Evaluating the impact of open data websites

Eleonore Fournier-Tombs

Faculty of Information, University of Toronto

Knowledge Media Design Institute

Over the past few years, the steady increase in the number of government open data websites has led to a call for appropriate evaluation tools. While some (Noveck, 2009) have expressed optimism as to the potential of government open data, others (Coglianese, 2009; Hindman, 2009) have been more hesitant. This paper therefore aims to answer the following question: *how does one evaluate the success of open data websites in reaching democratic objectives?* In doing so, it explores past academic studies and examines the researcher's experience with interpretive inquiry. Using Data.gov as an example, it argues that survey-based research, a common tool in information systems analysis, may not be suited to open data websites. Instead, it suggests a content analysis methodology, which hopes to inform future research on the subject.

Data.gov was created in 2009 by the United States Government's Chief Information Officer, Vivek Kundra (Lakhani et al, 2010). The website, which serves as an index for data stored on hundreds of other government websites, allows any person to

access wholesale and retail government data. The initiative followed United States President Barack Obama's publicly available Memorandum on Transparency and Open Government, in which he outlined the three objectives of his administration's Open Government Directive – transparency, participation and collaboration (Obama, 2010). A frontrunner in technologies facilitating open government, Data.gov was created using an agile systems development method, unusual for a government initiative (Lakhani et al, 2010). It also allows citizens to download the data and create novel data mashups. The website has the potential not only to facilitate government-to-government and government-to-citizen communications, but also to foster citizen-led technological innovation. (Lagace, 2010).

Beth Noveck, Obama's former Deputy Chief Technology Officer, claims that open government technologies "have the power to make government better, democracy stronger and citizens more powerful" (Noveck, 2009). Others are more skeptical. In a discussion of the Obama administration's open government processes as a political strategy, which he terms "fish bowl" transparency, Coglianese (2009) argues that a greater understanding of these measures on the behaviour of government, policy and public perceptions of legitimacy is needed. He adds that by placing too many expectations on transparency and technology, the United States Government risks disappointing the American public and increasing its cynicism.

Because the open data concept is recent, there is very little literature on initiatives at a national level and even less about how technology can impact the success of these platforms. One notable exception is the Harvard Business School's case study of

Data.gov (Lakhani et al, 2010). However, studies abound on the process of technological change or design within social, political and organizational contexts.

Walsham's (2009) synthesized change framework provides an interpretive analysis of information systems design within the context of organizational change that allows the researcher to understand the content, social context, social process and context/process linkage in play and how these are impacted by a proposed technology. This lens allows for in-depth understanding of the design of an information system and how it will impact the context in which it lies. Walsham and Sahay's (1996) application of the synthesized change framework to the study of the implementation of geographical information systems in India for wasteland management suggests that this framework can be used to analyze government data systems. Walsham positions himself as a social construction of technology (SCOT) theorist by contrasting interpretivism to positivism, arguing that information systems design should be seen as a social construction, rather than an objective process (Walsham, 1995).

Pinch and Bijker (1984) take a social constructivism approach to both technology and science, particularly when it comes to innovation. Their discussion of the creation of the bicycle within its social context, particularly in relation to gender roles, has been instrumental in shaping Science and Technology Studies (STS).

Winner (1980) argues that technologies often have inherent social or political implications. He explains, for example, that Robert Moses, the architect of the bridges on Long Island, purposefully made them low to prevent buses from passing under them, making it nearly impossible for low-income African-American citizens to access to parks and beaches used by the rich. He implies that while there is always a social context to the

design process of a technology, there may also be social implications designed into the technology, that continue to operate after the original context is transformed. Winner thus criticizes SCOT theorists for placing too much importance on the social and economic context and neglecting the study of the technology itself.

In another critique of Pinch and Bijker, Russell (1986) argues that their relativist approach and their focus on the interaction of social groups with the technology hinder their useful discussion of the social context of technological development. He adds that relativism implies political neutrality, such that all constructs are equal. As Winner (1980) makes clear, this cannot be the case, as constructs are often embedded in technologies rather than context-dependent.

A methodological response to the challenges of adopting both an interpretive stance when examining social systems and a positivist one when analyzing the technology itself has been developed by Deibert and other researchers at the University of Toronto's Citizen Lab (Deibert and Rohozinski, 2010). In a study on security breaches into the Office of the Dalai Lama's information system, Deibert and Rohozinski use fusion methodology, which combines grounded field research, technical investigation and data mining, analysis and visualization with the objective of understanding Internet data within its societal context.

This approach fits within the growing trend of mixed methods approaches to social science research, as explained by Luker (2008). She notes that three trends have reduced the relevance of purely interpretive or positivist frameworks for social research. The rise in popularity of Foucault, the creation of an info-glut, due to the enormous amount of information available to researchers, and the decline in linear processes and

thinking, in part due to the hyperlinked structure of the Internet, all favor mixed methods research, in which the researcher does not commit to a particular research ideology but rather chooses the frameworks and methods that will best answer the research questions.

While Foucault creates a base from which researchers could examine science and technology as human phenomena, and STS relativists, such as Pinch and Bijker (1984), further his ideas by viewing all science and technology innovations as human constructs, post-positivists such as Latour (1999) and Kuhn (1996) argue that while consideration of human conjectures and interpretation is crucial, there remains an objective truth. It must be noted that Latour, after being construed as a science and technology relativist (Latour, 1999), wrote a strongly worded article in the University of Chicago's Critical Inquiry in which he deplored the lack of objectivity in scientific publications and reporting (Latour, 2003). He joins Winner (1980) in a criticism of what may be termed overly interpretive theory, arguing that understanding phenomena as social constructs should not prevent the acceptance of scientific facts. In particular, he deplores the adoption of interpretive concepts by mainstream media in public debate over the existence of climate change and notes that some critical theorists have paradoxically adopted the stance that everything is a social construct. This article clearly positions him within the post-positivist ideology, arguing for a consideration of human elements in the search for truth. Kuhn (1996) also calls for balance between relativism and positivism. In *The Structure of Scientific Revolutions*, he argues that between revolutions, the scientific community operates under the assumption that it has a correct view of the world. Innovation or discovery, while initially repressed because of its subversive effect on the status quo, is eventually integrated by the field as a new equilibrium is reached. This shows, he argues, the

arbitrary nature of the scientific community's understanding of the world, which should be taken into account in STS.

The development of approaches fostering a societal understanding of technologies has also been spearheaded by Internet and infrastructure ethnographers, particularly Star (1999). She posits that studying infrastructure with an ethnographic lens can yield valuable insight into its design within an ecological and relational context. Star's work provides specific questions, or lenses, for interpretation of an infrastructure, such as the United States Government's open data system.

Open data systems more broadly can be viewed within e-governance, which uses online platforms to facilitate flows of data and information between governments and citizens, between governments and businesses and within governments themselves (Hunter and Tan, 2007). As explained by Hogge (2010) in a report for the Open Society Institute, open data goes beyond government transparency and access to information legislation by ensuring that government data is highly accessible and usable for its citizens. The Open Knowledge Foundation (2010) defines open government data as "data and information produced or commissioned by government or government controlled entities" in such as way that it can be "freely used, reused and redistributed by anyone". The Foundation identifies three reasons for the promotion of open government data: transparency, release of social and commercial value, and participatory governance.

Three individuals who have greatly contributed to the open government policy debates in the United States are Ellen Miller of the Sunlight Foundation (2010), Beth Novek (2009) and Vivek Kundra (2008). In its Sunlight Agenda 2010, the Sunlight Foundation (2010) notes that transparency fosters civic participation and that a key

feature of this transparency is access to and accessibility of information. The Agenda adds that data that the government chooses to publish should be available online and in interoperable formats. The Foundation thus espouses Barack Obama's transparency agenda. Miller states that: "core to the President's campaign for government transparency is the use of technology in ways that redefine what 'public information' means – that is online information, information that is as easily searchable as it is easily accessible," (Lakhani et al, 2010, 3). She also adds that a key feature of the Data.gov website is the capacity it gives to citizens to create aggregate data that affects their daily lives and increase their civic participation.

In a similar vein, Noveck has written extensively on the benefits of open data technology for democracy. Her main claim is that online technologies have provided the opportunity for collaborative governance and innovation, one in which citizens participate through distributed, open source channels (Noveck, 2009). This model has also been espoused by Benkler (2003), who notes that online technologies enable distributed production, which enhance economic efficiency. Hindman (2009) notes that this model also applies to governance, and that online democratic participation can improve policy-making.

Finally, Kundra is widely considered to be the initiator of the American open data system (Lakhani et al, 2010) at a national level. In public appearances and government reports, he makes it clear that his priority is "to create a runway, a platform for innovation" (Lagace, 2010). Like Noveck, he values giving citizens the tools for innovation in technology and participation in policy. In doing so, he places emphasis on the quality of the platform and the data it hosts.

The discussion above points to a need for evaluation methods and suggests an interpretive approach similar to an ethnographic case study (Star, 1999; Walsham, 2009; Yin, 2003). However, as we will see below, such research necessitates a level of access (Bryman, 2008; Luker, 2008; Walsham, 2006) that does not correspond to the open and distributed user base for open data websites. In other words, it is very difficult to have access to open data users outside of the platforms in which they participate.

Nevertheless, it has been found (Bryman, 2008; Luker, 2008; Yin, 2003) that user surveys can be a useful data collection method for interpretive research as they can lend themselves to both quantitative and qualitative analysis.

In June 2011, the researcher constructed an online survey in order to gather data on user perceptions of the website. The survey respondents were recruited from online open data forums based in the United States such as: the Google group Citizens for Open Government and the Open Government Meetup. An introductory message was posted and respondents were encouraged to complete the survey questions. Since the online survey was anonymous, it was not possible to compensate the participants. As noted by Groves et al (2004), however, individuals are more likely to participate in surveys that are on a topic of interest to them. It was therefore hoped that respondents would be interested in contributing to an enhanced understanding of open data practices.

However, response rates were too low to extract noteworthy conclusions from the data. The many surveys of various open source community participants (Fang and Neufeld, 2009; Roberts et al, 2006) suggest that while it is possible to survey members of an online community, this may be more appropriate when the community is clearly defined. Many researchers have been faced with low response rates to web-based surveys

(Fan and Yang, 2010). Fan and Yang identify a number of factors influencing response rates, including sampling methods. They further argue that characteristics of respondents, such as self-selection and anonymity, make it difficult to identify what biases might be present in the sample. After this attempt at data collection, therefore, it was determined that another method should be used.

New technologies provide large amounts of publicly available data that, some authors suggest, lends itself well to content analysis. As early as the 1990s (Willhelm, 1998), researchers have examined the value of online forums to carry out democratic tasks, such as deliberation and decision-making (Wright and Street, 2007; Janssen and Raphael, 2005; Gonzalez-Bailon, 2010).

Lasswell (1969) argues that content analysis, "the systematic, empirical studies of the messages transmitted in a process of communication" (57), is a valuable instrument for evaluating policy alternatives. His questions for content analysts are still relevant today: "Who, with what intentions, in what situations, with what assets, using what strategies, reaching what audiences, with what result?" (62)

West (2005) uses content analysis to evaluate the success of government websites at facilitating government-citizen information flows. He examines a number of variables, such as the presence of online publications, adherence to usability guidelines, email responsiveness and readability.

Other authors, such as Gonzalez-Bailon (2010), have used data produced through online discussions to evaluate the contribution of the platform to political deliberation. She proposes an instrument that measures the depth and breadth of the political

discussions on the social networking site Slashdot and notes a difference in the structure of discussions of a political nature compared to other types of discussions.

Although Data.gov has produced a number of offshoot online and in-person events and platforms that could be examined, the site itself also provides several data types that could be used in content analysis. This paper therefore proposes the combination of website content analysis and social networking forum discussion analysis in order to evaluate the effectiveness of Data.gov in reaching its objectives.

The following types of user-contributed data are available on the Data.gov website: forum posts; blog posts; blog post comments; raw dataset activity (community ratings of datasets, rater profiles, number of visits, number of downloads, comments about the datasets, contributor profiles); and interactive dataset comments.

Comments on the interactive datasets, for example, may provide insight into concerns that the users might have with the data. On a Survey of Patients' Hospital Experiences, for example, *Dziok* writes: "What's the timeframe for this data? Is it a continuous update? If it is how far behind does it lag? And where does the dataset begin?" In relation to the same dataset, *Andy* writes: "Response rates are very low for some surveys. Users should be aware of bias. I have observed some correlation between survey ratings and response rates." These two comments may be representative of a knowledgeable and engaged interaction with the datasets.

Similarly, under a forum topic entitled: "What datasets are you using or wish were available?", several users have explained how they use the data available on the site. *Elena Rapisardi*, an emergency management worker in Italy, writes: "What I realized is that it could be so incredibly difficult to gather and collect data such as: forest fires,

floodings, structures. […] During an emergency event, as I experienced during the last earthquake in my country, to have data and to have not [sic] can make the difference."

Although comments on Data.gov tend to be few and far between, they may yield information about the types of user that will engage with the datasets, as well as their needs and satisfaction levels.

As we have seen, a contextual and interpretive examination of the impact of government open data websites on democratic objectives has the potential to provide insights that could improve their design and thus strengthen democratic engagement. However, this study finds that user surveys, although relevant to many forms of information systems analysis, may not be the most appropriate data collection method, due to difficulties in respondent participation and sampling. Data.gov does provide platforms for user-generated content and discussions that may show user perceptions and use of the data. This data seems to lend itself well to content analysis, which has been used for both website and online deliberation analysis. Future research using publicly available data to evaluate open data websites will show whether this is an appropriate evaluation tool or if different types of data are needed. It is hoped that these findings will inform both academic and professional studies on open data websites and ultimately strengthen the quality of democratic engagement possible through new technologies.

References

Benkler, Y. (2003). Freedom in the commons: Towards a political economy of

information. Duke Law Journal, 52(6), 1245-1277.

Bryman, A. (2008). Social research methods. Oxford: Oxford University Press.

Coglianese, C. (2009). The transparency president? the Obama administration and open

government. Governance, 22(4), 529-544.

Deibert, R. J., & Rohozinski, R. (2010). Risking security: Policies and paradoxes of

cyberspace security. International Political Sociology, 4(1), 15-32.

Fang, Y. & Neufeld, D. (2009). Understanding sustained participation in open-source

software projects. Journal of Management Information Systems, 25 (4). 9-50.

Fan, W. & Yang, Z. (2010). Factors affecting response rates of the web survey: a

systematic review. Computers in Human Behavior, 26. 132-139.

Gonzalez-Bailon, S. (2010). The structure of political decision networks: a model for the

analysis of online deliberation. Journal of Information Technology, 25(2). 230-243.

Groves, R. et al. (2004). The role of topic interest in survey participation decisions.

Public Opinion Quarterly, 68(1). 2-31.

Her Majesty's Government, United Kingdom. (2010). Data.gov.uk. Retrieved October 16,

2010, from http://data.gov.uk/

Hindman, M. (2009). The myth of digital democracy. Princeton: Princeton University

Press.

Hogge, B. (2010). Open data study. Open Society Institute, Transparency and

Accountability Initiative, May 2010, October 16, 2010. Retrieved from

http://www.soros.org/initiatives/information/focus/communication/articles_publicati ons/publications/open-data-study-20100519

Hunter, M. G., & Tan, F. B. (2007). Strategic use of information technology for global organizations. Hershey: IGI Publishing.

Janssen, D. & Raphael, K. (2005). Online forums and deliberative democracy. Acta Politica, 40(3). 317-335.

Kuhn, T. S. (1996). The structure of scientific revolutions (3rd ed.). Chicago: University of Chicago Press.

Kundra, V. (2008). Creating the digital public square. Slideshare.Com. Retrieved from http://www.nascio.org/events/2008Annual/presentations/Kundra.pdf

Lagace, M. (2010). Data.gov: Matching government data with rapid innovation Harvard Business School Working Knowledge. Retrieved from http://hbswk.hbs.edu/item/6423.html

Lakhani, K., Austin, R., & Yi, Y. (2010). Data.gov. Harvard Business School, (May 13)

Lasswell, H. (1969). The uses of content analysis data in studying social change. Social Science Information, 7(1). 57-70.

Latour, B. (2004). Why has critique run out of steam? From matters of fact to matters of concern. Critical Inquiry, 30(2) Retrieved from http://criticalinquiry.uchicago.edu/issues/v30/30n2.Latour.html

Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford: Oxford University Press.

Luker, K. (2008). Salsa dancing into the social sciences: Research in an age of info-glut. Cambridge: Harvard University Press.

Noveck, B. S. (2009). Wiki government: How technology can make government better, democracy stronger, and citizens more powerful. Washington: Brookings Institution Press.

Obama, B. (2010). Transparency and open government: Memorandum for the heads of executive departments and agencies. Retrieved December 6, 2010, from http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/

Open Knowledge Foundation. (2010). Website. Retrieved December 6, 2010, from http://okfn.org/

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. Social Studies of Science, 14(3), 399-441.

Roberts, J. et al. (2006). Understanding the motivations, participation and performance of open source software developers: A longitudinal study of the Apache projects. Marshall School of Business Working Paper No. IOM 01-06.

Russell, S. (1986). The social construction of artifacts: A response to Pinch and Bijker. Social Studies of Science, 16(2), 331-346.

Star, S. L. (1999). The ethnography of infrastructure. American Behavioral Scientist, 43(3), 377.

Suchman, L. (1994). Do categories have politics? Computer Supported Cooperative Work (CSCW), 2(3), 177-190.

Sunlight Foundation. (2010). Sunlight agenda 2010. Retrieved December 6, 2010, from http://sunlightfoundation.com/policy/documents/agenda/

United States Government. (2010a). Data.gov. Retrieved October 16, 2010, 2010, from

    http://www.data.gov/

West, D. (2005). Digital government: technology and public sector performance.

    Princeton: Princeton University Press.

Walsham, G. & Sahay, S. (1996). Implementation of GIS in India: Organizational issues

    and implications. International Journal of Geographical Information Systems, 10(4),

    385-404.

Walsham, G. (2006). Doing interpretive research. European Journal of Information

    Systems, 15(3), 320-330.

Walsham, G. (2009). Interpreting information systems in organizations. Global Text.

Wilhelm, A. (1998). Virtual sounding boards: How deliberative is online political

    discussion? Information, Communication & Society, 1(3). 313-338.

Winner, L. (1980). Do artifacts have politics? Daedalus, 109(1), 121-136.

Wright, S. & Street, J. (2007). Democracy, deliberation and design: the case of online

    discussion forums. New Media and Society, 9(5). 849-869.

Yin, R. K. (2003). Case study research: Design and methods (3rd ed.). Thousand Oaks:

    Sage Publications.