Even as the 2008 campaign cycle recedes in memory, that extraordinary time remains ever present. The data collected by the primary and presidential campaigns provides an indelible record of the volunteer efforts and political preferences of millions of citizens around the country. Every action taken in response to an appeal to donate or volunteer among the thirteen million citizens on Obama’s email list, for example, is etched in a database now housed under the auspices of the Democratic Party. As are the 223 million pieces of personal information citizens provided to Obama’s millions of online and on-the-ground canvassers.

Data lies at the invisible backend of contemporary campaigning. A vast data industry and infrastructure makes possible both ‘Web 2.0’ politics and the revival, over the last decade, of old-fashioned shoe leather campaigning. The organizing efforts behind Obama’s campaign are best understood as the culmination of a decade-long struggle by both parties to build better databases, integrate online and field data, create better analytic tools to make sense of these traces of the electorate, and develop more sophisticated technical and organizational infrastructures to realize and control electoral participation. Obama achieved what many other campaigns have only wished for. The campaign started with its comprehensive knowledge of the electorate and sophisticated modeling techniques and then deployed its vast sea of volunteers online and on-the-ground to generate a real-time portrait of its support and persuade undecideds. Data informed all of its resource allocations, from where to dispatch staffers to whether to spend additional dollars on internet advertising or acquiring emails.

While data has long been a part of political practice, changes in the nature of political data and ways of gathering, storing, and acting upon it have undergone a revolution over the last decade. And, even as these data practices support political participation and mobilization, they come with a social cost. Political data is collected and traded on a vast and opaque market, with documented cases of breeches in security. The extent and nature of political data has the potential to threaten associational freedom, as citizens become increasingly aware that much of their online and offline behavior is subject to monitoring and act accordingly. Meanwhile, as data grows more sophisticated, so does targeted, persuasive communication that speaks narrowly
to individual interests instead of general concerns. Only the most well-endowed candidates can afford sophisticated data and the modeling required to make use of it, which truncates democratic competition. Finally, political data fundamentally creates ‘informational asymmetries’ between candidates, their supporters, and voters that undermine mechanisms of political representation.

Institutional political actors that are not official arms of the state, such as parties, candidates, and advocacy organizations, have wide latitude to store and collect political data. Given the unique challenges to democratic life posed by data practices, institutional and technical innovations that secure political privacy while promoting participation, deliberation, and competition are necessary. To this end, this brief essay outlines a proposal for a new political data repository, to be housed within the Federal Election Commission (FEC). It is designed to provide more transparency around political data, give citizens the agency to represent themselves to political elites, and set firm boundaries around what can be collected and how it can be used to ensure robust political speech and competition among candidates. While the prospects for this proposal succeeding are unclear, at the very least it may bring the key commercial and political actors to the table for serious discussions about political privacy.

Data and the Political Process

There are a number of excellent works that document the history of political data. Gathering data on the electorate has a time-honored history, from early civil society organizations circulating petitions to identify supporters to targeted, direct mail communication using commercial and public databases.

Beginning in the 1990s, and carrying through to the present, both parties used advances in computing technologies to build national-level databases. The Republicans were the acknowledged leader in this domain until the Democrats launched a national voter database effort during Howard Dean’s tenure as Party Chair. These databases blend public data (vehicle registration, party registration, real estate records, etc.) with commercial data (magazine subscriptions, credit histories, grocery club card histories, etc.). The Democrats, in turn, brought their national party database online for the first time during the 2007-2008 cycle. This enabled the tens of thousands of field staffers and volunteers canvassing the electorate for campaigns from governor to president to continually update the database from field offices scattered throughout the country. The data that parties and campaigns have access to - including those
provided by third party private vendors such as Catalist - is extensive. 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.iii Data alone is not valuable, however; it needs to be made meaningful for campaigns through models that identify likely supporters and swing voters. To develop these models, a host of private firms poll voters for their clients to figure out the demographic and lifestyle characteristics of likely supporters and undecided voters - which then gets mapped onto these databases to create a set of individual voter targets.

The uptake of the internet in campaigning over much of the last decade has, in turn, resulted in a wealth of new data for candidates. Obama’s effort was, by far, the most sophisticated political campaign to-date with respect to gathering and using internet data. The campaign continually watched user actions on its website to ‘optimize’ its applications, resulting in millions of additional dollars and signups to the email list. The campaign segmented its email communications to supporters based on their demographics and involvement with the campaign, and tracked user actions to design the best appeals. The campaign also continually assessed returns on investment for all of its applications using data, calculating to the penny what the cost of acquiring emails and running internet advertisements netted for the campaign. The campaign also gathered data from individuals who “friended” Obama on commercial social networking sites such as Facebook, and synched it with voter databases to organize early primary states and turn out voters. This synching of online data with voter databases was an important - if halting - area of innovation on the Obama campaign, but one that represents a new frontier in political data. For example, the campaign’s Facebook Connect application enabled supporters to urge their friends to cast their votes early, merging voter registration with social network data in a way that leveraged social relationships for the ends of the campaign.

Why Political Data and Threats to Privacy Undermine Democratic Norms

There is no doubt that contemporary campaigns use political data to realize and coordinate levels of civic participation and voter turnout not seen since the era of strong party politics. Given this, how can political data undermine democratic norms?

*The potential for data breeches:* While breeches in data are a concern in any context, the data held by campaigns, parties, and private firms reveals information about policy preferences and political ideology and as such are particularly sensitive for individuals. The courts have long held that unauthorized disclosures of personal data are intrusive.iv Meanwhile, the potential for
data breeches is heightened by the fact that political data is traded on a largely unregulated and international market. For instance, multi-national credit firms such as Experian service much of the political sector and the technical development of campaign tools and databases occurs in a global market. And, in many states reporting of breeches is entirely voluntary for political organizations. All of which means that there is significant potential for breeches to continue to occur - and citizens may have little if any knowledge of them.

**Associational freedom:** A tradition of Supreme Court jurisprudence argues that privacy is important with respect to protecting anonymous speech and freedom of association. This encompasses the right to have a space to develop one’s political views and voice outside of the glare of the public. Privacy helps ensure robust political debate by providing the opportunity for citizens to form their own viewpoints, craft arguments, and develop political identities free from surveillance and public pressure, all of which also preserves a space for fostering dissent from prevailing social norms. Political privacy, in this sense, has the social value of promoting a robust public sphere.

The logic behind associational privacy is that if citizens feel as if they are constantly being watched they will be less likely to state their opinions and be wary of the media, organizations, individuals, and consumer goods they choose to associate with. While candidates and parties are not official arms of the state - which mitigates concerns over state-sponsored repression and surveillance - it is unlikely that citizens distinguish between uses of political data in the context of electioneering and governance. Given this, the issue is that if citizens start to realize that their expressly political affiliations (membership in organizations), ideological engagements (magazine subscriptions), and even mundane lives (grocery club card purchases) are being aggregated to create a portrait of their beliefs that is then used for political purposes, they will alter their behavior accordingly.

**Democratic debate:** While freedom of association indirectly affects public speech in having the potential to foreclose those spaces apart for the formation of political views, data also serves to shape political debate much more directly. With unprecedented opportunities to tailor information environments to individual preferences - from what citizens see online, hear in a radio advertisement, encounter when they meet a neighbor canvassing at their door or through Facebook, or even see in video games - campaigns engage in persuasive and mobilizational forms of communication designed to narrowly appeal to individual preferences. While speaking
in these narrow ways to citizens on the basis of their interests and needs as gleaned through traces of their consumer and online data may produce turnout at the polls, it does not produce much in the way of conceptions of the public interest or general good. Meanwhile, entire portions of the electorate, particularly those unlikely to vote (highly correlated with lack of education and low socio-economic levels) are routinely ignored by campaigns.

Democratic competitiveness and representation: Outside of the impact that political data has on the public sphere, it adversely effects democratic competitiveness and mechanisms for political representation. Each party sets its own rules (for the Democrats, on a state-by-state basis) about providing candidates and advocacy organizations with access to its political data. What that means is that elite-supported candidates often have exclusive access to party political data for the conduct of their campaigns. While there are a number of private vendors of political data, these services are often cost prohibitive - which in turn privileges candidates that are backed by interest groups such as unions that provide their own data and resources. If non-institutional and non-wealthy candidates of the major parties are at a competitive disadvantage in primaries, minor party candidates are left entirely out of the equation with few resources and comparatively little organizational infrastructure.

At the same time, the erosion of political privacy also has the effect of truncating political representation. The nature and extent of political data creates informational asymmetries that rework the relationship between representatives and citizens. This occurs on two levels. First, through facilitating legislative gerrymandering, data enables representatives to actually choose whom they will represent based on demographics and political ideology. Many representatives no longer have to engage with the views of their constituents and balance them in elite political deliberations about national, not simply parochial, interests. They can simply advance their own politics in local and national forums given the ideological homogeneity of their districts.

Asymmetries in information between political actors and voters, in turn, facilitates the ability of elites to manipulate the electorate. For example, candidates and their agents - supporters they enlist to spread their message and generate data on their friends and neighbors - in many cases know more about those they are seeking to represent than citizens do about them. This makes voter outreach fundamentally transactional and manipulative, as campaigns and their supporters strive to craft their political speech to accord with what individual voters want to hear. While this is not entirely effective, of course, it allows candidates to present only select aspects
of themselves to voters, providing an incomplete portrait of their policy preferences. Meanwhile, political data refashions relations between citizens in similar ways, creating power dynamics given that volunteers have data on their peers. In the process, citizen-supporters become active pitchmen for potential representatives rather than standing on equal terms with their peers in civil society with respect to political elites.

Rethinking the Role of the State in Campaigns: An Innovation to Protect Privacy

Innovations are needed to protect political privacy and create more robust election practices. Toward this end, the Federal Elections Commission should create from extant databases one online, centralized political database on the electorate that candidates must use and that citizens can access. Housing this database under the auspices of the FEC has a number of advantages. The FEC is already the chief regulatory body that monitors federal elections, and this mandate can be expanded to include state campaigns. All candidates who are on the ballot will have access to this database for electoral purposes through an online system that their staffers can log into. Campaigns can then run canvassing operations using this database, but all data generated from personal contacts must be uploaded into this central repository - which is then accessible to all candidates. Requiring the use of one, centralized electoral database (and prohibiting alternatives) would improve data security by ensuring that there is only one repository for political data instead of the dozens that currently exist. Any breeches, meanwhile, will instantly be made public. This FEC data, in turn, will not be shared with any other agency or branch of government. Meanwhile, campaigns will not prohibited from collecting separate data using the internet, provided that it is individually-unidentifiable and disclosed. Supporters can also continue to use third party applications synched with campaign Websites, such as what is currently offered through Facebook Connect, provided that campaigns do not mine this data for information.

Importantly, this system would improve institutional transparency and give citizens more control over their personal data to protect their privacy. When this election database is culled together from extant data controlled by the parties and private firms, citizens will be able to access their profiles and edit and delete sensitive information. Meanwhile, aside from basic voter and party registration data, and extant spending disclosure laws, all of which are necessary for the conduct of elections, all other information will be opt-in and citizens will be able to specify on an ongoing basis the level of publicity they wish their data to have. Citizens will also
control their own degree of exposure to political actors. For example, if an individual supports a particular candidate, she can designate certain personal information (such as volunteer and donation history) to be revealed to only that candidate. She can also choose to share that information with the party of her choosing, and decide to retain this information across election cycles. What is important is that the citizen herself controls the visibility of and access to her data. This, in turn, will serve the end of associational freedom, as citizens can opt out of revealing such things as their magazine subscriptions and online behavior.

This central database, in turn, would foster more inclusive democratic debate, robust political competition, and stronger forms of political representation. Knowing less about the electorate would force campaigns to speak to more voters. It would also mitigate the targeting of political communication, as campaigns would have less knowledge of the electorate and be less able to draft narrow appeals based on demographic and affiliation interests discovered through data. Campaigns would have to rely on cruder measures of parceling out the electorate, which in turn would lead candidates to speak more to median voters and make appeals to more general interests. This plan would also help level the electoral playing field - setting insurgent candidacies and third parties on equal footing with major-party candidates in providing an information subsidy essentially on par with public financing of elections. Finally, one centralized database would create cruder technical means of redistricting, while helping to create conversational parity, restoring some semblance of informational balance to the interactions of candidates, canvassers, and voters.

---


\(^{vi}\) In the United States alone, political data has been sold to shore up the finances of failing firms, campaign Web sites have been hacked, questionable data practices have been exposed, and American voter data has turned up in surprising places overseas. For a review, see P.N. Howard and D. Kreiss (2009) “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.