furnish their data. For example, in 2003 the California Secretary of State created a Voter Privacy Task Force, the only one of its kind, charged with looking into the risks of potential misuse of this data. California more generally has been at the forefront of protecting data, both public and commercial. It even implemented the nation's first mandatory data breach law. Nonetheless, even in California companies such as Aristotle, the Washington DC-based political data company, occupy a gray area in state law. California courts have held that even though firms such as Aristotle are commercial, the ultimate end

[2] Even when there are tough rules in place governing public data, there are often only minimal restrictions on commercial entities, and these laws are generally left unenforced.

users of the data are political campaigns, which enjoy broad exemptions from regulatory efforts (K. Alexander, personal communication, March 4, 2009).

As is clear, campaign organizations, parties, consultants, and the firms that provide services to them enjoy broad latitude with respect to their use of political data. With exemptions from even the minimal laws governing the use of public and commercial data, these quasi-state entities have a free hand to collect data on and target the electorate. At the same time, the federal government is inadvertently making the job of firms, such as Aristotle, even easier. The federal "Help America Vote Act of 2002," designed to help prevent the election chaos in 2000 from happening again, requires states to create state-wide voter databases for administration purposes. While many states have been slow to implement the law, eventually it will result in the creation of one central site in each state to which parties and commercial data providers can go to receive public information on voters.

While maintaining voter lists and targeting the electorate are practices that date to the earliest days of party politics, the increasing sophistication of database technology is raising significant new concerns for democratic practice. Few could have foreseen how the storage capacities of digital technologies, coupled with the ability of new media platforms to generate real-time social and behavioral information, have exponentially multiplied the data collected on citizens.

As detailed in the next section, the lack of transparency with respect to data practices, the possibility for data breaches, the potential for deeper intrusions into the privacy of voters, and the thinning of democratic deliberation and representation all suggest the need to rethink the regulatory context around political data.

Campaign Data and Democratic Citizens

Contemporary political campaigns are built around data. And, in many respects, citizens endorse its use in conjunction with the electoral process. Many supporters not only voluntarily give data to political campaigns, but also, through canvassing, quite consciously generate data on their fellow citizens to help the candidates they support get elected. Consultants and scholars alike suggest that knowing and speaking to the issues that most concern voters increases political engagement and voter turnout. Given this, why should citizens and scholars care about the profusion of political data?

The Lack of Transparency in Political Data

As previously detailed, changes in political technologies greatly amplify the possibilities for collecting, storing, and acting on citizen data in electoral politics. Yet, few citizens or scholars know about the information practices of contemporary campaigns because this work is generally hidden from view. With wide latitude to collect, process, store, and share data, these entities and digital political consultants pursue their informational work without much in the way of oversight or accountability.

This basic lack of transparency exists at a time when, as our research suggests, political data is among the most valued commodities in the market for consultant services. Candidates hire consultants

who have access to data or contract out to firms, such as Catalist and Aristotle, for information on the electorate. Meanwhile, through new media tools that make online behavior and networks visible and that support virtual canvassing, campaigns are generating and capturing more sophisticated forms of data on supporters and voters.

Given this, with the proliferation of data has come a widening gulf between what is collected and what citizens know is being catalogued and stored. Political data receives scant public attention, in part because elected officials have little interest in holding hearings on practices that benefit them, and campaign consultants seldom speak to the press. Journalistic reports are far and few between and tend to focus on high-profile incidents, such as a Web site getting hacked or a breach in a database (Stephey, 2008; Zetter, 2003), rather than on the routine functioning of data systems. Many kinds of political organizations, including political parties, lobbyists, industry associations, non-profit groups, and political action committees, reveal little about their use of political data.

Even as there is little in the way of transparency with respect to data practices, much suggests that very few people have the expectation that campaigns will collect records of their grocery store purchases to profile them ideologically. Yet, the linkages between our credit card purchases, demographic profiles, and voter registration records are increasing as is the use of this information by actors we do not know for purposes that we cannot control (Clark, 1994; Howard, 2003). For example, credit card records about contraceptive, gun, and magazine purchases are valuable data for lobby groups working on abortion, firearm control, and other political issues (Howard, 2006; Carr, & Milstein, 2005; Sides & Karch, 2008). Even if citizens know about the information that has been collected on them—exceedingly rare, given the lack of transparency—there is no clear way for them to "opt out" of political datasets or manage the circulation and use of their private data.

Data Security

To an alarming extent, an increasing number of political databases are lost, accidentally exposed online, or hacked.

A decade of journalistic coverage suggests that breaches in political data occur (*Businessweek*, 2006; Grossman, Novak, & Roston, 2004; Sanders, 2001; Tynan, 2004; Zetter, 2003). Between 1990 and 2006, there were some 600 high-profile incidents of compromised data records in commercial and political contexts internationally, most of which occurred in the United States (Erickson & Howard, 2007). These breaches occurred due to such factors as administrative errors, accidental posting to Web sites, insider abuse or theft, missing or stolen hardware, and hacker attacks.

While most of these incidents involved private firms, since 2000, a number have involved the information infrastructure of political parties. In the United States, political data has been sold to shore up the finances of failing firms (Sanders, 2001), campaign Web sites have been hacked (Newsweek, 2008; Glendinning, 2008), questionable data practices have been exposed (Tynan, 2004), and American voter data has turned up in surprising places overseas (Grossman, Novak, & Roston 2004).

While breaches in data are a concern in any context, political data is particularly sensitive and a unique order of security threat and violation of voter trust. The data held by campaigns and parties is explicitly used for political purposes, and thus reveal levels of detail about policy preferences and ideological perspectives that are less relevant to the advertisers of consumer goods. Moreover, given that most people do not realize the degree to which the world of campaigns, elections, and political discourse is data driven, a breach is not just the loss of personal privacy but also an occasion for diminished faith in publicly supervised political processes.

Political Privacy and Freedom of Association

Political privacy, and with it a freedom of association, is uniquely at risk with the increasing sophistication and use of voter databases. As Hunter (2002) argues, a tradition of Supreme Court jurisprudence developed over the last half century protects anonymous speech and associational privacy. In doing so, the High Court articulated a general value of "political privacy" that protects "a space apart, free from the gaze and influence of other citizens, political parties, the government, and business interests, where individuals can develop and experiment with their political views."

In each of these domains, privacy helps ensure robust public debate by providing the opportunity for citizens to form their own viewpoints, craft arguments, and develop political identities free from state surveillance and public pressures to conform to social norms. The logic is that if citizens are constantly being watched, they will be less likely to state their opinions or freely choose with whom to associate. At the same time, privacy also provides a secure environment for social movements and activists to prepare for engagement with the state. A number of striking cases demonstrate the consequences of the loss of political privacy and the value of remaining anonymous. The Red Scare in the United States during the 1950s, for example, laid bare the hazards of providing the state with associational data.

While candidates and parties do not engage in these types of state-sponsored repression, given the lack of transparency around the data practices of campaigns, it is not always clear how data travels across legal contexts and organizational boundaries. Theoretically, the rules that separate constituent service data from voter profiling data are designed to prevent the state from building an organizational memory of citizen associations (C. Hoofnagle, personal communication, February 20, 2009). Yet, even while there are restrictions on how elected representatives can communicate as agents of the state, the massive amounts of data collected for electoral and advocacy campaigns does not simply disappear when a candidate wins office.

Perpetual campaign organizations and political parties hold the political information generated during electoral campaigns and make use of it to achieve policy aims, given that they are not considered arms of the state under the law. This is illustrated by the way that President Obama's extensive database of voter information, built up over the course of his primary and general election campaigns, changed hands soon after his election. This data is estimated to include the widely reported 13 million e-mail addresses, 7 million cell phone numbers, and data on 2.4 million Facebook and 2 million MyBO users, alongside the 223 million pieces of information that citizens volunteered about themselves to canvassers during the course of the election.

While the White House cannot directly use this information for political purposes (Rutenberg & Nagourney, 2009), Obama's campaign team built an entirely new organization within the Democratic Party to house this data. Organizing for America is an independent organization that coordinates its activities with the communications team of the administration in using the e-mail list to speak directly to the campaign's supporters on policy matters (ibid.). To date, Organizing for America has used the list to mobilize Obama's campaign supporters in conjunction with pressing Congress to pass Obama's 2009 stimulus plan and 2010 health care reform. This suggests the ways in which political data exists in perpetuity and can migrate to other contexts—often with little transparency to the movement of these databases and without the consent of those who are profiled. This is particularly troubling with respect to the ways that the affordances of digital media give strategic actors the ability to track social and intellectual associations.

Data and the Democratic Deficit

Outside of concerns over political privacy, in the hands of political consultants, data are a potent tool through which to segment the electorate. Public debate may be becoming more narrow as campaigns use detailed data on citizens to craft narrow political appeals to receptive audiences, while ignoring those either ideologically opposed to their message or unengaged in the political process. Public discourse suffers whenever political actors segment the population and narrowcast information, because such fragmentation allows them to decide which citizens are worth engaging. This is occurring to such an extent that Bennett & Manheim (2006) argue that digital media are facilitating a "one-step flow" of political communication. In their assessment, elite actors are "increasingly less likely to 'lead' because they are more likely to reinforce latent opinions than to reframe them" (p. 213).

This is clear in the ways that candidates and parties are most interested in identifying the people who will support them on Election Day. They will spend resources to engage citizens who might lean their way on a given issue and focus on creating narrow appeals that will have a maximal effect, given the political preferences they have identified through analytic modeling.

The corollary to this phenomenon is that candidates are free to ignore entire slices of the electorate. Campaign consultants have demonstrated that they will not spend significant resources engaging citizens who tend not to vote—often the urban poor and ethnic minorities (Howard, 2006). Why attempt to educate and energize people who have no history of voting and may seem politically marginal? In essence, while targeted communication may increase turnout among certain segments of the population, data also allows political actors to identify which citizens should be left entirely out of the conversation.

On another level, contemporary data practices narrow political representation. One way that this happens is through affecting electoral competition, as the best voter data files and analytic services are often only available to wealthy candidates and parties. While networked technologies have made it easier to collect vast amounts of information about people, integrating, storing, and analyzing that data remains expensive. This means that data (and knowledge) itself is a form of campaign asset that is most often

available to incumbents and large parties. Less well-funded and insurgent candidates challenging their party's standard-bearers often do not have access to the same intelligence on the electorate. Meanwhile, the power of well-financed lobbyists, advocacy organizations, and incumbents is greatly amplified by large relational datasets.

Another consequence of increasingly sophisticated data on the electorate is a problem that has drawn increasing scholarly attention: gerrymandering. Geographic information systems afford the layering of electorate data onto maps to allow elected officials to redraw congressional districts. A classic dilemma in democratic theory concerns the role of elected representatives: Should she be simply a mouthpiece for constituents, or does her election empower her to speak of her own views? With expansive political data on individual citizens, this dilemma is solved by enabling elected officials to actually embody the represented, creating districts that faithfully resemble officials' own views, as they can meticulously group voters in geographic space according to ideology, affiliation, and demographics.

Conclusion

There is little to suggest that data transparency and regulation will come about on its own.

The interests of our political leadership and large data mining firms are closely aligned against the state regulation that would secure transparency in data practices. Indeed, these actors pay much lip service to information security to avoid discussion of the much more troubling undermining of political privacy.

At the same time, the scholarly literature has celebrated the seemingly new forms of collective action and uncoordinated, bottom-up types of political participation made possible by digital media. Yet, the technical infrastructure that underpins this online action, and the professionals who work behind the scenes with data to identify voters and coordinate collective action, remain largely unanalyzed. Interfaces, databases, and consultancies form the largely invisible backend to Web 2.0 politics. The opaqueness of this industry has meant that there is little discussion of the consequences of these data practices for democratic life. Yet, this discussion could not be more timely. The lack of transparency and security in political data raises significant concerns for citizens. Political privacy and freedom of association have the potential to be undermined as databases grow more expansive, enduring, and centralized. We also have to seriously worry about a democratic deficit where only the wealthiest candidates can seriously compete, and where vast swaths of the citizenry remain unengaged in the political process.

While this article is intended to start the scholarly conversation on the generally invisible back end of Internet campaigning, based on a comparative work on political privacy (Howard & Kreiss, 2009), we have identified some first steps for policy makers and scholars. For example, as a start, data practices and storehouses of information should be made transparent and accessible. Citizens should have the means to find out what data is collected and stored about them and have the ability to opt out of political databases. Parties and candidates should be required to develop privacy statements that cover the full range of personal data they manage and clearly state how political information is used, who it is shared with, what it entails, and how it is gathered. Political data should be subject to the same breach reporting

requirements that many states currently mandate of commercial providers. Together, transparency and security would help to ensure that databases of citizens are more accurate, secure, private, and reliable, while raising public awareness of the flow of information about them.

Meanwhile, much more scholarly work is needed into the democratic effects of the data-driven polity. As this article suggests, advances in networked technologies have not only brought about opportunities for political participation in some domains but also the explosion of political data on individual citizens. As this data is carried across electoral cycles and continually updated, our profiles evolve alongside our political beliefs. In the process, we are marketed to and monitored to an unprecedented extent. Our social and intellectual associations are tracked for the political marketers and elected officials who can tailor their appeals to us, based on some of our most intimate characteristics. Accordingly, we lose opportunities for inclusive and substantive democratic debate.

Acknowledgments

This research was supported by the World Information Access Project (www.wiareport.org) and the Office of the Privacy Commissioner of Canada. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Office of the Privacy Commissioner of Canada. For their assistance and support of this project, the authors are grateful to Kim Alexander, Wendy Bolton, Greg Elmer, Chris Hoofnagle, Colin McKay and Gina Neff. In addition, we are grateful to the helpful staff of the Office of the Privacy Commissioner of Canada, the Office of the Privacy Commissioner of Australia, and the Information Commissioner's Office of the United Kingdom.

◇◇◇

Daniel Kreiss is a Fellow, Information Society Project and Postdoctoral Associate at Yale Law School. He is currently working on a book, forthcoming from Oxford University Press, that traces the history of digital media and political campaigning from Howard Dean to Barack Obama.

Philip N. Howard is an associate professor in the Department of Communication at the University of Washington. He is the author of New Media Campaigns and the Managed Citizen *(New York: Cambridge University Press, 2006) and* The Digital Origins of Dictatorship and Democracy: Information Technology and Political Islam *(New York: Oxford University Press, 2010).*

References

Agre, P.E. (2002). Real-time politics: The Internet and the political process. *The Information Society, 18* (5), 311–331.

Alexander, K., & Mills, S. (2004). Voter privacy in the digital age. California Voter Foundation. Available at http://www.calvoter.org

Ambinder, M. (2008, November 11). How to tell your VoteBuilders from your MyBOs, your Catalists from your VANs. *The Atlantic.* Available at http://marcambinder.theatlantic.com

Andrejevic, M. (2003). Monitored mobility in the era of mass customization. *Space & Culture. 6* (2), 132–150.

Baruh, L. (2007). Read at your own risk: Shrinkage of privacy and interactive media. *New Media Society, 9*(2), 187–211.

Bennett, W. L. (2008). Engineering consent: the persistence of a problematic communication regime. In P. Nardulli (Ed.), *Domestic Perspectives on Contemporary Democracy* (pp. 131–154). Chicago: University of Illinois Press.

Bennett, W.L., & Manheim, J. B. (2006). The one-step flow of communication. *The Annals of the American Academy of Political and Social Science, 608*(1), 213–232.

Bimber, B., & Davis, R. (2003). Campaigning online: The internet in U.S. elections. New York: Oxford University Press.

BusinessWeek. (2006, May 29). The snooping goes beyond phone calls. Retrieved September 27, 2010, from http://www.businessweek.com/magazine/content/06_22/b3986068.htm

Clark, R. (1994). The digital persona and its application to data surveillance. *The Information Society. 10* (2), 77–92.

Edsall, T. B. (2006, March 8). Democrats' data mining stirs an intraparty battle. *The Washington Post.* Retrieved September 28, 2010 from http://www.washingtonpost.com/wp-dyn/content/article/2006/03/07/AR2006030701860.html

Elmer, G. (2003). A diagram of panoptic surveillance. *New Media and Society, 5*(2), 231–247.

Erickson, K., & Howard, P.N. (2007). A case of mistaken identity? News accounts of hacker and organizational responsibility for compromised digital records. *Journal of Computer Mediated Communication, 12*(4), 1229–1247.

Foot, K., & Schneider, S.M. (2006). *Web Campaigning*. Cambridge, Mass: MIT Press.

Gandy, O. (2000). Exploring identity and identification in cyberspace. *Notre Dame Journal of Law, Ethics, & Public Policy*, *14*(2), 1085–1111.

Glendinning, L. (2008, November 7). Obama, McCain computers 'hacked' during election campaign. *The Guardian.* Retrieved September 28, 2010 from http://www.guardian.co.uk/global/2008/nov/07/obama-white-house-usa

Green, H. (2008, August 28). The candidates are monitoring your mouse. *Business Week*. Retrieved September 27, 2010, from http://www.businessweek.com/magazine/content/08_36/b4098022877194.htm

Grossman, L., Novak, V., & Roston, E. (2004, October 18). What your party knows about you. *TIME*. Retrieved September 28, 2010, from http://www.time.com/time/magazine/article/0,9171,995394,00.html

Howard, P. N. (2003). Digitizing the social contract: Producing American political culture in the age of new media. *The Communication Review, 6*(3), 213–245.

Howard, Philip N. (2006). *New Media Campaigns and the Managed Citizen*. New York: Cambridge University Press.

Howard, P. N., Carr, J., & Milstein, T. J. (2005). Digital technology and the market for political surveillance. *Surveillance and Society, 3*(1), 59–73.

Howard P. N., & Kreiss, D. Political parties & voter privacy: Australia, Canada, the United Kingdom, and United States in comparative perspective. World Information Access Project Working, Paper #2009.1. Seattle: University of Washington.

Hindman, M. (2007). 'Open-source Politics' reconsidered: Emerging patterns in online political participation. In V. Mayer-Schönberger & D. Lazer (Eds.) *Governance and Information Technology* (pp. 183–211). Cambridge, MA: The MIT Press.

Hunter, C.D. (2002). Political privacy and online politics: How e-campaigning threatens voter privacy. *First Monday 7* (2) Retrieved September 28, 2010, from http://ojphi.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/930/852

Jenkins, H. (2008, February 18). Obama and the "We" generation. *Confessions of an Aca-Fan.* Retrieved September 28, 2010, from http://www.henryjenkins.org/2008/02/obama_and_the_we_generation.html

Kreiss, D. (2009). Developing the 'good citizen': Digital artifacts, peer networks, and formal organization during the 2003–2004 Howard Dean Campaign. *The Journal of Information Technology and Politics*, *6*(3/4), 281–297.

Kreiss, D. (2010). Taking our country back?: Political consultants and the crafting of networked politics from Howard Dean to Barack Obama. Unpublished doctoral dissertation, Stanford University, California.

Kang, M. S. (2005). From broadcasting to narrowcasting: The emerging challenge for campaign finance law. *The George Washington Law Review*, *73*, 1701–1730.

Lees-Marshment, J. (2002). The marriage of politics and marketing. *Political Studies. 49*(4), 692–713.

Miller, T. (2008). My green crush. *Journal of Visual Culture*, *8*, 154–158.

Newsweek. (2008, November 5). Hackers and spending sprees. Retrieved September 28, 2010, from http://www.newsweek.com/id/167581

Plouffe, D. (2009). *The audacity to win: The inside story and lessons of Barack Obama's historic victory.* New York: Viking.

Polsby, N. (1983). *Consequences of party reform*. New York: Oxford University Press.

Rutenberg, J., & Nagourney, A. (2009, January 25). Melding Obama's Web to a Youtube presidency. *The New York Times*. Retrieved September 28, 2010, from http://www.nytimes.com/2009/01/26/us/politics/26grassroots.html?hp

Sabato, L. J. (1981). *The rise of political consultants: New ways of winning elections*. New York: Basic Books, Inc.

Sanders, E. (2001, March 21). Planned sale of Voter.com's data raises privacy concerns *The Los Angeles Times*. Retrieved September 28, 2010, from http://articles.latimes.com/2001/mar/08/business/fi-34937

Scammell, M. (1996). The odd couple: Marketing and magic. *European Journal of Marketing*, *10*, 114–126.

Scholz, T. (2008). Market ideology and the myths of 'Web 2.0.' *First Monday*, *13*(3).

Sides, J., & Karch, A. (2008). Messages that mobilize? Issue publics and the content of campaign organizing. *The Journal of Politics. 70*(2), 466–476.

Stephey, M. J. (2008). September 17). Sarah Palin's e-mail hacked. TIME. Retrieved September 30, 2010, from http://www.time.com/time/politics/article/0,8599,1842097,00.html

Stoller, M. (2008, January 24). Dems get new tools, new talent. *The Nation*. Retrieved September 28, 2010, from http://www.thenation.com/doc/20080211/stoller

Tynan, D. (2004, September 24). GOP Voter vault shipped overseas. *PC World*. Retrieved September 28, 2010, from http://www.pcworld.com/article/117930/gop_voter_vault_shipped_overseas.html

Turow, J. (2006). *Niche envy: Marketing discrimination in the digital age*. Cambridge, Mass: MIT Press.

Verini, J. (2007, December 13). Big Brother, Inc. *Vanity Fair*. Retrieved September 28, 2010, from http://www.vanityfair.com/politics/features/2007/12/aristotle200712

Whitman, J. M., & Perkins, J.W. (2003). The technological evolution of campaigns: A look at new and emerging practices. In R. P. Watson & C. C. Campbell (Eds.), *Campaigns and elections: Issues, concepts, and cases* (pp. 47–56). Washington, D.C.: Lynne Rienner Publishers.

Zetter, K. (2003, December 11). For sale: The American voter. *Wired*. Retrieved from September 28, 2010, http://www.wired.com/politics/security/news/2003/12/61543