

Constituency Campaigning in the Age of Data

Kaija Belfry Munroe, PhD
Professor, Quest University Canada

H. D. Munroe, PhD
Professor, Quest University Canada

This paper presents preliminary findings of a larger research project; please do not cite without permission.

Presented at the 2016 conference of the Canadian Political Science Association, May 31 - June 2, 2016, Calgary, Alberta. Please direct correspondence to kbm@questu.ca.

The application of “big data” by the Obama campaigns of 2008 and 2012 has been touted as a revolution in American politics (Issenberg, 2012a, b; Germany, 2009, 2014). Issenberg claims, for example, that the Obama team in 2012 was confident that they knew the names of the 69,456,897 American voters who had elected the president in 2008 (Issenberg, 2012b). Given the successes of Obama, it is unsurprising that all three major Canadian federal parties - the Conservatives, the Liberals, and the NDP - contested the 2015 general election armed with large-scale data management tools. Indeed, the Conservatives have been using a comprehensive voter database since 2004 (Flanagan, 2007). But are Canadian campaigns, which contest 338 simultaneous short, small, local elections instead of a single, year long, continent-wide presidential marathon, data-driven in the same way? What does it mean to run a data-driven campaign in Canada? If local campaigns in Canada can indeed be called data-driven, what are the implications for electoral competitiveness and for party structure? Is constituency campaigning undergoing a data *revolution*, or are applications of data - if any - simply accelerating campaigning-as-usual?

In this paper we present findings from a case study of the uses of data by all three major political parties in a single constituency during the 2015 election and develop a framework for analyzing that use, providing conceptual clarity in an otherwise murky literature. Empirically, we found considerable variation in how data was employed across the local campaigns, but that one of the three campaigns could indeed be considered data-driven within our framework. While the effect of data-driven campaigning on electoral outcomes remains unclear, we find that the emergence of data as a resource could have a significant centralizing impact on stratarchical party organization in Canada (Carty, 2004).

Local campaigning in the age of data

Local campaigning in Canada

Traditional examinations of local election campaigns in Canada either focus on the impacts of financial contributions on electoral success and/or party organization (Colletto et al., 2011, Eagles, 1988 and 2004), or on the significance of local-level campaigning - identification of support, communication of message, and mobilization of vote on election day - for electoral outcomes (Carty and Eagles, 1999). The literature highlights financial assets and volunteers as key campaign resources; increases in either of these have been found to positively correlate with electoral outcomes (Carty and Eagles, 1999).

Where does data fit within this context? A key question is whether data can be conceived of as a resource alongside volunteers and finances and, if so, what impact this has on campaigning. We suggest that some campaigns, particularly the Obama campaigns of 2008 and 2012, perceived data as such. While the Canadian campaign literature was largely established before the data innovations of the Obama 2008 and 2012 campaigns, Carty and Eagles (1999) did predict the impact of new technologies on campaigns:

“It seems to us equally plausible that the diffusion of computer technology, the availability of geographic information systems containing rich data on constituency electorates, and the spread of campaign

professionals into the trenches of constituency electoral battles, will strengthen these local forces, empowering rather than enslaving local party organizations.” (Carty and Eagles, 1999, p. 83).

Of particular interest here is the suggested implication of the increased employment of computer technology and data for the power relationship between local party organization (riding associations, candidates and electoral volunteers) and central party organizations. This relationship, described by Carty (2002, 2004) as a franchise model, is one in which both the local and national campaigns have their own spheres of activity in which they are largely autonomous, creating a “stratarchical” rather than hierarchical organizational structure. As Carty and Cross (2006) explain:

“The party in public office determines both parliamentary and electoral policy and disciplines its membership: the party on the ground determines just who becomes (and stays) a member of the party in public office” (2006, 98).

What is generally not discussed in this literature, however, is how local campaigns in Canada employ (or do not employ) data to inform campaign decision making, and how this plays out in the context of stratarchical party structure. Observers have noted the Conservative Party’s so-called data dominance (Flanagan, 2014) and the shift towards politics-as-marketing (Delacourt, 2013), but there is a lack of focused studies on the use of data across campaigns in this country.

Data-driven campaigning

What does it mean to use data in an electoral campaign? All campaigns gather and use information about the electorate, but not all campaigns necessarily use data, i.e. “a description of something that allows it to be recorded, analyzed, and reorganized” (Mayer-Schonberger and Cukier, 2014, 78). To transform information into data, one requires a process of recording that information “in a quantified format so it can be tabulated and analyzed.” (Mayer-Schonberger and Cukier, 2014, 78) The literature on applications of data in campaigning is nonetheless fragmented in large part because it either discusses technology writ large (Germany, 2009, 2014; Bimber, 2014; Medvic, 2009; Norquay, 2008), or because it focuses on a narrow collection or use of data for electoral purposes, particularly microtargeting (Barocas, 2012; Franz, 2013, Johnson, 2010; Bennett, 2015). Scholars have also highlighted the ominous implications of what Bennett (2015) terms “voter surveillance politics” (Jamieson 2013, Barocas, 2012; Franz, 2013) for the health of democracy. What is lacking, however, is a broader articulation of the concept of data-driven campaigning, which can incorporate the many uses and effects articulated in the literature.

Thus, we propose a three-fold conceptual framework for understanding data-driven campaigning:

1. Conceptualization of data as a resource: To be considered data-driven, the local campaign organization must understand data to be a resource in its own right.
2. Data generation by the campaign: The campaign must generate data, describing something in a format that allows discrete facts about the thing to be recorded, quantified, reorganized, and analyzed. The literature suggests that this may be done by:

- i. integrating existing data of various kinds with data gathered during the campaign (Johnson, 2010; Flanagan, 2014);
 - ii. generating probabilistic inferences about voters (Bimber, 2014; Barocas, 2012; Bennett, 2015); and/or,
 - iii. tracking the campaign's activities (Bimber, 2014; Germany, 2009, 2014).
3. Data employment in decision making: Data analysis must demonstrably guide campaign decision-making in some way (Bimber, 2014; Karpf, 2013; Johnson, 2010; Issenberg, 2012a, b).

Data as a resource

Almost unremarked in the discussions of the large-scale collection of voter data in American presidential politics (Issenberg, 2012a,b; Bennett, 2015; Rubenstein, 2012, Franz, 2013) is that the campaigns involved - notably the Obama campaigns of 2008 and 2012 - understood data as a resource in its own right on par with volunteer time or money. We do not assume that Canadian local campaigns necessarily do this, in contrast to Bennett (2015). Such an understanding is a prerequisite to generating and using data in a systematic manner in campaigning by, for example, developing infrastructure to manage it, or expending effort to collect it.

Data generation: Integration of known voter data

The easiest way for parties to generate data about voters is to integrate existing stores of information. National parties and local riding associations, for example, have long had information about donors, volunteers, party members, and people who could be relied upon to put up a lawn sign. Though the literature suggests that this information was digitized some decades ago in the United States (Medvic, 2011), it is only relatively recently that parties in the US and Canada began integrating their disparate repositories of this information into common databases that could link together what the party knew about a specific person as a donor, a volunteer, a party member, *and* a voter, and organize that information geographically so that it could be sorted and searched by riding or polling division (Issenberg, 2012a; Flanagan, 2014). The rise of integrative data infrastructures has transformed voter contact into a two-way exchange, in which identifying voters as confirmed supporters, confirmed opponents, or undecided becomes equally or more important than conveying a particular message to them (Johnson, 2010.)

The defining feature of integrative data creation is that it rests on the collection and aggregation of 'hard' facts, collected through official records (Elections Canada's voter records), voter contact (doorknocking or phonebanking) or party records (donor or membership information). These data are systematically collected and recorded using infrastructure conceived for that purpose.

Data generation: Inference of unknown voter data

In the absence of hard data about each voter's preferences, campaigns may try to infer those preferences by analyzing other data about voters. This is generally what is understood by microtargeting. All campaigning is targeted along both geographic and demographic lines (Flanagan, 2014; Ridout et al, 2012). Microtargeting simply implies finer scales of geographic or

demographic resolution, and indeed the scale referred to by the term is decreasing: In 2004 and 2008, it meant small but still aggregate demographic groups (Norquay, 2008; Hardy et al, 2010). By 2012, the term was being used to refer to an ideal of campaigning at the scale of individuals (see Franz, 2013, Rubinstein, 2012; Issenberg, 2012b; Jamieson, 2013). As available information about voters becomes ever more detailed, it becomes possible to calculate, based on hundreds of discrete data points about *every* voter, their probability of voting and their probability of supporting one candidate or another. The predictive voter modelling undertaken by the 2012 Obama campaign is the point of reference for this kind of data creation, which takes a relatively small number of hard voter identifications and examines hundreds of points of data about those voters in search of patterns from which ‘soft’ probabilistic identifications can be interpolated for all voters with similar demographic profiles. (Bimber, 2014; Issenberg, 2012b).

Data generation: Tracking of own activity

As campaigns increasingly use customized IT platforms in the routine operation of their campaigns (Medvic, 2009; Germany, 2009, 2014), it becomes increasingly easy to automatically generate data about about one’s own campaign, donors, and volunteers (Bimber, 2014; Norquay, 2008). One of the major advances of the 2008 and 2012 Obama campaigns, for example, was the development of web-based volunteer management systems that made it easy for the campaign to direct volunteer effort while automatically feeding data about those volunteers’ activities back to the campaign (Bimber, 2014; Issenberg, 2012b; Norquay, 2008; Germany, 2009, 2014) This, in turn, gave the campaign an ability to monitor and analyze its own progress and effectiveness (Issenberg, 2012a,b; Germany, 2014).

Data and decision-making

Finally, a “data-driven” campaign is one in which decisions are guided by the use of data rather than by instinct, guesswork, intuition, tradition, or rules of thumb. It is the use of data *per se* that is definitive: the three forms of data creation described here - integration, inference and tracking of activities - are independent of one another, and we deliberately make no judgements about their relative importance. What matters is that the local campaign conceptualizes data as a resource, goes to some effort to collect and manage it, and makes decisions about its activities (where to canvass, whether to send direct mail and to whom, etc) based on analysis of that data. A local campaign that received copious and detailed analyses from the central party, but made no use of them in its decision-making, could hardly be said to be data-driven. Many scholars note a shift from intuition-based campaigning to a culture of evidence and testing (Bimber, 2014; Karpf, 2013; Johnson, 2010) that may be following the turn towards professionalization noted by Carty and Eagles (1999); as such, we had some basis for expecting to see such an approach to campaigning at the local level.

Methodology

In order to analyze the use of data in campaigns across multiple parties, we chose a single-case comparative methodology that focused on the three front-running parties in one riding during the 2015 federal election. This allowed for the use of qualitative research methods, which we felt best suited to in-depth analysis of campaign activities. Specifically, we engaged in participant observation with the Conservative, Liberal and NDP campaigns in the riding of West Vancouver-Sunshine Coast-Sea-to-Sky in British Columbia. Different researchers engaged

with the Liberals and the Conservative campaigns to provide those parties greater comfort that information would not be shared until after the election (The NDP was not bothered). We took part in data training sessions, phone-banked, doorknocked, organized GOTV, hung out in headquarters, and scrutineered with each campaign, depending on their methods. The goal was to collect data in order to use the framework to answer the research questions above, namely to understand whether local campaigns in this election were data-driven and whether this impacted the relationship of the local party association with the central organization. We bolstered this participatory research with elite interviews with campaign staff and volunteers.

The riding of West Vancouver-Sunshine Coast-Sea-to-Sky was selected for two reasons. First, both authors are faculty members at the only university in the riding; our professional standing and the familiarity of all three parties with the university, as well as previous acquaintance with some of the significant players in both the Liberal and Conservative campaigns, meant that we were able to establish the trust needed to participate in the campaigns, document their methods and interview their staff and volunteers, under the condition that no information would be shared with the other campaigns during the writ period. Second, the riding is a challenging one for campaigning: it is made up of five urban centres, linked together by rural areas dotted with small communities. The main urban nodes of the riding are West Vancouver, which holds almost 50% of the electorate; Squamish, Whistler, and Pemberton (in the Sea-to-Sky corridor); and Gibsons on the Sunshine Coast. Pemberton is about a two hour drive from West Vancouver, while Gibsons is only accessible by ferry from West Vancouver. The dispersed nature of the riding means that no candidate is likely to have significant name recognition in every community, and that all candidates needed to be able to campaign effectively in communities other than their homes (West Vancouver for the Liberals and Conservatives; Gibsons for the NDP.) We suspected that this would increase the incentives for all campaigns to use data to increase their effectiveness.

West Vancouver-Sunshine Coast-Sea to Sky was perceived as a safe seat by the Conservative party, a battleground riding by the Liberal party, and a faint hope riding by the NDP. In 2011, John Weston, the Conservative Candidate, was re-elected comfortably with 45.53% of the vote. The Liberal and NDP candidates received 22.47% and 23.59%, respectively, with the Greens following at 7.06%. In 2015, however, the Liberal party had recruited the popular former mayor of West Vancouver as a candidate. Believing that this would allow them to break into the traditionally conservative areas of West Vancouver, and given that the rest of the riding has historically leaned towards the left, the Liberals thus perceived the riding as a winnable seat. Meanwhile, electoral redistribution prior to the 2015 election led to the removal of Powell River, a mill town with a significant union base, from the riding; this deprived the NDP of its main support centre and the riding entered the faint hope or “orange wave” category for that party. Thus, West Vancouver-Sea-to-Sky-Sunshine Coast offered a unique opportunity and also larger methodological challenge; the three parties studied related to the riding in three different ways – one perceived a safe riding, one a battleground and one a faint hope. This means, however, that the way the local campaigns behaved might differ due to the interest the central party had in the campaign based on how likely they were to win. Our analysis, therefore, can only be said to demonstrate how these three parties used data under these particular conditions and cannot be viewed as universal across the three parties. There were, however, clues in the actions of these campaigns about the parties’ overall strategies related to the use of data and we have tried to parse these out as much as possible.

Observations

The conceptual framework above provides an important tool for collating and analyzing the observations from the participatory research and interviews. Below, we analyze each party in relation to our criteria for a data-driven campaign discussed above. As we will demonstrate, there was significant variation across campaigns as to how data was employed and whether campaigns were truly “data-driven”.

Conceiving of data as a resource

Our observations indicate that, on a national level, the Conservatives, Liberals, and NDP clearly perceive data about voters, supporters, and volunteers as a resource. All three have invested considerable effort in building national infrastructures to manage large data holdings, and the national campaign provides tools and training to local campaigns to enable them to gather more such data and make use of it. While these infrastructures are broadly similar, at least from a user perspective, there are some notable differences in capability.

What are these infrastructures? The best known is the Conservative Party’s Constituent Information Management System (CIMS), which dates from 2004 (Flanagan, 2007). The Liberal party used a platform called Liberalist, which appears to have been in limited service in the 2011 election but for which many functions were apparently in full-scale operation for the first time in 2015. The NDP also entered the 2015 election with a new platform, tested in by-elections, called Populus. This replaced an earlier system known as NDPVotes. While technical back-end details of these systems were not available to us, from a user standpoint they are highly similar: A web-based interface allows an authenticated user to search for voters by name in their assigned riding, and for each voter to see a record of that person’s contacts with the party (have they ever volunteered, have they ever donated, have they taken a lawn sign), the party’s contacts with that person (either simply whether or not the person is a supporter, or a comprehensive list of how and when they were contacted); and their voting history. The interface also includes volunteer management tools: in most cases, for example, users can sign up for phone banks, door-knocking, or other events, and the database will generate whatever forms of voter data are necessary to support the event - such as worksheets that can be printed out and carried by volunteers while door-knocking. These worksheets list the names of voters at each address on a specified route, and offer volunteers a simple, standardized way to record that voter’s level of support - for example, by circling an icon of a happy face for supporters, a neutral face for undecided voters, and a frowning face for hostile voters in the case of CIMS.

All of the platforms appear to assign different levels of access to users based on their roles in the campaign organization; users in leadership positions have access to functions which simply do not appear on the screen of campaign volunteers. Among these are the ability to generate reports from the database to list, for example, all people who have either volunteered or who have been identified as a strong supporter within the last five years in a specified geographic area in the riding. Because of our positions within each campaign, our direct experience with these higher-level functions varied; nonetheless, we observed enough to know the broad outlines of their capabilities. Though the platforms are largely similar, there are some minor variations. CIMS, for example, includes fields that allow for information about issues of concern to individual voters to be recorded. Because of the tight integration between the provincial and

federal NDP, at least in BC, Populus tracks a person's voting history across both federal and provincial elections (and presumably tracks party support and engagement across both levels, as well.) Liberalist, at least from our observations, seems to have a much richer set of features for organizing and managing volunteers - including an embedded email interface to make it easier to contact volunteers or supporters *en masse*. Finally, all of the platforms share one key characteristic: they are centrally-managed and national in scope, such that information entered into them anywhere in Canada, by any branch of the party, is retained by the national party organization and is accessible by any other branch of the party depending only on clearance.

The fundamental input for these databases is the national list of electors, which is regularly produced and distributed by Elections Canada. The national parties also acquire some data through bulk purchase (telephone numbers, for example, may be purchased from telephone companies; market research firms such as Environics have long offered a variety of demographic market segmentation datasets); such data can be automatically merged with the list of electors. The lion's share of the data in these systems, however, must be collected by local campaign organizations. The degree to which they do so is variable, and the quality and extent of data available to a particular local campaign depends greatly on how good a job previous campaigns in that riding have done of data-gathering.

In our sample, the local organizations varied considerably in the extent to which their leadership conceived of data as a resource independent of (volunteer) time or money. Those which did not think of data in this way still had the national data infrastructure at their disposal, but did relatively little either to contribute to it or exploit it. For the local Liberal campaign, data was a crucial resource. The campaign manager had previously worked in the 'data office' of the national organization and under his leadership all levels of the campaign seemed geared toward data collection and utilization. The local campaign team included a data manager, and considerable effort was put into both gathering and acting upon data about voters and about the campaign itself. These efforts seemed closely integrated with those of the national party, as will be discussed below.

Volunteers in the Conservative campaign, meanwhile, referred to the party's data advantage in conversation, and clearly thought that this was an asset to them in the election. When asked, they clarified that the data they meant was about individual supporters - email addresses, for example, and data about issues that people cared about, made it easy for the campaign to attempt to reproduce the same winning coalition of voters from 2011. This understanding of data as a resource was evident in the campaign's door knocking and phone canvassing, in which we were told the purpose of the interaction was to determine if a person was a Conservative supporter or not, rather than to persuade them to vote Conservative.

In the NDP campaign in the riding, while the campaign leadership saw the national party's database infrastructure as being useful for tracking supporters and volunteers, there was little effort put into collecting data. Volunteers doing door-to-door canvassing, for example, were given worksheets generated by Populus, but it was clear that the campaign saw their efforts as being primarily about persuading voters on the doorstep. Data gathered about supporters was only half-heartedly entered into Populus. Our observations suggested that in the eyes of the local campaign, Populus was little more than a fancy address book, and that the dominant paradigm was one of unstructured information rather than structured data. The campaign did not conceive of the data itself as a resource in the same way as the Liberals or Conservatives.

Creating data

The campaigns that we observed created data in various ways, but only one - the Liberal campaign - engaged in all three modes of data creation we describe: integrating known voter information, inferring unknown voter information, and tracking one's own campaign activities.

Integrating known voter data

All parties' databases support integrative data creation in that they allow for new information gleaned about that specific person through direct contact to be recorded as a data point. They also allow for tracking of voter behaviour over time, notably whether or not a specific individual has voted in past elections (as parties can now systematically track election-day reporting from Elections Canada and retain that information indefinitely.) In our sample, the Conservative and the Liberal campaigns invested considerable effort in creating known voter data - so much so that data-gathering seemed to be the primary activity of both campaigns!

The Conservative and Liberal campaigns had local volunteers canvass voters both door-to-door and by phone with the goal of identifying voters as supporters, hostile, or undecided. This was complemented by paid telephone canvassing. For the Liberal campaign, this was organized and paid for by the local campaign. Respondents in the Conservative campaign, meanwhile indicated that the national party runs a paid phone bank which often calls known supporters (based on CIMS data), but that this effort was usually directed at battleground ridings rather than apparently safe seats like ours. The local campaign's efforts during the writ period, however, produced uneven results: We observed that it was very hard for volunteers to learn a person's name without alienating them on the doorstep, so the accuracy of the data recorded - which is in theory linked to names on the list of electors - could be suspect. In particular, the Conservative volunteers seemed unconcerned about the potential of recording incorrect data.

Two things make this voter identification effort noteworthy in comparison to traditional practices: First, the results are recorded as data, i.e. in a quantified, systematic fashion intended to enable easy tabulation, reorganization, and analysis. Second, data generated in the writ period is combined with existing data, such as an individual's voting history, past party membership, past donations, and other contacts with the party. Such contacts need not necessarily be election-related: Respondents informed us that Conservative Members of Parliament are instructed to record data based on interactions with constituents, as well as constituent responses to leading surveys placed in parliamentary mailings available to MPs. It is this ease of analysis and chronological depth that distinguish integrative data from more traditional methods of recording the results of voter contact efforts.

Both the Liberal and Conservative efforts at data-gathering were supported by a mobile application running on volunteers' smartphones or tablets. This is a logical extension of canvassing as a data-gathering exercise: Volunteers are able to enter data about voters directly into CIMS or Liberalist (using mobile apps called CIMS2GO or MiniVAN, respectively) in the field, rather than recording data on paper and then entering it manually into the database later. Both apps presented volunteers doing door-to-door canvassing with a list of addresses on a given street, and for any address, a list of the voters known to reside there. For each voter, the volunteer can record whether or not they are a supporter or not (by labelling as supporters or

not in MiniVAN, or happy/neutral/frowning face icons in CIMS2GO). Both apps had the ability to record other information about the voters, such as whether or not they wanted a lawn sign or should be approached to donate, but in the case of the Conservative campaign, volunteers were told not to bother with these fields. The Liberals, on the other hand, wanted their volunteers to collect as much information as possible, including ethnicity and whether they had children under six. Both apps carefully structure what kinds of data volunteers can enter: it was possible to correct mistaken entries in CIMS2GO, for example, but in neither app could a volunteer delete the name of a voter from a specified address if that person no longer lived there, nor add new names. Cims2go and Minivan were also clearly designed to be used for other kinds of volunteer efforts, including phone banking (both apps could generate phone numbers to call in target areas of the riding as determined by a local campaign organizer) or sign placing; in the Conservative campaign, however, its use was restricted to door-to-door canvassing and tracking voters on Election Day using Elections Canada's bingo sheets at polling stations.

In the NDP campaign, we observed little systematic effort to generate new data by integrating existing voter information in Populus with data gathered through door to door canvassing. Walk sheets were apparently generated for door-knocking, but there was relatively little emphasis placed on recording voters' level of support. This is done in Populus using a system of marks in which 1 indicated a strong NDP supporter, and 2L, 2C, 2B, and 2G indicating (we believe) a voter for whom the NDP was a second choice to the Liberals, Conservative, Bloc Québécois, or Greens, respectively. The NDP campaign appeared to place much greater emphasis on raising the visibility of their campaign through mainstreeting (having the candidate meet voters in busy urban areas) or burma shaving (waving signs at high traffic intersections) than the other campaigns.

Inferring unknown voter data

We neither observed, nor had any indications from well-placed respondents, that the local Conservative or NDP campaigns were making use of any kind of predictive voter modelling. Indeed, though our locally-based approach makes it impossible to know for certain the full extent of a national campaigns efforts, it appears that neither of these parties have data infrastructure designed to allow such voter modelling to be carried out. While CIMS does allow a numerical score ranging from negative 15 (indicating a hostile voter) to positive 15 (indicating a committed supporter) to be assigned to each individual voter, these scores are based on interactions with that voter - quite different from a probabilistic assessment that is inferred from other data about that voter.

Liberalist is thus quite different from CIMS and Populus in that its integrative data generation functions are a basis for a much more ambitious program of predictive voter modelling. The database includes a separate field to record data generated by predictive voter modelling, which ranks all voters into ten tiers - with tier 1 denoting voters who are all but certain to be Liberal supporters, and tier 10 denoting voters who are all but certain to be hostile to the Liberal party under any circumstances. These predictions were made on the basis of both their past contacts with the party and voting behaviour, information gathered by voter contact during the campaign, and other demographic and individual data - the full extent of which our respondents were unable or unwilling to discuss, but which seemed to include commercially purchased bulk data. The net result is that every voter has a predictive score assigned to them. Hard IDs based on direct voter contact are then used to check the accuracy of the predictive model itself.

Predictive voter modelling was usually referred to simply as “analytics”, defined by a key respondent as “the use of data... all the data we collect and process, in order to predict behaviour.” Like the Obama campaign of 2012, both a person’s probability of voting and their probability of being a Liberal supporter were calculated. In a sense, this allowed the Liberal party to not just rival but to exceed the Conservative party’s extensive collection of integrated voter data.

Tracking campaign activity

The integrative platforms of all three parties allow incidental tracking of the campaign’s efforts - the Conservative party, for example, uses CIMS to coordinate between central and local phone banks, since both phone banks update the same database and generate their call lists by filtering out voters who have already been contacted. This is not the same, however, as deliberately developing data about one’s own campaign for its own sake. Both the Conservative and Liberal parties have infrastructure for doing this, though only the Liberal campaign seemed to make use of it.

For the Conservative party, most of the capability to track the efforts of campaign volunteers comes from CIMS2GO. The app, which was designed in-house, is clearly intended to be the tool with which volunteers engage with the campaign: in addition to generating walk lists and allowing on the fly data entry from door knocking, and similar functions for phonebanking, the app also tracks and displays how many doors the volunteer has knocked, how many calls they have made, and how many signs they have handed out. What’s more, this information shared with all other volunteers in that riding in the form of a leaderboard - quite similar to a tool used by the Obama campaign to motivate its volunteers (Issenberg, 2012b). In addition, CIMS2GO uses the location-aware feature of most smartphones continuously while in use - suggesting that the exact geographic position of the volunteer could also be tracked. Given all of this, we were surprised that the local Conservative campaign did not appear to use any of these features to track their campaign efforts. So far as we observed, progress in door-to-door canvassing was recorded simply by highlighting streets that had been canvassed on wall maps of the riding; the only tracking of phone banking appeared to be recording which voters had been called and reached, rather than when or by whom.

The Liberal campaign, on the other hand, was able to generate quite rich data about their progress towards internal goals because of the features of Liberalist and MiniVAN. Moreover, the central party could also track their effort - distributing regular summaries of which local campaigns were leaders, and which were laggards, for overall number of doors knocked, voters contacted, etc. At any given time, the local and national campaigns could readily determine which areas of the riding had not been canvassed and, when new paid staff became available, dispatch them to those areas. By Election Day, the local campaign knew exactly how many identified supporters they had, and because of this decided to include not only confirmed supporters but also unconfirmed but highly probable Liberal supporters in their voter mobilization effort in certain areas.

The local NDP campaign, meanwhile, kept track of their efforts in a more traditional way (much like the Conservatives): Highlighting maps on walls and using Populus, albeit in a less systematic way, to keep track of which voters had been contacted. Given that the campaign was

significantly smaller in scale than either the Liberals or the Conservatives, there was little need for any more systematic effort.

Data-driven decision-making

As we note above, merely recognizing that data is a resource and then expending effort to generate it does not necessarily a data-driven campaign make. What matters is that the campaign leadership make decisions about what to do based on analysis of data - and in this we found significant variations between the three campaigns we observed. All campaigns began by looking at polling divisions where they had significant vote share in previous elections as a guide to focus their efforts; not all campaigns went significantly beyond this in terms of making decisions based on data.

The rhetoric of data dominance notwithstanding, the Conservative campaign was data-driven in an almost accidental way: Gathering voter identification data was a significant preoccupation of the campaign, and the fact that software filtered certain addresses or phone numbers out of the walk or call sheets for volunteers, regardless of the reason, represents a use of data to shape the tactical effort of the campaign. The voter data that was gathered was used to drive an Election Day get out the vote effort in which Elections Canada records of who had voted were fed directly into CIMS via CIMS2GO. The local campaign then directed phone calls to confirmed supporters who had not yet voted in an effort to turn them out at the polls. Overall, though, the Conservative campaign was a decidedly traditional one: Decisions about where to canvass, for example, were based largely on intuitive judgements of where the party's supporters were, with West Vancouver and more affluent areas of a few other communities getting the vast majority of attention. Consequently, while the campaign had access to significant amounts of data, it would be a stretch to say it was "data-driven".

The Liberal campaign, by contrast, was heavily shaped by voter data. The national party had a pre-determined number of supporters that they believed the local campaign would need to identify in order to carry the riding, and voter contact efforts were geared towards this target. The local campaign would receive daily lists of priority polling divisions in the riding, developed by the analytics team at central party offices, and these lists would be used to direct the day's canvassing. The party believed that it was the fact of contact with a voter that mattered, not the message that was delivered, and as a result, in the words of a key respondent, was that "the bulk of the effort on our analytics team is about who to say it to, or even who to talk to." Entire towns were ignored until the end of the writ period because analytics suggested that canvassing there wasn't the most efficient allocation of scarce volunteer resources. What the campaign did, and its awareness of what it had already done, was heavily reliant on systematically gathered and quantified data, and all of it built to a heavy voter mobilization effort on Election Day that relied on systematic use of hard voter IDs, predictive models, and tracking of the volunteers engaged in the ongoing get-out-the-vote effort.

Although the NDP campaign did use Populus to generate a list of target polls for mobilization efforts on Election Day, it is unclear if this list was much different from the target polls the campaign identified at the outset of the writ period based purely on past election results. The local campaign as we observed it did not seem to be driven by data in any significant way. This may well have been an effect of the riding being a faint hope for the NDP; we are aware that considerable effort was invested in other ridings. Nonetheless, the observed variations in the

conceptualization, generation, and use of data among the local campaigns in our riding give rise to some interesting implications.

Discussion and implications

Ultimately, this research demonstrates that “data-driven” campaigning is occurring in Canada, and that its extent varies between parties, and likely within parties as well. Certainly, all of the central organizations of the three national parties conceive of data as a resource, but their capacity to create data is dependent on volunteer or staff willingness and technological skill at the local level.

In our riding, the Liberal campaign was the only one being provided significant support in both financial and human resources from the national campaign. The campaign was consequently highly professional with at least four paid staff members, including a professional campaign manager and deputy campaign manager. These experienced operatives were well versed with Liberalist and understood data as a long term resource that would serve the party in future elections. We were told, however, that not all Liberal campaigns (particularly those in faint-hope ridings or safe seats) would have been so focused on data creation and utilization. The Conservatives and the NDP, on the other hand, did not receive support in the form of paid staff from headquarters and their campaigns relied exclusively on volunteers and volunteering parliamentary staff. While data was clearly important to the Conservatives, both Conservative and NDP campaigns appeared shaped by long-term volunteers doing things the way they had always been done. In addition, the bulk of the volunteers in this riding for all parties were seniors, leading us to suspect that generational factors likely also affect the extent to which local campaigns generate and exploit data. Consequently, it seems likely that all parties will see variation in the use of data across ridings depending on the demographics of volunteers and the professionalization of the campaign. This further supports our finding that while data is conceived by national parties as an independent resource, it - like volunteers and money - is unevenly distributed among local campaigns.

As Coletto et al (2011) noted with respect to financial resources, the way in which data is managed has implications for party structure. It is clear that the datafication of election campaigns in Canada was a strategic decision of the central wings of all parties we observed. We were told by respondents that the main national parties had all hired consultants from the past Obama campaigns to help them learn how to engage with data as a resource. Moreover, while the burden of gathering voter identification data rests with local campaigns, the data in each party’s infrastructure is owned and managed by the national party, and access is granted to local entities at the national party’s prerogative. These infrastructures are designed in ways that increased the level of surveillance and control of local campaigns by the central party organization. We argue therefore that data-driven campaigning has the capacity to fundamentally change the relationship between the local riding association and the central organization of the party.

This was evident in the Liberal campaign, which was by far the most data-driven in this riding and also the most influenced by the national organization. The list of priority polls received daily from national headquarters determined where local volunteers were sent to canvass. While this appeared highly efficient, it also meant that local knowledge that would have

previously been paramount to campaign decision-making was ignored. In Squamish, for example, the highest density polling division - composed of townhouses and apartments inhabited by middle income Canadians to whom the party's messaging was chiefly directed - was not extensively canvassed. Polling divisions consisting of single family homes - which would have been middle class in Ontario, but are highly affluent in the Vancouver area, and therefore more likely to hold Conservative supporters - were canvassed extensively. In Whistler, meanwhile, volunteers were often sent to polling divisions with significant numbers of vacation homes inhabited only for a few weeks per year, and not to the neighbourhoods where residents actually lived. Had the campaign sought the knowledge of local volunteers instead of relying on numbers sent from Ottawa, this could have been avoided. Nonetheless, it appeared that the central party's determinations of where canvassing should take place were sacrosanct.

Meanwhile, the highly detailed tracking of local campaign activity allowed the central party to attempt to micromanage other aspects of the campaign. As previously noted, the central party distribute regular comparisons of different local campaigns' progress in door-knocking, phone canvassing, and similar activities. One respondent said he could see no other purpose of such statistics then to "berate us" for appearing to lag behind other ridings (despite the obvious differences between high-density urban ridings brimming with youthful volunteers and geographically dispersed ridings with fewer, and older volunteers.) For the Liberals, the biggest protector of local autonomy was their highly regarded professional campaign manager, whose reputation for winning afforded him considerable latitude from the directives of the national campaign. But, since he himself was a paid party operative from outside of the local riding association, that is hardly evidence of local empowerment. While the Conservative national organization appeared to have little interest in controlling the local campaign, meanwhile, its technological infrastructure clearly allows for similar tracking and micromanaging.

Thus, while Carty and Eagles (1999) were correct in predicting that computer technology, GIS-linked data, and professional operatives would affect the way campaigns operate at the local level in Canada, we have found, contrary to their prediction, that this appears to increase the power of the central organization in proportion to the perceived importance of data at the local level. As campaigns become more data-driven, then, it is possible that Canadian parties will cease to be conventionally stratarchical, at least where elections are concerned.

But will local Canadian election campaigns all tend towards the systematic collection and exploitation of data? Obviously, a widespread perception that large-scale exploitation of data produces electoral victories will push all parties towards this outcome - regardless of the actual effectiveness of such tactics. This study can be merely an exploratory foray into the question of the empirical effectiveness of various forms of data-driven campaigning, but we can nonetheless make some preliminary observations on its implications for electoral competitiveness and on the broader question of whether data-driven campaigning is a revolutionary change from past practice.

The drive towards data is in part a reaction to the scarcity of other campaign resources, notably volunteers. Respondents in both the Liberal and Conservative campaigns noted a general decline in the number of volunteers for election campaigns over the last two decades. At the same time, respondents from all parties indicated that their parties had a renewed interest in volunteer-intensive forms of voter contact, notably door-to-door canvassing. (Bimber, 2014, notes a similar trend in American politics, likely in response to experimental findings indicating

that these have the highest impact on voter turnout; see Green and Gerber, 2008). To employ such tactics with a limited pool of volunteers requires an ability to target their efforts with as much precision as possible using either integrated or inferred voter data. Such data can also yield benefits in raising funds from small donors, as the successive Obama campaigns (Bimber, 2014) and the Conservative party (Flanagan, 2014) have discovered. Applying data about voters and about one's own campaign can also allow parties to campaign effectively even in areas where they have no local volunteers at all. We observed professional staff from outside the riding being dispatched to canvass communities they had never visited before, armed only with data about which houses to visit and which ones to skip. Regardless of the accuracy of that data, it enabled the staffers to take on a task that might have been seen as impossible, or at least highly uncomfortable, in the past. In all of these examples, data allowed parties to do more with less.

It is unclear, however, whether all campaigns should seek to become data-driven. Data is not cost-free; collecting and analyzing it requires the expenditure of either money or time, and the competitive advantages and efficiency gains (if any) may not be worthwhile for all campaigns. For a campaign with very modest financial and volunteer resources, like the NDP campaign we observed, proceeding by intuition and rules of thumb may be a better strategy if it allows more resources to be allocated towards other campaign activities. The Conservative campaign, meanwhile, is a cautionary tale in the opposite vein: Merely having access to vast troves of data about voters, meticulously built up over many electoral cycles, is no guarantee of victory. A campaign must use it to good effect, and even when such use is as basic as directing the GOTV effort on Election Day, voter data must be accurate if it is to be useful at all. Inaccurate data creates the illusion of certainty, which may be worse than uncertainty.

Moreover, even skillful application of data of any kind is seen even by professionals as offering only marginal gains. As one respondent put it, the job of a campaign is three-fold: raise the profile of the candidate; identify supporters; and turn out the vote on Election Day. "If you do a poor job of the first two", our respondent observed, "then you have no choice but to rely on analytics." It seems unlikely that the Liberal party's victory either locally (winning 55% of the vote) or nationally was entirely due to data-driven campaigning. Did the party succeed due to more traditional political forces such as opponent unpopularity, strategic voting and leadership style or message salience? If, as Flanagan observes, elections are won at the margins, then one might expect that intensive use of data would allow parties to eke out victories in a few close ridings (Flanagan, 2014). How much, however, does this matter when weighed against the many other factors that shape electoral outcomes?

This brings us to the larger question with which we began this project: is data revolutionizing campaigning at the local level in Canada? We remain skeptical about the efficacy of data-driven campaigning because in the Canadian context, elections take place in constituencies that are significantly smaller than the scale upon which the original innovators, the Obama campaign staff, were working. The scope of US presidential campaigns is such that probabilistic modelling becomes the only way that campaigns can attempt to "know" all voters; in a typical Canadian riding of under 120,000 people, on the other hand, this is not only possible but the main goal of any campaign. Moreover, some of the uses of data are unnecessary: the Obama campaign expended great effort in tailoring messaging to particular segments of the population in particular areas and determining on which channels and when to best buy ads for those areas. This is hardly necessary in a riding like ours with only one local radio station and

no local television. Thus, data can help Canadian campaigns overcome deficiencies in volunteers or time, but the effort and finances involved might be better allocated elsewhere; moreover, no amount of predictive modelling will actually persuade voters. That requires the same sort of direct connection and community engagement as seen in Carty's old-style politics of Tecumseh Corners (Carty, 2002). We do not believe it is a coincidence that the woman who won the 2015 election in our riding was the longtime mayor of the largest community, well known and respected for her previous work and well-connected to the political elite there. Thus, while data is clearly a useful tool for Canadian campaigns, it is unclear whether data-driven campaigning actually represents a major departure from the past and, thus, a revolutionary change. Indeed, given what we observed about the increased power of the central organization at the expense of the local grassroots, we think that if anything, it might be better described as a counter-revolution.

Works Cited:

- Barocas, Solon. "The Price of Precision: Voter Microtargeting and Its Potential Harms to the Democratic Process" *PLEAD'12* - Proceedings of the first edition workshop on Politics, elections and data. Maui, Hawai'i, USA. November 2, 2012
- Bennett, Colin. "Trends in Voter Surveillance in Western Societies: Privacy Intrusions and Democratic Implications." *Surveillance and Society*, 13 (3/4): 370-384, 2015
- Bimber, Bruce. "Digital Media in the Obama Campaigns of 2008 and 2012: Adaptation to the Personalized Political Communication Environment." *Journal of Information Technology & Politics* 11, no. 2 (April 3, 2014): 130–50. doi:10.1080/19331681.2014.895691.
- Carty, R. Kenneth. "Parties as Franchise Systems The Stratarchical Organizational Imperative." *Party Politics* 10, no. 1 (January 1, 2004): 5–24. doi:10.1177/1354068804039118.
- Carty, R. Kenneth, and William Cross. "Can Stratarchically Organized Parties Be Democratic? The Canadian Case." *Journal of Elections, Public Opinion and Parties* 16, no. 2 (July 1, 2006): 93–114. doi:10.1080/13689880600715912.
- Carty, R. Kenneth, and Munroe Eagles. "Do Local Campaigns Matter? Campaign Spending, the Local Canvass and Party Support in Canada." *Electoral Studies* 18, no. 1 (March 1999): 69–87. doi:10.1016/S0261-3794(98)00044-4.
- Carty, Roland Kenneth. "The Politics of Tecumseh Corners: Canadian Political Parties as Franchise Organizations." *Canadian Journal of Political Science/Revue Canadienne de Science Politique* 35, no. 04 (December 2002): 723–45. doi:10.1017/S0008423902778402.
- Coletto, David, Harold J. Jansen, and Lisa Young. "Stratarchical Party Organization and Party Finance in Canada." *Canadian Journal of Political Science/Revue Canadienne de Science Politique* 44, no. 01 (March 2011): 111–36. doi:10.1017/S0008423910001034.
- Delacourt, Susan. *Shopping for Votes: How Politicians Choose Us and We Choose Them*. 2nd Printing edition. Madeira Park, BC: Douglas & McIntyre, 2013.
- Eagles, Munroe. "Money and Votes in Canada: Campaign Spending and Parliamentary Election Outcomes, 1984 and 1988." *Canadian Public Policy* 19, no. 4 (1993): 432–49.
- . "The Effectiveness of Local Campaign Spending in the 1993 and 1997 Federal Elections in Canada." *Canadian Journal of Political Science/Revue Canadienne de Science Politique* 37, no. 01 (March 2004): 117–36. doi:10.1017/S0008423904040065.

- Flanagan, Tom. Harper's Team: Behind the Scenes in the Conservative Rise to Power. McGill-Queen's University Press: 2007
- Flanagan, Tom. *Winning Power: Canadian Campaigning in the Twenty-First Century*. McGill-Queen's Press - MQUP, 2014.
- Franz, Michael M. "Targeting Campaign Messages: Good for Campaigns but Bad for America?" ch 8 pp 113 to 131 in Ridout, Travis ed New Directions in Media and Politics Routledge: 2013
- Germany, Julie Barko. "The Online Revolution" ch 8 pp 147-160 in Johnson, Dennis W. Ed Campaigning for President 2008: Strategy and Tactics. New Voices and New Techniques Routledge: 2009
- Germany, Julie "Advances in Campaign Technology" ch 5 pp 81-91 in Johnson, Dennis W. Campaigning for President 2012: Strategy and Tactics. Routledge: 2014
- Green, Donald P. and Alan S. Gerber. Get out the Vote: How to Increase Voter Turnout (2nd edition) Brookings Institution: 2008
- Hardy, Bruce W., Chris Adasiewicz, Kate Kenski, and Kathleen Hall Jamieson. "Spending Differences and the Role of Microtargeting in the 2008 Campaign". Paper presented to the 2010 annual convention of the American Political Science Association.
- Issenberg, Sasha. *The Victory Lab: The Secret Science of Winning Campaigns*. Crown, 2012(a).
- Issenberg, Sasha. "How Obama Used Big Data to Rally Voters." *MIT Technology Review*, December 16, 2012 (b).
<https://www.technologyreview.com/s/508836/how-obama-used-big-data-to-rally-voters/>.
- Jamieson, Kathleen Hall. "Messages, Micro-targeting, and New Media Technologies" *The Forum*, Vol. 11 No. 3, pp 429-435, 2013
- Johnson, Dennis W. Campaigning in the Twenty-first Century: A Whole New Ballgame? Routledge: 2010
- Karpf, David. "The Internet and American Political Campaigns" *The Forum*, Vol. 11 No. 3, pp 413-428, 2013
- Mayer-Schonberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Reprint edition. Boston: Eamon Dolan/Mariner Books, 2014.
- Medvic, Stephen K. "Campaign Management and Organization: The Use and Impact of Information and Communication Technology." In *New Directions in Campaigns and Elections*. Routledge, 2011.
- Nickerson, David W., and Todd Rogers. "Political Campaigns and Big Data." *The Journal of Economic Perspectives* 28, no. 2 (2014): 51–73.
- Norquay, Geoff. "Organizing without an Organization: The Obama Networking Revolution." *Policy Options*, October 2008.
- Ridout, Travis N., Michael Franz, Kenneth M. Goldstein, and William J. Feltus. "Separation by Television Program: Understanding the Targeting of Political Advertising in Presidential Elections" *Political Communication*, Vol. 29, pp 1-23, 2012
- Rubenstein, Ira S. "Voter Privacy in the Age of Big Data" *Wisconsin Law Review* p. 861, 2014