From Stories to Evidence: How Mining Data Can Promote Innovation in the Nonprofit Sector

Michael Lenczner and Susan Phillips

> " *Endlich ist die große Ungewißheit aller Datis im Kriege eine* "
> *eigentümliche Schwierigkeit, weil alles Handeln gewissermaßen*
> *in einem bloßen Dämmerlicht verrichtet wird, was noch dazu*
> *nicht selten wie eine Nebel- oder Mondscheinbeleuchtung den*
> *Dingen einen übertriebenen Umfang, ein groteskes Ansehen gibt.*
>
> (The great uncertainty of all data in war is a peculiar difficulty,
> because all action must, to a certain extent, be planned in a
> mere twilight, which in addition not infrequently – like the
> effect of a fog or moonlight – gives to things exaggerated
> dimensions and unnatural appearance.)
>
> Carl von Clausewitz
> Prussian General and Military Theorist (1780–1831)
> in *Vom Kriege* (*On War*)

Being a director at a nonprofit organization often means making guesses instead of properly informed decisions. One source of the "information fog" is fragmented funding. Nonprofit organizations have multiple types of funders, most of whom are not their direct beneficiaries. Predicting funder behaviour is therefore more of an art than a science. Planning for the future, setting goals, and making decisions all suffer in the nonprofit sector because of a lack of timely and accurate information.

This article examines the opportunities to use newly available digitized information to address this information deficit. It shows how the rich, variegated and fast-changing landscape of information available online can be collected, combined, and repurposed in order to deliver it in actionable forms to decision makers across the nonprofit sector. This information can significantly improve planning decisions and enhance the effectiveness of the sector.

The article concludes that a cultural shift is required in order for the nonprofit sector to exploit the opportunities presented by digital information. Nonprofits and funders are enjoined to increase their numeracy and to find creative ways to use data as part of their evaluation, planning and decision making. Researchers need to be adventurous in their use of quantitative information and specifically should employ linked datasets in order to explore previously unanswerable research and policy questions. The producers of data need to collect and publish their information in ways that facilitate reuse. Finally, funders need to support a variety of projects that seek to exploit these new opportunities.

How Mining Data Can Promote Innovation in the Nonprofit Sector

Michael Lenczner and Susan Phillips

Introduction

Carl von Clausewitz, a Prussian general of the 19th century, coined the term "fog of war" (tinyurl.com/28qrg7q). The term refers to the difficulty a military commander faces when obliged to make a decision with information that is incomplete and not validated. We believe that it is not an exaggeration to say that this is the habitual situation faced by directors of nonprofit organizations, as well as by their funders, as they try to accomplish their missions in sustainable ways.

Being a director at a nonprofit organization (NPO) often means making guesses instead of properly informed decisions. One cause of the information fog is fragmented funding. NPOs are usually reliant on a multiplicity of funding sources, including government programs (at federal, provincial, and regional/municipal levels), donations from individuals, corporate sponsorships, and support from private and public foundations. The challenge of planning for the future is exacerbated by the difficulty of predicting the priorities and behaviour of funders, especially considering that government funding partners, frequently a significant source of revenue, always have a real chance of changing in the next election. The instability and competitiveness of financing means that an increased ability of a NPO to achieve its mission does not necessarily result in increased funding. This forces many NPO leaders to split their attention between the mission of the organization and its funding, which hampers good, integrated decision making.

A second factor is the need to identify not only potential competitors but likely collaborators. Given the magnitude of the social innovations that many NPOs are trying to achieve, they increasingly need to actively collaborate with a wide range of other organizations, funders, and governments. Obtaining enough information to sort out which organizations, among the myriad of possibilities, are likely to be effective partners is a significant challenge. This challenge is magnified by a multiplicity of stakeholders who may have competing views of the inherent value and means of achievement of the NPO's mission, and all of whom expect to have current information on progress toward and impact of the mission (Tschirhart and Bielefeld, 2012; tinyurl.com/7wwdvkq). In an environment in which the legitimacy and activities of the charitable sector are being closely scrutinized – for efficiency, impact and political activity – the ability of NPOs to be more effective producers and consumers of quality information is paramount. The need to oper-

ate with better access and clarity of information applies equally to funders as they seek to deploy their limited resources in strategic ways with maximum benefit.

This article examines the opportunities to use newly available digitized information for more effective decision making in the nonprofit and charitable sector. It shows how the rich, variegated, and fast-changing landscape of information, often currently restricted to silos, can be collected, combined, and repurposed in order to present previously unavailable information to decision-makers across the sector, and thus greatly enhance the potential for social innovation.

Nonprofit Information Today

That good decisions and the ability to create solid plans for the future rely on good information is a central tenet of strategic management. Unfortunately, the types and sources of information available to Canadian NPOs are very limited. The main source of information gathered on and available to Canadian charities and public is the T3010 annual tax returns that are collected and published by the Canada Revenue Agency (CRA; cra-arc.gc.ca). This database is supplemented by a variety of specialized resources including the Canada Survey of Giving, Volunteering and Participating (CSGVP; tinyurl.com/6q4nopl), the Canadian Human Resource Council for the Nonprofit Sector (hrcouncil.ca), and websites of various umbrella organizations that represent different components of the sector. Although each of these is a useful resource for specific purposes, they are inadequate to address the challenges outlined above.

The information collected by the CRA illustrates some of the shortcomings of the existing data. Despite the scope of the CRA charity database, it was designed with internal CRA administrative and compliance purposes in mind, and so its application is limited. Although researchers can use this dataset to get a better high-level picture of the charitable sector (which only comprises half of the nonprofit sector), it is not particularly user friendly and does not help directors or funders in decision-making. Recent projects such as Place2Give (place2give.com), Open Charity (opencharity.ca), and Donate2Charity (donate2charities.ca) have undertaken to display this information with much better visualizations. In addition to better packaged T3010 data, Imagine Canada's new portal, CharityFocus (charityfocus.ca), encourages NPOs to provide narratives that "tell their stories" about their activities and impacts. Still, the core tax dataset is restrictive in that it provides only broad

How Mining Data Can Promote Innovation in the Nonprofit Sector

Michael Lenczner and Susan Phillips

descriptions of the human resources and financial activities of individual charities and can offer little insight into the networks, clients, collaborators, government partners, and funders of NPOs.

Beyond these traditional sources, there is a bewilderingly vast and diverse world of information about the sector that is becoming accessible. Everything from the annual general reports of rural volunteer centres to the public financial accounts of federal departments is a potential source of actionable intelligence. When any of the entities related to the nonprofit sector publishes its information online, that information becomes exploitable in new ways. For instance, many, if not most, funders now publish information online as to whom and what they fund. Their reports are mainly designed with print publication as the principal media, however, and the information comes in heterogeneous formats and layouts, at different time intervals, with different levels of detail, and is published under different licenses. There has been no effort to standardize these datasets – until the open data movement of the last few years.

The Possibilities of Digital Information

The impact of the Internet is due not only to the vast amount of information available and the speed at which it can be accessed, but also to the digital format of the information. Digital information possesses special qualities that enable machines to manipulate it easily in various ways. An example is that it is possible to download web documents in various formats (Adobe PDF, Microsoft Word, Text Format, etc.) and run a program that will return the number of words contained in each document. This task, which in the era of print publishing would require a person counting out every word, is now one that every high-school student takes for granted as being possible at the click of a button. The effect of these digital qualities is that information can be sliced, categorized, and linked to other pieces of data.

The digital qualities of data are exemplified and exploited by the open data movement (tinyurl.com/qdffwj). Open data is a new practice whereby governments publish their data online in formats and under licenses that encourage "repurposing." These data cannot include information about individuals so as not to compromise individual privacy. The formats can easily be treated by machines (think of Excel sheets as opposed to PDF reports). Both the datasets that have been "opened" by government as well as the rest of their online informa-

tion can be publicly – and for the most part, legally – accessed in ways not controlled by their creators. They are now being combined or "mashed up," and put to creative uses by social innovators with possibilities that are just beginning to be imagined (see Davies and Bawa, 2012: tinyurl.com/73t3nww; Sonvilla-Weiss, 2010: tinyurl.com/72e8f4f). Open data policies are not transformative on their own (Cole, 2012; tinyurl.com/7tuss52), but they point the way towards increased application of public information.

Many of the early experiments with aggregated government data were focused on local governments, often led by community activists with encouragement by municipal governments. These attempts aimed to promote citizen participation and community empowerment. Early projects include websites that combine and map crime and census data allowing people to compare neighbourhoods by a range of safety and socio-economic indicators. As a wider range of datasets are made available, a second generation of applications is starting to emerge beyond basic mapping. These applications use multiple sources of data and have a potential to offer more in-depth analysis of more complicated issues (Bhushan, 2012; tinyurl.com/7jus3ax).

It is now possible for "data entrepreneurs" to: i) find diverse online data, including non-government sources; ii) gather these data regardless of the initial intent of the publishers; and iii) combine, process, and apply them in legal ways never imagined by the publishers. A new Montreal-based company, Ajah, is at the leading edge of this emerging practice of data aggregation and repurposing for civil society, and serves as a good case study of the potential of open, combined data for nonprofits.

Ajah: Putting the Data to Work

Ajah (ajah.ca), which was founded in 2010 by one of the authors, Michael Lenczner (CEO), focuses on collecting information that is published online and transforming this information into useful services for Canadian NPOs. Ajah's primary service is an online research tool that helps NPOs identify possible funders, evaluate them, and determine how best to approach them. In order to successfully approach funders, NPOs need to acquire information about public and private foundations, government funding programs, as well as publicly and privately owned corporations. Ajah employs some of the online sources of information published by and about each of these types of entities.

How Mining Data Can Promote Innovation in the Nonprofit Sector

Michael Lenczner and Susan Phillips

Ajah uses the T3010 CRA tax files of Canada's 86,000 charities to extract detailed information about foundations and identify their funding recipients. The company collects information about federal and provincial funding through various sources, including the federal proactive disclosure reports and provincial reports and databases (e.g., the Alberta Lottery Fund, Ontario's Trillium Foundation, British Columbia's Government Gaming Report). Corporate donations are tracked through the automatic collection and parsing of annual general reports of charities, as well as through manual research.

The digital publication of information means that, for the most part, it does not need to be acquired and processed manually. Rather, it is possible to write scripts that fetch the files, extract information that these scripts have been told to expect in certain places (e.g., dates, organization names, partial or full addresses, grant amounts), and store the information in a database. In the same way that digital text files made counting words much easier by orders of magnitude, the "scraping" of these digital reports and storing them into a database allows for quick and powerful analysis. Mapping grant recipients and identifying patterns and trends become trivial tasks.

If there is additional funding information, such as the type (e.g., operational, capital costs), purpose (e.g., arts, culture, sports, environment), or duration (single year versus multi-year), it is possible to perform further analysis. It is also possible to overlay this information against an external data set, such as Statistics Canada demographic data or political ridings. This would make it possible to explore the relationships between funding and socio-economic indicators or the possibilities of political patronage, for instance.

However, the real potential lies in combining solitary sources of information and cross-referencing them. Information from multiple sources can be connected to a specific funder or NPO. Suddenly, it is possible to see a much more complete picture of an NPO when its T3010 return is linked to the information about its program descriptions and the grants it has received. Such information can be found in foundation, corporate, and government records, and scraped for content. Furthermore, it becomes possible to detect correlations between changes in the funding behaviour of a specific funder and changes in recipient charities by examining their economic profiles. The impact of funding cuts by a group of funders (e.g., federal funders) can be examined and contrasted with the corresponding beha-

viour of another group of funders (e.g., provincial funders or foundations). In time, it should even become possible to model the effects of different funding policies or economic events.

Because computer programs can be set to run automatically, a scraping program can be set to check every day for a new copy of a report or form. When new information is published it is automatically downloaded, compared against its previous version, and added to the database. This whole process occurs without any human intervention; intervention is only required when funders publish incorrect or insufficient information that a machine cannot properly categorize.

As a result of this process, an extraordinary database – the largest of its type in Canada – has been compiled on the nonprofit sector and it is automatically updated and properly cross-referenced with minimal human intervention. This database can be used to answer a wide variety of specific questions, or to create and power specific tools, such as Ajah's funder research service.

Opportunities for the Nonprofit Sector

Besides its usefulness to NPOs, this new world of data also creates opportunities for funders, policymakers, and researchers, presenting important opportunities for all to improve analysis, planning, and decision making.

In the case of charities, the clearest opportunity is to use this information to benefit resource planning, specifically the search for diversified, stable funding. There is a wealth of unexamined information about most charities' primary funding partners. If properly analyzed, it could give them the ability to identify new funders, better predict their behaviour, and make more robust resource plans. This information can also be used by directors to decide what programming to develop and at what scale, enabling them to avoid the mission drift that can occur from chasing the most obvious funding opportunities.

Although there is less information on social impact, there are significant opportunities for NPOs to improve their program evaluation and reporting of outcomes. Both funders and the broader public are looking for evidence that resources allocated to the nonprofit sector are having the desired impact, and they are actively seeking out such evidence. To respond to this demand, NPOs are trying to find new ways to tell their stories.

How Mining Data Can Promote Innovation in the Nonprofit Sector
Michael Lenczner and Susan Phillips

NPOs seeking to position themselves well could combine their qualitative stories with effective use of quantitative data – both their own and what is publicly available – to provide more complete and satisfying accounts.

By using digital data, funders have opportunities to improve their analysis and their decision making. Connected datasets allow funders to address a wide range of questions: the impact of their grants, how they fit into the funding landscape of a locale, or how best to leverage other funders. With a clearer picture of the revenues and financing mixes of its recipients, funders are empowered to make better decisions. Easily accessible data on who is funding what in a city or region might provide the impetus to advance the formation of regional networks and collaboration among funders that has been talked about for the last decade or so.

Equally significant opportunities exist for researchers and policymakers. In terms of scholarly research on the nonprofit sector, Canada lags behind the United Kingdom and particularly the United States where there is a cottage industry in analyses of the 990 form, which is the American equivalent of the T3010. With few exceptions, Canadian researchers are just beginning to use T3010 data, and the advent of digital data represents a leap forward. It presents opportunities for research that is informed by and can address real-world challenges, and that can be injected into policy and organizational decision-making in a timely manner.

Consider Ajah's recent research that addressed the question of how extensively Canadian charities use social media. Using conventional methods, researchers would have devised a phone or web survey, drawn a sample of a few hundred or even a few thousand charities, hoped for an adequate response rate of perhaps 30 percent, and analyzed the data with a small horde of research assistants, over the span of a several months. Instead, in partnership with the marketing agency Stephen Thomas, Ajah simply wrote a program that checked the websites of 22,000 charities in two days and identified their Twitter, Facebook, and other social media accounts. These data were then linked to the charities' T3010 financial information in order to permit the incorporation of an economic dimension into the analysis. The linked data will be made available to researchers and will be the subject of both academic papers and resources for community organizations later this year. A future version of this report could easily include the content of those social media accounts and allow for detailed content analysis.

What is Required

While the opportunities noted here would not require tremendous resources, it is unlikely that the stakeholders in the nonprofit sector will fully pursue them. Too few of them have either the awareness or the capacity to use quantitative data. In order for these opportunities to be realized, a moderate but real shift in culture is required.

On the whole, NPOs lack a capacity for numeracy. Also, many will protest that quantitative approaches have severe limitations. Bearing in mind the real limitations of the available quantitative data, there are still demonstrable advantages to employing such data. This is why Carleton University's new Masters in Philanthropy and Nonprofit Leadership (carleton.ca/sppa), to begin in spring 2013, includes a component on quantitative as well as qualitative research. This new Masters is the first program of its kind in Canada and it is intended to help nonprofit leaders be more strategic and innovative, which includes being good users of available research.

Funders also need to begin to employ better evidence in their decision making. They should model the impact of their decisions on the financial viability of recipients and the funders' networks, and use this to inform their decisions. Being better informed will permit funders to minimize the risks associated with more "creative," impact-oriented grantmaking (Anheier and Leat, 2006; tinyurl.com/cbq77r7) such as making larger grants to promote transformative and durable social innovation. Yet, many funders collect data only for accountability purposes, rather than learning, and, like NPOs, lack the skills and capacity to make good use of it (Hall et al., 2003; tinyurl.com/blsvppk). They, too, must take the requisite step of developing the capacity of being both good consumers and producers of data, and they need to use it strategically to improve their effectiveness.

Canadian researchers have not been particularly adventurous in developing large-scale, empirical analyses of the nonprofit sector. Researchers need to move aggressively towards employing quantitative data and be creative in finding data sources that contain information relevant to their topics. Linking disparate datasets – such as the matching of social media information with T3010 returns – allows researchers to analyze more complex problems. In addition, granting councils, foundations, and other interested parties have not supported or encouraged such research for any sustained period. Universities could be much more active collab-

How Mining Data Can Promote Innovation in the Nonprofit Sector

Michael Lenczner and Susan Phillips

orators with the sector in producing and using quality evidence and in providing training in the relevant skills. Internships that would allow graduate students to spend time working in NPOs could be specifically directed towards enhancing capacities for data gathering and analysis. "Executive-in-residence" opportunities could be hosted by universities for senior staff of NPOs or policy-makers to enable them to develop more creative uses of data for both organizational and policy making.

Finally, the producers of data about the nonprofit sector need to collect and publish their data in ways that facilitate reuse. These producers include governments at all levels, foundations, corporate donors, and NPOs themselves. Data collection and publication could be improved: errors in the data reduced and data published in easier-to-use formats, in non-aggregated or non-summary form, with explicit permission for reuse. For example, the International Aid Transparency Initiative (aidtransparency.net) is "a voluntary, multi-stakeholder initiative" that encourages international aid donors to publish their funding information in an agreed-upon format. Data producers should even go beyond facilitating access to encouraging re-use. For example, the New Zealand Charity Commission provides a public interface to query their database as well as sponsors a competition to make the best "mashups" or reuses of their data (tinyurl.com/bodkp8s).

There also needs to be an ongoing dialogue between data users and producers in order to identify areas where new datasets or modifications of existing datasets are required. An example of this type of collaboration is the newly formed T3010 User Group, which is composed of NPOs, academic researchers, and vendors to the nonprofit sector. Improvements in collection and publication should be encouraged in a similar manner to the Voluntary Sector Reporting Awards (tinyurl.com/65ljdeg) provided by the Queen's Centre for Governance to recognize excellence in transparency and good governance.

Conclusion

The technique of identifying, collecting, and connecting datasets relevant to the nonprofit sector, as employed by Ajah, cannot completely dispel the "fog" in which nonprofits and their partners operate. Measuring social impact is a complex challenge that will not be resolved in the near future. But, systematic use of funding data can provide the necessary information to illu-

minate the objectives and patterns of funders, thereby allowing charities to reduce the energy and guesswork of fundraising, and to more effectively plan their programs. If we are able to take advantage of the opportunities presented by this financing and other digital information, we can expect improved planning by NPOs, more informed decision-making by funders, and researchers and analysts furthering our knowledge of the sector. At least some patches of fog may be lifted.

Acknowledgements

About the Authors

Michael Lenczner is a Montreal-based entrepreneur who works in both the for-profit and nonprofit sectors. He has worked on community technologies since 2000, community data since 2003, and on open government data since 2005. He co-founded Ile Sans Fil, CivicAccess.ca, Montréal Ouvert, serves on the board of several nonprofits, and is a frequent partner in academic-community collaborations. He is the CEO and founder of Ajah, a company that builds online tools for the Canadian nonprofit sector.

Susan Phillips is Professor and Director of the School of Public Policy and Administration at Carleton University in Ottawa, Canada, where she has led the creation of Canada's first Masters in Philanthropy and Nonprofit Leadership. Internationally recognized for her research in this field, Susan is currently co-editing the Routledge Companion to Philanthropy, the first international handbook in philanthropy, and she is beginning a major SSHRC-funded comparative study of community foundations. She is a frequent advisor to nonprofits, foundations, and governments.