ERECTING THE PUBLIC INFORMATION MARKETPLACE

Alon Peled

Hebrew University of Jerusalem, Israel, Alon.Peled@post.harvard.edu

Follow this and additional works at: http://aisel.aisnet.org/mcis2010

Recommended Citation

http://aisel.aisnet.org/mcis2010/67

This material is brought to you by the Mediterranean Conference on Information Systems (MCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in MCIS 2010 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
ERECTING THE PUBLIC INFORMATION MARKETPLACE

Alon Peled, Hebrew University of Jerusalem, Israel, Alon.Peled@post.harvard.edu

Abstract

How can we incentivize public agencies to share electronic information effectively? I first argue that the Obama Administration’s Web 2.0 methodology of addressing this question (http://www.data.gov) will fail because (1) Federal agencies will not willingly surrender precious information assets that are the source of their political power; (2) valuable data has a currency value in the information age—it is not free; and (3) a Web 2.0 approach to sharing public sector data ignores (and sometimes exacerbates) important problems including how to delete data about citizens, how to create historical versions of the digital image of the citizen, and how to uncover the hidden and powerful algorithms that govern the production of public sector data.

Instead of Obama’s Open Data approach, I propose to redefine the public sector as that which holds the legitimate monopoly on primordial data (such as social security IDs). Based on this, I have developed a new information sharing architecture based on three key ideas: (1) bureaucratic politics must take primacy over technology; (2) bureaucratic language can be automated to facilitate information-sharing transactions; and (3) information-sharing transactions can be monetized as buy and sell transactions.

This paper also presents the theoretical foundations for a computational linguistics project to erect a Public Information Marketplace, and discusses several political, legal, economic, ethical, regulatory and technical challenges for building such a marketplace. The paper concludes with a concrete political strategy on how to advance the proposed new information marketplace.

Keywords: Information Sharing, Open Data, Electronic Trading, Public Administration

1 Portions of this paper are currently under adjudication for publication in the Journal of Public Administration Research and Theory (JPART) and Public Administration Review. A small section of this paper was published in Peled, Alon. “The Electronic Mountain: A Tale of Two Tels”. American Review of Public Administration, Vol. 37, No. 4 (December 2007): 458-478
1 THE PUBLIC SECTOR'S ELECTRONIC INFORMATION SHARING PROBLEM

The private sector is accumulating huge volumes of digital data. For example, Google alone holds over 200 times the amount of data stored in the Library of Congress. Several American Federal entities such as the National Security Agency (NSA) and the Department of Energy (DOE) store even larger volumes of digital data than Google.\(^2\) The digitization of data in numerous domains further intensifies this data-accumulation trend. Thus, for example, in 2000, a survey estimated that the Department of Defense's (DoD) biological bank held 176.5M files and 282M biological samples and that this biological storage house was growing annually by 20M files—this before the intensification of national security efforts after September 11 2001 (Anderlik and Rothstein 2001 415).

Today, the public sector only "sees" the virtual image of the citizen rather than interacting with the physical citizen herself. Agencies fiercely battle each other to acquire a legal monopoly over new information assets. An increasing number of Federal agencies posses a sophisticated high-end computing gear and technical knowledge, and process huge amounts of digital data so they can replace human intuition with statistics and business intelligence (Ayers 2007 10, 31, 33, 43-45, 69). Sometimes, the value of information assets is discovered when these assets are stolen. For example, in 2008, the German government purchased for $6.2M a stolen CD with information about the foreign bank accounts of rich Germans. Many rich Germans then voluntarily confessed about their foreign bank accounts. Over the next two years, the German government collected an additional 180M euro in tax revenues thanks to this stolen CD (i.e., thus, the $6.2M investment yielded a 4500% growth) (Jolly 2010; Landler 2008).

In my previous work, I redefined the public sector as the collection of institutions whose administrative staff successfully upholds a claim on the monopoly of the legitimate process of producing, updating and disseminating primordial data. Primordial data are primary data elements (ex: social security identifier) about the entire population that serve to link and update all other data sets.\(^3\) For example, in the USA, one is not legally dead until his or her name lands inside the infamous Social Security Administration (SSA) “Master Death File”, which, in turn, updates numerous other databases (GAO 1994, 2007). I explained that each generation redefines the essence of the public administration based on its needs. For example, Woodrow Wilson, in 1887, sought to protect the public sector from corrupt politicians and therefore defined public administration as laying outside the proper sphere of politics (Wilson 1887). Likewise, after the tumultuous 1960s, social equity became an important mission for government, and scholars followed suit by inserting this new mission into the definition of the public sector (Frederickson 1971). More recent definitions have been influenced by globalization and theories of social capital (Bozeman 1987; Pierre 1995).

\(^2\) NSA tracks, records and analyzes close to 70% of all electronic communications on earth via Project ECHELON (formed in 1971) and DoE has been collecting and analyzing data on simulations of nuclear weapons tests since 1972.

Our generation, I argued, must redefine the public sector as an information-processing machine due to three reasons. First, the public sector’s data elements are ubiquitous, ranging from the obvious (ex: a driver’s license ID) to the obscure (ex: Kentucky’s Bonus Wildlife Management Area Quota Hunt Deer Permit ID). Ownership of a particular data element is assigned or revoked by law. For example, a congressional act assigned ownership of Energy Guide labels to the DOE and removed this responsibility from state-level bureaucracies. Agencies gain significant new resources when they become the owners of a new data element. For example, the Department of Agriculture (DOA) won such new resources to develop standards for organically produced food products. Second, the public sector accumulates an ever-growing stockpile of digital data elements that it must maintain and disseminate. Finally, an information food chain governs our modern economy. This chain is founded on the public sector’s primordial data elements. In short, the public sector is the “great infocrator,” a term which combines the ideas of “information,” “creator” [of data elements] and “dictator” [ruling over primordial data elements]. There is no escape from this infocrator. Every life change and event is captured in a data element owned by a public agency. Moreover, by law citizens must provide timely information to update these data elements. Resultantly, the public sector then becomes the owner of the most valuable information assets.

Yet, public organizations fail—time and again—to share their precious information assets with other public agencies. For example, Federal agencies’ inability to share counterterrorism information was the US government’s greatest failure before 9/11. The President and Congress have since enacted eleven presidential directives and congressional acts instructing the new Department of Homeland Security (DHS) and other agencies to share security information (GAO 2006). The Government Accountability Office (GAO) has designated homeland security information-sharing as a high-risk area (GAO 2005, 2007). These efforts, however, have not yielded adequate results, as recently evidenced by the USA intelligence community's failure to fuse into a coherent story all the discrete pieces of information regarding the Christmas 2009 airline attack. Addressing this crucial issue, a White House review concluded that counterterrorism IT did not sufficiently enable the correlation of data that would have helped avert this attack (Elliott and Baldor 2009; White House 2010). It is clear that information sharing has not significantly improved since 9/11, and that federal agencies are still failing to "connect the dots". Therefore, how can we incentivize public institutions to share information more effectively?

2 A CRITIQUE OF OBAMA’S OPEN DATA APPROACH TO INCREASE ELECTRONIC INFORMATION

The Obama Administration has adopted an "Open Data" approach to address this aforementioned question. On his first day as President, Obama signed the Open Government Directive instructing all Federal agencies to post on the web as much data as possible. Specifically, this directive instructed each Federal agency to publish on the web three high-value data sets. In March 2009, Obama appointed Vivek Kundra as the first Federal Chief Information Officer (CIO) and Kundra, in turn,

4 Wherever and whenever used, the terms "citizens" in this paper refers to all people living in the state (citizens and residents alike) about which the Federal government collects and stores data.
launched www.data.gov. The key idea behind this initiative was to foster a web 2.0/social networking environment in which Federal agencies release valuable data. The public, in turn, Kundra argued, will then build applications to analyze and display the released Federal data. Released data and open source applications are then meshed up to increase governmental transparency and accountability and to improve public services (Economist 2010 10-11; Hoover 2009 ; Noveck 2009).

I argue that this approach is the wrong panacea for Federal information sharing problems due to three reasons. First, information assets are the most important sources of political power for Federal agencies; hence these agencies are unlikely to release these assets. Instead, I believe that agencies will "comply" with Obama's Presidential Directive by inundating data.gov with large volumes of unimportant data while carefully and jealously guarding more valuable data. Second, in the information age, the value of data is not $0. In the commercial marketplace, the most successful vendors, such as Apple, eBay and Amazon, are the ones who managed to create marketplaces where bits and bytes of data (including digital versions of songs, applications and movies) are exchanged for a price—even if a very low price measured sometimes in fraction of cents. I believe that, if the Obama Administration will continue to press hard to promote its www.data.gov initiative, a dangerous downwards spiral will evolve. Agencies will increasingly release large quantities of useless data. Software developers will discover that the released data is useless and will develop less open source applications to manipulate this data. Governmental agencies will then be "disappointed" to discover that so little is done with their released data and will release even less important data. Ultimately, both the government and the public will lose interest in www.data.gov. Thirdly, www.data.gov does not address, and, in some cases, exacerbates, the key problems that government must tackle in the information age. These problems include questions such as: How, when and why does the government delete data about citizens? How can a citizen know what his or her "digital image" looked like at a given point in time (say, for example, two years ago) when the government made an important decision about him or her (for example, to deny the citizen a Federal job)? How can citizens correct inaccurate personal data across multiple Federal computers? How can Federal decision makers know which data assets are more important than others? How can the Federal government uncover, inspect, and correct some of the hidden algorithms that command the process of manufacturing and using data about citizens? How can we detect if certain sources of data were not created, released, and manipulated with the explicit intention of hurting certain groups, interests or individuals? Unpleased with Obama's Web 2.0 approach to increasing electronic information sharing, in the next section I will develop an alternative approach to expand Federal electronic information sharing via a new public information marketplace.

3 ERECTING THE PUBLIC INFORMATION MARKETPLACE

I propose that public data be thought of as a "contested commodity." A contested commodity is a product or service where the process of its commoditization is likely to yield benefits such as the expansion of trade, the settlement of flexible and

---

5 Radin (1996) defines and explains the concept of a "contested commodity."
competitive prices and the uprooting of black marketeering behavior. Nonetheless, this product is "contested" because people are left with an uneasy, gnawing feeling that it is not ethical to place such a product within the context of a marketplace. Typically contested commodities involve, in one aspect or another, the process of "commodification of life" itself (Parry 2004, 265). Examples of "contested commodities" include trading in bodily organs (live and cadaver organs), prostitution, surrogate motherhood services, babies, the bodily relics of saints in medieval ages, detailed marriage contracts and genetics information. I argue that digital data about the so called "virtual citizen" is another type of a contested commodity.

I further propose that government can carefully create an internal public information marketplace within which public agencies will buy and sell this contested commodity. Selfish agencies which place their narrow bureaucratic interests above everything else will be drawn to trade data in this new marketplace. Over time, this marketplace will create powerful incentives for public agencies to discover and unlock internal information assets. A creative political and regulatory framework along with a new controlled computer natural language can be created to launch this public information marketplace. I contend that a complete political, ethical, economic, legal, regulatory and technical infrastructure can be erected to unleash a new "information ecology" within the Federal government. The erection of this new public information trading marketplace will save billions of dollars as doomed-to-fail Federal database-integration projects will no longer be needed. In this proposed marketplace, just-in-time data assets will become available to agencies. These assets will then empower a more nimble and smarter government whose actions and decisions are more transparent to the public.

The digital data that the government collects about us, the citizens, falls under one of the following three categories. Data that is both predictive in nature and was extracted unintentionally (ex: genetic data discovered via a routine blood test) belongs to the "non-tradable data" category. Conversely, data such as the license number of a car or a gun belongs to the "tradable data" category. In between these two categories, huge quantities of public data belong to the third category which I term the "contested data" category. On the one hand, subjecting these pieces of data to the powers of a marketplace will yield enormous benefits. These benefits include the ability to create—for the first time in the history of the Federal government—a real and concrete price tag for data, and the constant and instant movement of data within government as the exclusive and objective measurement for the value of public information assets. In addition, Federal employees will easily learn to master this new data marketplace because it will merely extend their usual and customary competitive bureaucratic fights into a new domain. On the other hand, we must be careful and disable some marketplace behavior due to the nature of the data traded. Thus, for example, we may opt not to allow the advertisement of certain pieces of data (ex: potentially frightening data about contaminated former military bases) while permitting the trading in such data between, say, the Environment Protection Agency (EPA) and the Department of Defense (DOD).

---

6 The cleanup of old military bases contaminated by toxins is a good case in point. See: (GAO 2009, 2009).
The proposed new public information marketplace will allow agencies to set prices for their data assets based exclusively on external demand (i.e., from other agencies). This marketplace will deter fraud as the reputation of the agencies trading in data will be at stake. As with the online eBay or Amazon marketplaces, an agency that cheats one time will pay dearly for its unruly behavior in countless future transactions. A network of external computer servers relying on a computerized version of the bureaucratic language (similar to the Dutch RINIS system7) will facilitate and support the buy and sell transactions. Micro-transaction engines that already exist commercially will empower agencies to execute trillions of buy and sell data transactions vis-à-vis each other and execute these transactions in split seconds. Multiple information products will be bought and sold in this new marketplace. These products can broadly be classified into three categories. First, primary pieces of data will be bought and sold (for example: The CIA will be able to purchase data from other agencies about suspects and their whereabouts)8. Second agencies will buy and sell business intelligence types of information products (for example, EPA can sell a new predictive information product that index the probability of a former defense bases to suffer from one or more contaminations). Finally, agencies will even be able to buy and sell futures contracts in order to sponsor new information technology projects. Thus, for example, an agency will be able to pay for the hardware and software of a new information system by selling to another agency the rights to harvest and analyze the so-called data-exhaust or electronic footprints of citizens who will use the new system once it becomes operational. Gradually, carefully, and over time, this electronic public information marketplace will be extended to allow private sector organizations to join the trading and, later, even citizens will be able to buy and sell data vis-à-vis the government.

Important IT trends converge today to create a window of opportunity for the erection of the proposed public information marketplace. Today, Federal agencies already breathe and think digital data. Digital data itself is the perfect mobile, versatile and tradable product as evident by the success of electronic commercial marketplaces such as Apple and eBay. It is also much easier today to convert data from one format to another. Finally the cost of hardware, software and bandwidth continues to decline. Similar commercial marketplaces already exist in the private sector (Ayers 2007 134-135).

---


8 During the 18 months preceding 9/1/2001, the CIA searched for data about the whereabouts of Mohamed Atta. During this very same period, almost a dozen public and private databases held information about Atta’s activities. For example, the FBI and DMV had information about four incidents in which Atta’s vehicle was stopped in Florida. The INS (today USCIS) database held five reports on Atta’s travel from and to the United States, and three about his domestic travels. The FAA’s Airman Registration and Aircraft Registry databases contained 14 entries about Atta’s efforts to acquire a commercial pilot certificate and to train in a large jet flight simulator in six different air fields in three states. Private sector databases also held multiple records on Atta’s apartment leases, cell phone purchases, aviation fuel acquisitions, new bank accounts, and transfers of money into these accounts. All these pieces of data were utterly useless from the perspective of the agencies owning them. Hence, for even the smallest credit, these agencies would have been happy to sell them to the CIA (Bright et al. 2001 ; FBI 2007 ; National Commission 2004 ; Inspector General 2002 )
4 CHALLENGES EN-ROUTE TO ERECTING THE NEW PUBLIC INFORMATION MARKETPLACE

Critics are likely to raise powerful challenges and problems en-route to implementing successfully the proposed public information marketplace. Legislators who support this proposal will need to overcome the incorrect yet powerful "public data must be free and it belongs to the citizens" concept that the Obama's administration is promoting. Journalists may argue that trading in public data will encourage unruly behavior. Worst still, if my proposal is successful, these critics may argue, the public will quickly lose its altruistic ideals how public data can be more than just currency (Cohen 1989-1990 26). Ethicists may suggest that my too-efficient "just-in-time-data" marketplace will propel a new type of big brother government. Bureaucrats from politically-weak agencies might be concerned that data-strong ministries (such as the Social Security Administration and the IRS) will exploit the new system to become even stronger by extracting and manipulating the data assets of weaker agencies. Similar concerns were voiced regarding how the information revolution makes stronger countries and corporations even stronger at the expense of other weak (and weakening) countries and corporations (Braunschvig 1998 ; Premkumar 2003). Private sector leaders will become alarmed that the new marketplace might create a too-knowledgeable government. Hence, these critics will further argue, this proposed information marketplace will alter (for the worst) the delicate balance of (information) power among citizens, the third sector, private corporations and the Federal government. Legal experts might be concerned that the necessary legislation required to support the new marketplace would be delayed in Congress for many years as in the case of legislating trade in another contested commodity—human organs (Cohen 1989-1990 7-8). Economists are likely to ask how initial prices for data assets will be set, how pricing algorithms will be developed, and how specific incidents of fraud and black marketeering behavior will be dealt with. Finally, software architects might argue that a universal computerized bureaucratic language is not feasible because the various computer "languages" that Federal agencies "speak" today are too different to find a common ground.

Nonetheless, solutions exist to address these challenges. The naïve and altruistic www.data.gov model is not as successful as its architects would like the public to believe. Over time, politicians, bureaucrats and the general public will discover the truth about data.gov and become more open to listening and accepting alternative approaches such as the information marketplace. The decentralized public information marketplace will provide adequate privacy protections because the citizen's composite virtual data-image will be scattered across scores of agencies rather than stored within a single centralized database. Hence, thieves and abusers of virtual data will be compelled to leave behind electronic footprints in multiple databases, each located in a different agency.

The new public information marketplace will provide incentives for weak agencies to discover and harness their hidden data assets and to cooperate with each other in order to bring new information products to the Federal data bazaar. In an identical manner, during the medieval ages, the politically weak but internally well-networked Armenian and Jewish trading communities discovered how to benefit from newly created markets and fairs (Braudel 1992 165). We can also use some of the profits of the new marketplace to create a "super fund" to reduce the entrance cost for weak
agencies into the information marketplace. In addition, legislators can declare certain highly valuable data elements that participate in virtually every trade (such as social security numbers, passport numbers and vehicle license plate numbers) as shared and non tradable data in order to eliminate the built-in advantage of agencies such as the Social Security Administration that own these data elements.

New types of Federal units and jobs will be created (ex: departments in charge of developing the data pricing algorithms for a given agency) while older units and jobs will be eliminated (ex: units that, today, are in charge of negotiating data exchange agreements vis-à-vis other agencies). Most certainly, data trading will trigger some unruly behavior and uncertainty regarding when and how to acquire data. Still, overall, free marketplaces are very efficient in tackling such behavior on their own terms. On the legislative and regulatory front, we can adopt the Canadian Voluntary Codes (VC) model to start data trading immediately while buying some time for legislators (Canada’s Office of Consumer Affairs (OCA) 2007). Finally, over the past decade, huge progress was achieved in computerizing the bureaucratic language in various domains. Thus, for example, in a previous publication I analyzed the technical success of deNovis, a small startup, which converted the immensely complex Federal medical claims-payment adjudication rules into an English-like controlled natural language.\(^9\) The deNovis case study merely illustrates that experts correctly predicted—already a decade ago—that embedding translation engines within a smart network backbone is a better approach for electronic information sharing than attempts to build gigantic "system of systems" databases or campaigns to mandate technical standards (Bonometti, Smith, and White 1998 ; Leebaert1998\(^{10}\)). Interestingly, among other benefits, my proposal will also resolve the problem of versioning and re-constructing the point-in-time image of the virtual citizen that the Obama's open data approach fails to address.

5 A POLITICAL STRATEGY TO UNLEASH THE NEW FEDERAL INFORMATION BAZAAR

Important public sector reforms sometimes begin with a "crazy idea" such as the one proposed here. In the next phase, a group of scholars will increase the credibility and validity of this idea by building a common reservoir of metaphors that challenge the dominant "Open Data" paradigm. Following this, an unusual and broad coalition of supporters can be built to advance the Federal Information Marketplace initiative. As demonstrated by the sad history of the failed campaign to advance trading in live and cadaver organs, the automatic support of economists and legal experts who habitually support every free marketplace initiative will be insufficient to make the new public data marketplace initiative successful (Crespi 1994 18-19). Therefore, the political coalition that aims to advance the creation of the new Federal data marketplace must include politicians, ethicists, journalists, senior bureaucrats, and technical experts. It is also possible to divide and rule the opponents’ camp. Thus, for example, there will be a number of Federal agencies that would be delighted to see the weakening of the SSA’s data monopoly (via the agency's command of social security numbers). It is also possible to first experiment with a prototype of the proposed public information marketplace in a "public goods" arena such as sharing data about cleaning the

---

environment. The advantage of launching data trading in this type of arena is that, conversely from other security domains, the fundamental assumption here is that all data ought to be shared. We also must not aspire to solve in advance all types of problems. Marketplaces, on their own, are very efficient in creating innovative solutions to problems that initially appear insolvable (ex: how to provide price-discounts for agencies that buy from another agency large bulk amounts of data rather than individual pieces of data). Marketplaces are also very effective, when left alone, in terms of eliminating black marketeering and fraudulent behavior as demonstrated in the case of trading in human organs (Kaserman 2001; Schwindt and Vining 1986).

Scholars who studied public information systems concur that the threshold for success for the proposed new public information marketplace is very low as Federal agencies continue to fail miserably in sharing electronic data (Vann 2005 143-145, 150; White 2007 5-8,11). In fact, some of the trading mechanisms I proposed above already exist today as quasi-legal and informal arrangements within the Federal government (GAO 2008). Once successful, the marketplace will completely change the "information ecology" within the Federal government. Agencies will cease to collect some unimportant data and will begin collecting other types of data. In a similar manner, we will also see more Federal agencies attempting to create and bring to the marketplace new and exciting business intelligence types of information products. Every age in history had its unique "royal merchandise" (Braudel 1992 169). Our royal merchandise is digital information. The proposed new Federal information marketplace where agencies can freely buy and sell data and information products from each other is merely the institutional mechanism to unleash the power of the 21st century's royal merchandise inside the Federal government.
References


