Issues in TECHNOLOGY Innovation

Number 14 December 2011

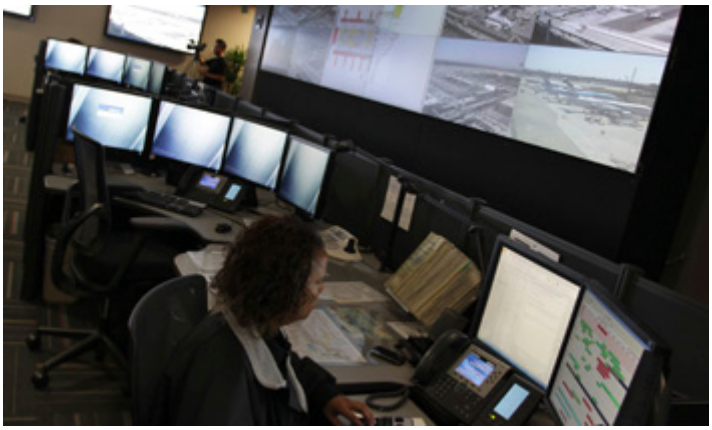
Government-wide Information Sharing for Democratic Accountability*

The politics of using unique identifiers to track powerful political players across government databases may be less dismal than widely believed.

J.H. Snider

ABSTRACT

In a representative democracy, average citizens should be able to easily monitor the public actions of their representatives and the politically powerful who seek to influence those representatives. New semantic web technologies make such monitoring more cost effective to do than ever before. But while these technologies have been widely used to monitor the weak, they have not been used to monitor the powerful, who often cite privacy and cost concerns as excuses to avoid such monitoring. This paper recommends asking the president of the United States to 1) use the new technologies so the American people can more easily monitor his public, official actions, and 2) serve as a showcase for Congress and the rest of the executive branch of government-wide information sharing for democratic accountability.



Reuters/Jonathan Alcorn

Issues in Technology Innovation

The Center for Technology Innovation at Brookings has launched this paper series to seek and analyze public policy developments in technology innovation.

The Center for Technology Innovation

Founded in 2010, the Center for Technology Innovation at Brookings is at the forefront of shaping public debate on technology innovation and developing data-driven scholarship to enhance understanding of technology's legal, economic, social, and governance ramifications.

* This paper draws heavily on J.H. Snider's paper, Connecting the Dots for Democratic Accountability (Washington, DC: iSolon.org, October 22, 2010). Whereas this paper focuses on government-wide individual identifiers, that paper focused on government-wide organizational identifiers.

Introduction

If public officials can easily track citizens (aka "potential terrorists") across thousands of government databases, why cannot citizens do the same for public officials (aka "democratic representatives") and the powerful political players that seek to influence them?

J.H. Snider is president of iSolon.org and a network fellow at Harvard's Edmond J. Safra Center for Ethics

Tracking the Politically Weak Vs. Powerful

Consider a public official with a security clearance who wants to do a background check on a citizen. He can enter the citizen's name in a simple query and find any suspicious activity and relationships for that citizen gathered government-wide, including by tens of thousands of local police, fire, and transportation departments. The result of this government-wide information sharing is that the public official can "connect the dots" and find patterns of suspicious behavior otherwise undetectable.

Now consider an investigative reporter who wants to investigate the official and nominally public actions of elected representatives and those who seek to influence those representatives. To do his job, the reporter, like the public official, needs to search across thousands of government databases, albeit for different information. For example, to track the official actions of a member of Congress, including possible inappropriate influences on and beneficiaries of those actions, the list of databases is long, including Congressional campaign contributions, lobbying, gifts, travel, office expenses, correspondence to Federal agencies, floor and committee votes, floor and committee remarks, press releases, bill text and sponsorships, reservations for Capitol Hill meeting rooms on behalf of private organizations, personal finances, regulatory appointments, and employment pre- and post-Congress. Federal agency databases, including agency contracts, awards, permits, leases, and licenses; comments in regulatory proceedings; advisory committee appointments; and various ethics disclosures, may also prove insightful.

The information categories listed above are conceptual, not physical. Physically, they may each be subdivided into many separate databases. For example, the U.S. House and U.S. Senate keep physically separate records; each of the dozens of Federal agencies keeps its own Congressional correspondence; and licenses/leases to use oil, timber, grazing, spectrum, building, and other public assets are physically scattered across hundreds of different Federal databases.

If an investigative reporter could search government-wide for the names of individual public officials with the same ease that public officials can search for potential terrorists, the reporter would undoubtedly be able to do his job much more efficiently and effectively.

The Semantic Web Breakthrough

Emerging so-called "semantic web" technologies make searching across government

databases and websites easier than ever before imaginable. A vivid simple example is product ratings. Google created a simple set of product rating metadata (metadata is data about data) for any website to tag its product ratings for easier search. Now anyone who enters a product in a Google search can find all the ratings for that product scattered over the hundreds of millions of websites that Google scans.

NIEM does have the potential to be used for open government applications.

Semantic web technologies also promote relationship mapping, as illustrated by Facebook's social metadata, which makes it possible to track friendships and other types of relationships among individuals. Tracking relationships for democratic accountability is especially important because embarrassing political influence is typically laundered via intermediaries.

If government databases tagged the name of public officials with a unique identifier, the same type of decentralized search and relationship mapping could be used to track the official actions of those officials.

NIEM and UCORE

To a remarkable degree, the Federal government's National Information Exchange Model (NIEM) already makes feasible highly ambitious government-wide information search. Launched after 9/11 to facilitate government-wide sharing of information about potential terrorists, NIEM has greatly expanded its sphere of sharing in the last few years.

NIEM creates standardized metadata that make it possible for different government databases to easily share information. The various sets of metadata may be called ontologies, and the ontologies are organized hierarchically. At the top of the hierarchy are ontologies shared by particular domains of knowledge such as the CIA and Department of Justice. As of November 2011, NIEM had twelve domains: Biometrics; Chemical, Biological, Radiological, Nuclear (CBRN); Children, Youth, and Family Services; Cyber; Emergency Management; Immigration; Infrastructure Protection; Intelligence; International Trade; Justice; Maritime; Screening; and Human Services.

At the bottom of the hierarchy is a core set of ontologies describing attributes such as who, what, when, and where that are shared by virtually all exchanges of data. A popular version of this NIEM core is called Universal Core (UCORE).[2]

Despite NIEM's mission to facilitate information sharing, all its initial domains of knowledge were associated with secrecy, notably national security and criminal justice. Its newer domains, such as healthcare and family services, have also predominantly shared information only with a privileged set of users. Open government, in contrast, is based on the democratic norm that government

[2] The current versions of UCORE and NIEM, UCORE 2.x and NIEM 2.x, work together but are not fully integrated. NIEM 3.x and UCORE 3.x are expected to be fully integrated. NIEM/UCORE 3.x may also add RDFa and OWL functionality. Some UCORE 2.x implementations on military intranets have already added RDFa and OWL functionality.

information is equally available to all citizens.

However, NIEM does have the potential to be used for open government applications. For example, it was used to implement Recovery.gov, which tracks Federal contract expenditures. The reason NIEM remains strikingly weak in the area of open government probably has less to do with technology than politics.

The Core of the Core

From the perspective of open government, Who metadata may be viewed as the core of the core. This is because Who metadata is essential to holding public officials publicly accountable. If you cannot distinguish between the thousands of John Does with the same name and different permutations of the same name in government databases, government-wide search isn't very useful.

So why not apply a Who ontology to public officials and the powerful political players who seek to influence them? The standard response, other than the increasingly implausible "it is too expensive to do," is that this would violate the privacy of U.S. citizens. The last thing we want to do is make it easy for the U.S. government to track all the government interactions of its 300+ million citizens. This hearkens back to the image of the Federal government's discredited Total Information Awareness program, which was unpopular even among the most ardent advocates of the Patriot Act.

But this argument turns out to be merely a convenient straw man. Not all applications of a Who ontology for open government are controversial. Indeed, some would be very, very popular among the American public. If not for the open government community's insular, short-term, copycat, and advocacy oriented culture, which has resulted in a striking lack of public policy creativity (a controversial assessment given the many highly promoted open government innovations in recent years), it would be hard to imagine how this straw man argument could have remained unchallenged for so long.

Technically Easy But Politically Hard: A Who Ontology for Congress

A simple, cost effective, and publicly popular application of a Who ontology would be to track the official actions of the 535 members of Congress. As part of the Federal government's Personal Identity Verification (PIV) system, more than 5 million Federal employees and contractors already have unique personal identifiers, including a unique numeric identifier linked to a unique biometric identifier (such as a fingerprint or iris scan). These identifiers are encoded on a smartcard that Federal employees and contractors must increasingly use to access Federal facilities and computer systems. The long-term goal is to prevent any access to Federal facilities or computer systems without the use of such a device. States, localities, and businesses have also started to implement the global unique ID (GUID) incorporated in the

Federal government's PIV system. Another unique personal identifier for members of Congress, such as the bioguide URL for each member of Congress issued by the Clerk of the U.S. House of Representatives, could be used for this purpose.[3]

The metadata describing the unique identifier associated with each member of Congress would be attached to every official Congressional action. For example, when a member sent an official and thus nominally public letter to any government agency, the unique ID would be attached as metadata to his or her electronic signature on the letter. Each letter could also include what, when, and where metadata. What would specify a letter, When a date stamp, and Where an address (both from and to).

A Who-What-When-Where ontology may include many standardized subcategories of metadata. For example, in addition to a postal address, Where could include an email address, a website, and a set of GPS coordinates. Similarly, When could include a point in time or a period in time.

A journalist would then be able to enter a simple query, say, to track all the correspondence of a member of Congress to the 50+ federal agencies during a particular period of time.

The problem with this exceedingly simple yet powerful Congressional application is clearly not technology, cost, or public appeal. The problem is that most members of Congress would never agree to it, except in response to an unlikely popular uprising, because information is power and they wouldn't want potential challengers and other troublemakers to have easy access to such information.

Winning the President's assent to attaching standardized who-what-when-where metadata to his official public actions is probably a lot more politically realistic than getting Congress to do the same.

More Politically Feasible: A Who Ontology for the President

I propose a different initial application of a Who ontology for democratic accountability, one which is the ultimate combination of simplicity and power. The proposal avoids the complexity and political difficulties of mandating the use of unique IDs for hundreds of millions of Americans or millions of incorporated organizations. It doesn't even bother with the relatively small number of members of Congress. Instead, it initially seeks to implement a Who ontology for only a single person: the president of the U.S. The reason is that winning the president's assent to attaching standardized who-what-when-where metadata to his official public actions is probably a lot more politically realistic than getting Congress to do the same.

President Obama has demonstrated great willingness to endorse open government practices. This proposal is a natural extension of those policies and one he could easily implement with minimal need to cajole the bureaucracy or seek any

[3] For reasons of privacy, unique personal identifiers for democratic accountability purposes should be different from identifiers, such as social security numbers and tax identification numbers, used for other purposes. The GUID may be well suited to take on this democratic function for not only major recipients of government largesse (e.g., government employees and contractors) but also major political players (e.g., registered lobbyists and large campaign donors).

additional Congressional appropriation. Moreover, this pathetically simple and affordable technology would provide the public with a qualitatively improved tool to track his official and nominally public actions on their behalf. To assure accountability, Federal entities that receive the president's official correspondence, including the embedded who-what-when-where metadata, would be required to post it on their public websites, although it could be buried almost anywhere on their websites because the public would primarily access the correspondence via search engines.

Once this simple application demonstrated the utility of the Who-What-When-Where ontology, it could gradually be expanded in ways the public would not only find uncontroversial but would enthusiastically endorse. One way to think of this expansion is in terms of the core of the core: the number of government officials who could be tracked in this way. For example, it could be expanded to all senior White House officials and heads of the various departments that constitute the president's cabinet. The Plum book, published by Congress, which contains the list of the most senior positions in the Federal government, over 7,000 political appointees, would be a good target universe.

Since all Federal GUIDs in the PIV system are attached to unique identifiers for Federal agencies, if you know a Federal employee's GUID, you also know the agency and sub-agency associated with the employee. The agency identifiers are derived from the codes used by the U.S. Department of the Treasury for structuring budget data by agency and sub-agency.

Perhaps the president's actions would even embarrass the heads of the U.S. House and Senate into copying him, with pressure building for similar practices to trickle down to Congressional committee chairs and ultimately even rank-and-file members of Congress and their staffs.

Another type of expansion involves the type of information exchanged. Exchanging correspondence is an especially easy case. Complications arise when the president submits data to an external database not under his direct control, such as filings with the Federal Election Commission or the U.S. Office of Government Ethics. The solution would be to have the president simultaneously transmit his information to both the external database and an internal White House database published to the web. The White House database could duplicate the information on the external database and then add the new metadata, or it could only add the new metadata and use it to point to the relevant information in the external database. An analogy for this type of dual reporting is a retailer who enters sales data once but automatically submits it separately to state authorities (for paying sales tax) and its own accounting systems. Although less than ideal, the White House's dual information reporting system would set an example for other Federal entities and encourage them to design their databases for efficient government-wide information sharing.

Another type of expansion would be to include in addition to who-what-when-where metadata the metadata associated with a particular domain of knowledge. This, as we have seen, reflects the hierarchical structure of NIEM. Such domains of

knowledge tend to be orders of magnitude more complex than the simple who-what-when-where information focused on in this paper. But politically, when it comes to open government domains, there may be fewer obstacles to their implementation. This may help explain why the complex ontology for government budget reporting, XBRL, is actively being considered by both Congress and federal agencies. As implemented by the U.S. Securities & Exchange Commission for tagging corporate financial filings, XBRL has more than 18,000 standardized metadata tags based on Generally Accepted Accounting Principles (GAAP).

The president needs to bravely set an example for all Americans by countering the natural tendency toward information accountability NIMBYism.

Although the focus here is on information sharing for democratic accountability, information sharing has also proved a powerful way to reduce costs across both governmental and non-governmental enterprises. Information sharing has often led to massive economies of scale and reduced costs in every part of the information cycle from creation to distribution to acquisition. Indeed, reducing costs is the major argument for standardization, not just standardized metadata, in almost every sphere where it is employed. The countless private and public organizations devoted to standardization are testimony to this. The irony is that high cost is the most popular excuse among politicians not to implement open government proposals. But it's actually the weakest excuse and should be used as an argument for rather than against information sharing for democratic accountability.

The Politics of a Who Ontology for Democratic Accountability

Unfortunately, a natural human instinct in all spheres of endeavor is to avoid accountability. Americans, including elected public officials and the powerful political players who influence them, all love accountability for others but not themselves. But this information accountability NIMBYism is bad for our democracy. It is also probably the best explanation for why our ability to track the official actions of the politically powerful, including our elected representatives, remains so primitive.[4]

The reason tracking potential terrorists has been so effectively implemented across tens of thousands of government agencies from the local to the Federal level is not that tracking terrorists is technologically simpler and less expensive than tracking the politically powerful. Indeed, tracking the official actions of elected officials is technologically trivial and economically a pittance by comparison. Nor is it that potential terrorists are less resistant to being tracked than the politically powerful. It is that potential terrorists aren't a powerful political group and cannot effectively engage in information accountability NIMBY politics.

The president needs to bravely set an example for all Americans by countering the natural tendency toward information accountability NIMBYism. No one else is better positioned, by setting a personal example, to usher in the new era of semantically

[4] For a discussion of such NIMBYism in Congress, see J.H. Snider, The Dismal Politics of Legislative Transparency, *Journal of Information Technology & Democracy*, Spring 2009.

enriched open government data. Since the president's official and public actions are already so closely tracked by a large and sophisticated press corps, he also has the least to lose.

Given the conflict of interest politicians have in designing democratic information systems, government-wide information sharing for democratic accountability that moves beyond the president and his direct sphere of control involves many difficult governance issues, which are not addressed here. This is the realm where the NIEM governance model breaks down. More generally, it is the fundamental problem associated with open government public policies that are supposed to generate democratic accountability for the politically powerful as well as the weak. Ultimately, the governance problem can only be solved through the creation of new checks & balances institutions. By showing what's technically and economically trivial to do to improve our democratic information systems, the president can cast a light on the centrality of governance issues.

Conclusion

The ability to easily track the official actions of elected representatives and the powerful political players who seek to influence them is essential to a healthy democracy. But endemic to democracy is that such powerful political players do not want their actions tracked. Since information is power, those with such power have no rational incentive to give it up. But for the sake of our democracy, they must be forced to do so. Progress, made possible by the emergence of semantic web technologies, should not be held hostage to their anti-democratic interests. The president, among all government officials, is best positioned to usher in the new era of technology-based democratic accountability by setting a personal example and pointing to the difficult governance issues that need to be addressed.

The Center for Technology Innovation
The Brookings Institution
1775 Massachusetts Ave., NW
Washington, DC 20036
Tel: 202.797.6090
Fax: 202.797.6144
http://www.brookings.edu/techinnovation

Editor
Christine Jacobs

Production & Layout
John S Seo

Tell us what you think of this *Issues in Technology Innovation*.

E-mail your comments to techinnovation@brookings.edu