When Transparency and Collaboration Collide: The USA Open Data Program

Alon Peled
Department of Political Science and Public Administration, The Hebrew University, Jerusalem, Israel.
E-mail: Alon.Peled@post.harvard.edu

President Obama’s inaugural flagship Open Data program emphasizes the values of transparency, participation, and collaboration in governmental work. The Open Data performance data analysis, published here for the first time, proposes that most federal agencies have adopted a passive–aggressive attitude toward this program by appearing to cooperate with the program while in fact effectively ignoring it. The analysis further suggests that a tiny group of agencies are the only “real players” in the Data.gov web arena. This research highlights the contradiction between Open Data’s transparency goal (“All data must be freed”) and federal agencies’ goal of collaborating with each other through data trade. The research also suggests that agencies comprehended that Open Data is likely to exacerbate three critical, back-office data-integration problems: inclusion, confusion, and diffusion. The article concludes with a proposal to develop an alternative Federal Information Marketplace (FIM) to incentivize agencies to improve data sharing.

On August 29, 2005, Hurricane Katrina made landfall near the Louisiana–Mississippi border. Despite early warnings, Katrina was the most destructive hurricane ever to hit the United States, with 1,250 casualties and damages estimated at $125 billion. The Congressional Bipartisan Select Committee appointed to investigate the preparation and response to Katrina was troubled by the lack of information sharing among federal and state agencies in the aftermath of the hurricane. Agencies were found to have made duplicative plans to evacuate the Superdome (where thousands of survivors found temporary shelter), resulting in a 24-hr delay to the evacuation. Likewise, agencies directed survivors to hospitals that were already overcrowded with patients sent by other agencies. Desperate survivors could not find comprehensive information regarding when, how, and to whom emergency supplies were being delivered. Life-sustaining generators were housed in lower floors susceptible to flooding because emergency plans containing instructions for their relocation in case of flooding were not publicized. Exasperated, the Select Committee (2006, p. 1) noted that the federal government remained the largest purchaser of information technology in the world and yet was “woefully incapable of storing, moving and accessing information.” The Select Committee then added:

Better information would have been an optimal weapon against Katrina. Information sent to the right people at the right place at the right time. Information moved within agencies, across departments, and between jurisdictions of government as well. Seamlessly. Securely. Efficiently… One would think we could share information by now. But Katrina again proved we cannot. (p. 1)

These strong words serve as a reminder that information-sharing problems continue to haunt the American federal government. Why exactly are federal agencies so reluctant to share their electronic data, and how can federal electronic data sharing be improved?

President Obama’s Open Government Program proposed one solution to this problem. The Open Government Directive (OGD) ordered agencies to publish all nonclassified datasets on the web; however, an analysis of the Open Data program’s performance, published here for the first time, suggests that most agencies effectively ignored the program. It is suggested that President Obama’s blitzkrieg transparency campaign collided with existing interagency collaboration practices through data trade. The article concludes with a suggestion to strengthen such practices via the erection of a formal Federal Information Marketplace (FIM).

Open Data’s Blitzkrieg Campaign to Free Federal Data

While campaigning, Obama promised to reverse the post-9/11 “retreat from openness” (Roberts, 2006, p. 19). During the 77 days between election day and inauguration day, the Obama–Biden transition crew commissioned a team to prepare the Open Government initiative. This team identified organizations that were willing to support a transparency
Agencies had much to lose in this openness onslaught. Officials feared that the new transparency policy would be costly, discourage their troops from holding a frank internal discussion, and impact their ability to utilize data as bargaining chips with other agencies (Coglanese, 2009; Meier & Bohte, 2007).

Open-government architects then launched a blitzkrieg-style campaign to enforce openness on agencies. On his first full day in office (January 21, 2009), at the height of the worst economic crisis America had experienced since the Great Depression, President Obama signed three memorandums and two executive orders. Four of these five documents promoted open government (White House, 2009). In March 2008, Vivek Kundra was appointed as the first federal Chief Information Officer (CIO). In early 2009, Washington’s rank-and-file bureaucrats were invited to provide the new administration with direct input (instead of commenting via their agencies). On March 19, 2009, Eric Holder, the Attorney General, reversed the Ashcroft memorandum of October 12, 2001, which advised agencies to employ caution in cooperating with the Freedom of Information Act (FOIA), declaring that “in the face of doubt, openness prevails” (Holder, 2009, p. 1). A barrage of open-government initiatives surfaced, including eRulemaking, IT Dashboard, Recovery.gov, and USAspending.gov. The administration showcased open-government stories and ensured that senior appointees adhered to open-government principles (Millar, 2011).

On May 21, 2009, the 120-day anniversary of the President’s Open Government Memorandum, a team headed by the CIOs of both the Department of the Interior and the Environment Protection Agency (EPA), launched www.data.gov (Office of Management and Budget, 2009). The architects of data.gov aspired to make this site much more than a new online index. Rather, they were determined to make data.gov the premier web-publishing location for all important federal datasets. Then, on December 8, 2009, the Office of Management and Budget published the OGD. The OGD specifically demanded that agencies “identify additional high value information not yet available and establish a reasonable timeline for publication online in open formats with specific target dates” (Office of Management and Budget, 2009, p. 7). However, the OGD did not offer guidelines regarding the definition of a “dataset” and, as shown in the next section, agencies quickly learned how to “slice and dice” their databases to appear as if they were in compliance with the OGD.

Through these executive acts, senior-level appointments, and positive government rhetoric, Open Data grew from an abstract idea to a powerful doctrine. Additional Open Data supporters joined the government including Beth Simone Noveck, a law professor and social networking entrepreneur, who was appointed Deputy Chief Technology Officer of the White House Office of Science and Technology Policy. Noveck (2009) argued that common people’s expertise combined with the limitless potential of social networking would create a new federal knowledge ecosystem. For Open Data enthusiasts technology is the solution to any problem: For example, when the U.S. Government Accountability Office (GAO) (2010b) noted that data on data.gov was growing stale, Kundra promised more technology to resolve the problem.

Open Data architects created a public-relations buzz that continues to characterize the program to date. For example, they announced that all agencies succeeded in releasing high-value datasets by the designated deadline (Data.gov, 2010). Nonetheless, a consortium of volunteers conducted its own evaluation and discovered that the initial dataset submissions were a “mixed bag,” meaning that they contained both high-value and non-high-value datasets (McDermott, 2010, p. 402). More recently, Open Data architects have highlighted the international open data movement that data.gov has spawned (Kundra, 2011). Scholars uncritically published facts released to them by Open Data (Lakhani, Austin, & Yi, 2010). Although a handful of scholars did predict that the Open Data program would not achieve its goals, they lacked sufficient empirical data to support this claim (Coglanese, 2009; Harper, 2010; Hendler, 2009, 2010).

In addition, Open Government proponents introduced evaluation of an Open Government project, the “Open Government Dashboard,” based on a self-assessment questionnaire completed by the agencies. Unsurprisingly, agencies awarded themselves with high marks despite significant discrepancies in degrees of compliance with the assessment criteria. A comparison of the self-assessments of two agencies illustrates this inconsistency. Both the U.S. General Services Administration and the Department of Veteran Affairs (VA) awarded themselves high marks on the questionnaire for meeting project expectations. However, to meet project expectations, the General Services Administration created an elaborate website with hundreds of new web pages and documents available via the site. The VA “complied” with the OGD’s dictates by indexing three PDF documents and two links to older information that existed on the VA’s website prior to the launch of the Open Data program. None of the agencies awarded themselves red (the lowest) “Fails to Meet Expectations” marks.
FIG. 1. Growth of indexed high-value datasets and downloads to/from open data.

It has been challenging to develop an independent method to empirically measure the impact of Open Data. A single scholar examined the first batch of released datasets and concluded that there was little high-value data in these datasets; however, since this early study, not a single empirical study of Open Data has been published (Harper, 2010). Data.gov did not respond to inquiries to release data about itself; however, it was possible to measure how agencies truly responded to the program by downloading some performance data that Open Data has published (data.gov/metric). Anticipating that this data may one day be removed, the research team downloaded weekly and during 1 full year (May 16, 2010–May 8, 2011) “snapshots-in-time” of the performance data. This study therefore is the first quantitative assessment of data.gov’s performance (McDermott, 2010, p. 402).

Open Data in Numbers (May 2010–May 2011)

An initial empirical examination supports Open Data’s assumptions. Open Data architects successfully recruited agencies to index data via data.gov (during the surveyed 52 weeks, the number of participating agencies grew from 105–169). Over the same period, participating agencies more than doubled the number of datasets that they indexed via data.gov (3,252 datasets in May 2011 as compared to only 1,283 datasets in May 2010). The public registered over 1-million downloads of datasets. Logarithmic panel regression analysis provides additional support for Open Data’s fundamental assumption: An increase by 1% in the number of indexed datasets is correlated with an increase in the number of downloads.

Yet, a closer examination reveals a different story. Figure 1 presents Open Data’s history in pictorial form, and shows that most agencies met the minimal OGD requirements by publishing a small number of datasets (the dotted line; the scale of 0–2,000 measures the aggregated number of high-value datasets agencies indexed on Open Data’s website) and did virtually nothing thereafter. The general public consequently discovered that data.gov is of little value and the number of downloads from data.gov did not significantly increase (the solid line; the scale of 0–60,000 measures the weekly number of downloads of all datasets. Note that without direct access to the raw Open Data weblogs, it is impossible to determine whether these downloads were carried out by government employees or private individuals). Although there is no objective reference point to determine if these download numbers are relatively low or high, note that Open Data architects themselves recently expressed disappointment regarding the general public’s response to data.gov. For example, Noveck (2011), at the conclusion of her 2-year White House term, called on “mediators” and “interpreters” with data-analysis skills to step forward and help the public discover the value of governmental data. She also proposed to educate citizens regarding the economic opportunities buried in the heaps of newly freed governmental data, contrary to the “give them data and they will come” belief that characterized the earlier days of the Open Data program (Millar, 2011; Noveck, 2011).

A further examination of the spike in the number of high-value datasets in September 2010 also supports the argument that most agencies provided minimal data to data.gov. The spike is principally accounted for by an increase in mid-September 2010 in EPA high-value datasets from 330 to 1,729. These datasets (and their corresponding downloads) comprised numerous Toxics Release Inventory reports, each reporting data for a single state and a single year (from...
1986–2010), which the EPA had been publishing on its own website since 1998. In addition, the EPA’s CIO was one of the two CIOs placed in charge of launching data.gov, providing further motivation for the EPA to support the project. In fact, since its inception, data.gov has grown to become the playground of a tiny group of agencies. Kundra (2011) emphasized that during its first 2 years, data.gov had grown from 47 to 389,681 datasets and applications. However, in May 2011, only 5 of 169 participating agencies accounted for 99.37% of all datasets and applications (378,488 of 380,879, as measured on May 8 2011), as shown in Figure 2. Almost all dataset uploads were geographical applications (98.90%, or 376,678 of 380,879 datasets). These geographical applications were data maps that existed, in one form or another, on the CENSUS website, the U.S. Geological Survey (USGS) website, and the National Oceanic and Atmospheric Administration (NOAA) website, before data.gov was launched. Moreover, the CENSUS and the USGS remain the only two agencies displaying a “negative ratio” on indexed datasets and geographic applications to number of downloads. The CENSUS, for example, has indexed almost 14 times more datasets and geographical applications than the number of downloads received via data.gov. This is likely due to the fact that the public can find similar data elsewhere for free (e.g., via Google maps) or can directly access the data on the websites of the CENSUS, the USGS, and the NOAA. Even excluding these geographical postings, data.gov remains the domain of a very small number of agencies. Five agencies jointly command 58% of all indexed nongeographical datasets (with the EPA alone owning more than 47% of all nongeographical indexed datasets). Likewise, four agencies jointly possess 58% of all nongeographical applications indexed on data.gov. The same is true for high-value datasets (i.e., the EPA alone commands almost 70% of all these datasets). Similarly, four agencies own over 56% of all high-value applications. The small club of active data.gov players includes the EPA, the CENSUS, the USGS, the NOAA, and to a lesser extent, NASA and the Department of Defense (DOD). Moreover, almost all agencies were equally unlikely to update previously uploaded datasets or to upload new datasets. Figure 3 measures the median number of idle days since an agency last updated an existing dataset or uploaded a new dataset. The y axis measures the number of days since an agency last updated data or uploaded new datasets. The strong linear growth displayed in Figure 3 shows that agencies increasingly declined to update datasets and to upload new datasets. In fact, as of May 2011, the average participating agency had not returned to data.gov for 222 days since its last data.gov transaction. In fact, almost 30% of all agencies (50 of 169) have remained idle in data.gov for 1 full year or more.

Jointly, Figures 1, 2, and 3 suggest that only a handful of agencies remain active in data.gov and that the average agency did the minimal required to qualify as a “participating [Open Data] agency” and shortly afterward stopped cooperating with the Open Data program. The trends described earlier have intensified in recent months, evidenced by the fact that in the beginning of 2011, Open Data architects further lowered the “participating agency” requirements. On average, the 42 agencies that joined data.gov between May and December 2010 published seven times more datasets on their first appearance in data.gov than did the 22 participating agencies that joined data.gov between January and May 2011. Most important, about 80% of all agencies (no matter when they joined...
Federal Data as the Ultimate Source of Agencies’ Political Power

Why did significant political pressure fail to convince most federal agencies to publish more data on the web? Why did President Obama not succeed in mobilizing the vast political power, energy, and good will that he commanded at the beginning of his presidency to compel agencies to publish more data on the web? Scholars who have been writing about this topic for almost 4 decades now have argued that computing resources (including electronic datasets) are inextricably tied to political power struggles between bureaucratic agencies and that this fight is fueled by three primary factors. According to this line of argument, a large bureaucracy such as the American federal government contains numerous departments and agencies, each with its own agenda and objectives, competing over resources, influence, and autonomy. Departments and agencies battle over what kind of computing equipment to acquire, what datasets to build, and how to standardize and regulate access to these computing resources. Agencies fight so ferociously over computer resources for three reasons. First, the information contained within computers often determines which organizational factions will gain or lose power relative to others. Second, computing infrastructure is expensive, and therefore those who control it govern a considerable portion of the organizational resources. Third, many people perceive those who are engaged in computing as sophisticated and professional; hence, computing brings “effective power” to those who command it. For example, bureaucrats know when and how to release data, stories, and forecasts as part of their efforts to increase their agencies’ autonomy and reputation (Hilgartner, 2000, p. 20; Krause & Douglas, 2005, p. 23; Moffitt, 2010; Rourke & Edward, 1961, p. 127; Schneider & Ingram, 1990, pp. 517–519; Stone, 2002, pp. 305–323; Weiss & Gruber, 1984).

Scholars have further argued that computers reinforce the existing organizational status quo by providing already powerful departments and agencies with opportunities to decide how much to invest in computing, how to control access to data, and which priorities to promote while developing new computer systems and databases. Frequently, powerful bureaucrats who are not technically skilled manipulate the acquisition and application of technical skill within their organizations to ensure that computer technology operates on behalf of a predetermined agenda. Clumsy systems and inaccessible databases are sometimes produced, modified, and “enhanced” due solely to the fact that their outputs stand to increase the political power of already powerful departments and agencies (Danziger, Dutton, Kling, & Kraemer, 1982; Danziger & Kraemer, 1986; Dunlop & Kling, 1991; Dutton & Kraemer, 1977; King, 1983; Kling, 1980, 1983; Kling & Iacono, 1984, 1989; Kling & Scacchi, 1982; Kraemer, 1980;
grants.gov (U.S. Government Accountability Office, 2011b, on the volume of online data transactions it receives from its partner of grants.gov (a joint online-application program the Department of Health and Human Services, the management partner of grants.gov (a joint online-application program for federal grants) recently requested to restructure the program on a fee-for-service payment model by which each of the 25 participating agencies would pay its share based on the volume of online data transactions it receives from grants.gov (U.S. Government Accountability Office, 2011b, p. 40). Second, agencies pay for data with their own data. For example, the DOD agreed to share biometrics data with the DOJ in exchange for acquiring access to the FBI’s biometrics database. The DOD also successfully acquired ownership over the Department of State’s small biometrics database. In return, the DOD later agreed to sign a Memorandum of Understanding document with the Department of State that granted the latter access to the much larger and more valuable DOD biometrics database. Third, an agency can receive special federal funding for providing data to another agency, such as when the SSA provided data to support CMS’ Medicare Part C and Part D. Last, Congress compensates an agency for data-integration services provided to other agencies by extracting and transferring valuable budgetary and personnel resources from the receiving agencies to the providing agency. For example, the new Terrorist Screening Center established its Terrorist Threat Integration Center Identities Database partially by taking over data, functions, and personnel from the Department of State’s older TIPOFF counterterrorist program, the FBI Watchlist Unit, and the Transportation Security Agency’s No-Fly and Selectee list program (Secretary of State, Attorney General, Secretary of Homeland Security, & Director of Central Intelligence, 2011).

Note that agencies are not the only actors who seek to benefit from the information products they labor so hard to create. The popular web culture itself has evolved from its infant “everything-for-free” phase to its current for-profit condition. Successful companies such as Apple, Amazon, and eBay have developed marketplaces where digital products are cheap, but not free-of-charge. These companies reap significant financial gain because even a cheap product yields handsome profits when multiplied by billions of split-second transactions. The penalty for driving down the price of any digitized commodity to $0 is immediate and harsh. For example, Europeans failed in their first attempt to build an electronic-emissions marketplace because they issued too many allowances, to a point where trading in the valueless new commodity practically stopped (European Climate Exchange, 2010; European Union Emission Trading System, 2010). In short, it is unfair and politically unfeasible to demand federal agencies to give for free the information products they labored hard to create.

However, if agencies refuse to release valuable datasets to data.gov, why did they not inundate the Open Data site with countless less valuable datasets that they own and cannot trade with other agencies? After all, in the age of social networking, it is easy and cheap to generate abundant quantities of valueless data. Only a small number of agencies have evidenced this fact: The U.S. General Service Administration posted on data.gov a searchable version of its 1,700-page, semiannual regulatory agenda, including “nuggets” such as the administration’s proposal to set standards for a “retirement home for chimpanzees used in federal research” (Noveck, 2009, p. 125). Similarly, the EPA, the Federal Communications Commission, and the U.S. Forest Service published datasets containing millions of
nonsearchable public “comments.” However, these are the exception; so why does the data analysis suggest that overall agencies significantly reduced the indexing of valueless datasets on data.gov?

I propose that federal agencies quickly comprehended that cooperation with Open Data—even the uploading of less valuable datasets—had the potential to exacerbate three data-integration problems that arise before data are ready to be released to the public. First, the problem of inclusion concerns what data are included or excluded from the datasets that agencies index on data.gov. For example, distorted, incomplete, or misleading data could be intentionally released to damage or promote the cause of certain groups.

Second, the problem of confusion refers to the fact that it is almost impossible to rectify bad data once they are published on the web because they trickle downstream to numerous databases that are not under the control of a single entity. In contrast, it is far easier to resolve bad data upstream. For example, almost 2 decades ago, the SSA successfully created its Master Death File that directly feeds numerous other databases to ensure that the deceased no longer receive correspondence.

Last, the problem of diffusion is created by the fact that Open Data did not offer a mechanism to time-version datasets, to create historical integration points among composite datasets, or to allow data to retire or expire. Knowing how the citizen’s composite virtual self (i.e., data woven across federal databases) appeared to government officials at a given point in time is critical to understanding why, for example, citizens were arrested or deemed as worthy of receiving welfare benefits. Without such time-versioned mechanisms, data.gov is a useless index where one cannot connect datasets in a meaningful way. Open Data architects celebrated the idea that “data, once posted, cannot later be sequestered or taken down” (Noveck, 2009, p. 126). This “post it forever and delete nothing” argument ignores the problem of deriving meaning from historical data. The National Archives and Records Administration and the GAO admitted that this problem is risking the data quality of all federal databases (U.S. Government Accountability Office, 2010a; National Archives and Records Administration, 2006).

At no stage of the Open Data initiative did the Obama Administration prohibit common interagency data-trading practices as illegal; hence, agencies remain highly motivated to protect precious and tradable information assets from the OGD’s requirement to “free” high-value datasets. At the same time, the Open Data program exacerbated agencies “back office” problems with data quality. Over many years, agencies have fine-tuned routines to tackle the problems of inclusion, confusion, and diffusion so that they can account for data released to the public. Therefore, agencies were reluctant to support a program such as data.gov that threatened their ability to trade and exacerbated their “back office” data-quality problems, which could lead to an inundation of citizens’ inquiries regarding the quality of their published data. In other words, Open Data architects offered federal agencies a bad “trade”: Open Data would receive fame and glory for “freeing governmental data,” federal agencies would be robbed of their ability to trade data and would pass on to them the tedious, thankless, and hard work of resolving data-quality problems that the Open Data program itself exacerbates. The technical managers of federal agencies understood the true cost of supporting Open Data and therefore reduced their involvement with Open Data to a minimum.

Future of the Open Data Program

Via the manipulation of the “zeitgeist of democratizing data,” the nontested and unchallenged Open Data program has become an almost religious dogma (Lakhani et al., 2010, p. 5). Open Data architects have continued to claim that Open Data successfully mobilized the expertise of the masses, harnessed the hidden desire of agencies to share data, and gave birth to a community of innovators across the executive branch of government. In the Open Data jargon, individuals are “called to arms” to support the Open Data program, everything associated with “the government of the past” is brushed away, and the spirit of openness and innovation is contagious across government (Noveck, 2011). The Open Data program continuously promotes an imaginary world in which organizations and people happily and altruistically share data and applications (Mahoney, 2000).

However, this “one-size-fits-all” monolithic approach continues to ignore both the nature of federal agencies as “political creatures” that bargain with each other and the unique features of federal data as a commodity. This article argues that federal agencies will never “free” precious information assets that define their political status and bargaining power vis-à-vis other agencies. Agencies can and do conceal information to barter knowledge with each other. As Open Data continues to pressure agencies to “free” data, most agencies have adopted a passive–aggressive stance to the program by indexing a minimal quantity of mostly useless data while locking more valuable datasets inside closed database “gardens.” Consequently, software developers have discovered the valueless nature of most “freed” federal data. Initially, they enthusiastically developed applications for newly released federal data. However, these early applications were plucking the “low-hanging fruit,” which was quickly depleted; after which developers lost interest in the harder task of sifting through freed, valueless federal data. Similarly, in New Zealand, a project similar to data.gov died prematurely after software developers discovered that facts were dumped into cyberspace without purpose and refused to develop applications for them (The Economist, 2010b).

Ultimately, information assets are the source of political power in Washington, DC. Agencies have been trading data with each other long before the onset of the Internet age (U.S. Government Accountability Office, 1991); hence, they refuse to “free” valuable information. All this is old news. Control over what information is acquired and how, where, when, and why to disseminate data has always been the ultimate asset
bureaucrats possess in their internal bargaining games. Open Data architects disdainfully called this argument ‘the nasty, brutish, and short worldview of government’ (Noveck, 2009, p. 190). However, senior Washington bureaucrats understood that in the name of transparency, Open Data architects were demanding that they deprive their agencies of possession of that which defines their very being in the new information era: electronic data. Disinclined to openly confront the Obama administration’s flagship data.gov program, agencies adopted a passive–aggressive stance toward the program.

Excuses for refusing to cooperate with Open Data proliferate; the OGD possesses a long list of such excuses (e.g., national security, personal privacy, breached confidentiality, and even “damage to other genuinely compelling interests”). Senior agency officials are skilled veterans who know all too well how to provide excuses that will exempt them from releasing data. Open Data advocates accused allegedly incompetent federal information technology bureaucrats of failing to understand the Open Data program (Noveck, 2009, pp. 123, 125). However, the empirical findings of this article suggest that precisely because senior bureaucrats did understand the principles of Open Data, they were reluctant to cooperate with the program. In other words, Open Data’s transparency goal clashed head on with federal agencies’ own data trading agendas. Recently, Kundra resigned from his CIO position to protest the harsh cuts in the Open Data budget (Wadhwa, 2011). However, Open Data’s history-in-numbers as told in this article have suggested that this program had failed to accomplish its goals long before the program’s budget was slashed.

Open Data provides a fascinating case study which highlights three new research avenues. First, this case study raises a series of questions regarding the definition and meaning of the concepts of “record” and “database” in governmental work. According to the Privacy Act of 1974, a “record” is any item, collection, or grouping of information about an individual that is maintained by an agency and contains his or her name or another personal identifier. The Privacy Act then defines a “system of records” as a group of records under the control of any agency from which information is retrieved by the name of the individual or by an individual identifier. The GAO recently acknowledged that these definitions are hopelessly arcane in the age where agencies acquire data from various social networking channels (U.S. Government Accounting Office, 2010a). This article suggests that even after new data are acquired and stored on federal computer systems, agencies command various means to “slice and dice” the data to appear (or not appear) as if they are complying with various transparency policies. This topic requires further empirical research to propose ways to redefine the fundamental concepts of database and record in governmental work.

Second, the Open Data case study casts doubt on the vitality of the political control hypothesis in public policy scholarship. According to this hypothesis, politicians can effectively coerce federal agencies to free their digital data troves. By this argument, the president (via presidential committees) and politicians (via Congress) possess the means to change bureaucratic behavior (Epstein & O’Halloran, 1999; Huber & Shipan, 2002) by engineering the direction of new agencies (Bawn, 1995, p. 64), stacking the deck in favor of certain interests while constructing new agencies (McCubbins, Noll, & Weingast, 1984, 1987, 1989), and mirroring a political agenda in the structure of a new agency and its work processes (Balla & Wright, 2001; Yackee & Yackee, 2006; Zegart, 2004). However, as narrated by this article and its findings, bureaucrats have found new information technology techniques to escape the scrutiny of politicians and to appear in compliance with directives while adhering to their own agenda. Scholars must pay closer attention to the role information technology plays in this “cat-and-mouse game” between politicians seeking to control bureaucrats and bureaucrats developing new means to evade such control.

Third, and most important, the failings of the Open Data program emphasize the necessity to erect a formal internal Federal Information Marketplace (FIM) to incentivize agencies to improve data sharing via trade. For almost a decade now, politicians have attempted in vain to coerce federal agencies to improve electronic data sharing. FIM will ensure increased information sharing within the federal government by accounting for bureaucrats’ unwillingness to relinquish power and information sources for free. FIM will adhere to Madison’s old maxim to build a good government based on “multiplicity of interests” and the principle that “ambition must be made to counteract ambition” in governmental work (Hamilton, Madison, & Jay, 1961, p. 322 [Federalist 51]). Public-sector scholars and information technology practitioners can partner to create an FIM prototype to demonstrate that data trading can improve information sharing. Returning to the Hurricane Katrina information-sharing example that opened this article, only an information marketplace in which self-interested actors exchange real-time data can ensure that information will be sent to the right people at the right place at the right time. If FIM is established to deliver on this promise, it will become the federal government’s third-most important computer innovation in history, following the American government’s contribution to the inventions of the computer and the Internet.

References


