

©IFIP, 2010. This is the authors' version of the work. It is posted here by permission of IFIP for your personal use; not for redistribution. The definitive version is Dawes, S.S., and Helbig, N. (2010). Information Strategies for Open Government: Challenges and Prospects for Deriving Public Value from Government Transparency. In *Electronic Government: Lecture Notes in Computer Science*, M.A. Wimmer et al. (Eds.): EGOV 2010, LNCS 6228, pp. 50–60.

Information Strategies for Open Government: Challenges and Prospects for Deriving Public Value from Government Transparency

Sharon S. Dawes and Natalie Helbig

Center for Technology in Government
University at Albany/SUNY

sdawes@ctg.albany.edu; nhelbig@ctg.albany.edu

Abstract. Information-based strategies to promote open government offer many opportunities to generate social and economic value through public use of government information. Public and political expectations for the success of these strategies are high but they confront the challenges of making government data “fit for use” by a variety of users outside the government. Research findings from a study of public use of land records demonstrates the inherent complexity of public use of government information, while research from information science, management information systems, and e-government offer perspectives on key factors associated with effective information use. The paper concludes with practical recommendations for information-based open government strategies as well as areas for future research.

Keywords: Government Information Strategies, Open Government, Transparency, Public Value

1 Introduction

In 2009, the Obama Administration outlined a set of open government principles for the US government that encompass three fundamental goals: collaboration, participation, and transparency. This paper examines public access to government information as a strategy for achieving transparency. Transparency initiatives generally serve one of two goals. The first is to provide citizens and other stakeholders with a “window” into what government is doing and how it works in order to hold elected officials and public agencies accountable for their decisions and actions. The second goal is to release government data to the public so that taxpayer-supported information can be used to generate social and economic value.

In order to achieve this second transparency goal, the 2009 Open Government Directive (OGD) [11] requires all federal government agencies to post previously internal electronic datasets on a publicly available web site, Data.gov. In addition, the OGD requires all federal government agencies to create an agency-specific Open

Government Plan and to make it available to the public over the World Wide Web. These plans make an agency's mission, activities, and results more visible and understandable to the public. They provide important context and metadata through reports, service summaries, links to major programs, and downloadable datasets.

Recovery.gov is a companion initiative created to track the spending associated with the \$787 billion American Reinvestment and Recovery Act. Recovery.gov is the first effort of its kind to collect information about spending and performance across a variety of different program areas and from multiple levels of government, as well as from non-profit and business organizations that receive federal funds.

All of these information-based strategies also incorporate Web 2.0 tools to facilitate online public dialogs to solicit feedback, questions, and recommendations for improvements. Taken together, these initiatives do more than open government data to public scrutiny; they also actively encourage businesses, civic organizations, and individuals to use government information for their own purposes. These uses especially foster the second goal associated with opening government information – to enable and encourage information-rich applications outside the government that generate economic and social value

The success of these transparency strategies rests heavily on easy access to public information resources. Public information resources are defined as the data, information content, systems, and information services that emanate from the day-to-day administration of government programs [7]. Generally, the use of public information resources extends well beyond the government itself to include a very diverse multi-stakeholder society. These stakeholders represent loosely connected communities of interest, where the object of concern is the information in government systems, not the systems themselves or the technologies that comprise them. The societal value of these public information resources is derived primarily from unpredicted and flexible uses of the data content by all stakeholders [3].

Data.gov and similar initiatives offer public access to information resources that are distributed among different government organizations, locations, or custodians. Data.gov provides electronic access to raw, machine-readable information about government finances, program performance, and decisions. Its goal is to allow people and organizations outside government to find, download, analyze, compare, integrate, and combine these datasets with other information that so that they provide greater value to the public. However, it is important to remember that these datasets are defined and collected in different ways by different programs and organizations. They come from a variety of different systems and processes and represent different time frames and other essential characteristics. Most come from existing information systems that were designed for specific operational purposes. Few were created with public use in mind. Some agencies provide good meta data and other contextual information, others little or none. While quickly getting data out in the open is an important goal of this initiative, the value of the data for any particular use depends on making these characteristics easy for users to find and understand.

While thousands of these datasets are now available and being put to a variety of uses outside the government, a number of valid criticisms have been made in terms of basic usability, weak application of stewardship principles, lack of data feedback and

improvement mechanisms, and inadequate metadata [5]. Both the criticisms and the potential benefits of information-based open government strategies are receiving much current attention, but they are not new or unique. Past research on public use of government information resources has much to offer that can help refine, improve, and enhance current efforts.

The remainder of this paper is organized as follows. First, we present the results of a case study about the public use of land records to illustrate and assess the benefits, barriers, and policy choices associated with public access to government information. Next we discuss the implications of the case findings from the perspectives of information science, management information, and e-government research. We conclude with a discussion of practical recommendations for encouraging value creation and suggest future research directions.

2 Case study: use and value of land records and parcel data

The following case summarizes an in-depth assessment of the prospects for using a common public information resource, land records, for a variety of public and private purposes [4]. The case holistically examines the challenges that confront information-based strategies to derive social and economic value from government data by looking at the full range of information characteristics and uses by both government and external users. The study was conducted in New York State from September 2004 through February 2005 to assess the uses, users, and value of land records and associated detailed parcel information collected. Parcel information is collected at the municipal (town) level, generally when property is sold or sub-divided, for the primary purpose of maintaining real property ownership laws and tax administration. The data is used for this primary purpose by the town, the county of which it is a part, and by the state government. The most detailed information on all properties is maintained at the town level, a subset of data about each property is reported every year to the county level, and a smaller set is reported annually by the counties to the state. The same data also has great utility for many secondary uses. These secondary uses represent the latent economic and social value of the data, beyond the value of its primary use in tax administration. The purpose of the study was to reveal this latent value, identify barriers to its realization, and make recommendations for policies and practices that would encourage its development.

The study data consist of 35 interviews plus official publications and web sites prepared by federal, state, local, nonprofit, and private sector organizations. Interviewees were selected from five demographically, operationally, and geographically diverse counties in New York State to provide an understanding of the range of logical, purposive flows of data from original data collectors to other organizations and users. The study began with the Real Property Tax Official in each County who was asked to suggest people to be interviewed who use or otherwise interact with that county's parcel data. State agencies that collect and use parcel-level data as a major part of their responsibilities were also interviewed. Tape-recorded face-to-face or telephone interviews covered the following topics from the perspective

of the organization represented by each interviewee: organizational mission and programs; collection, dissemination, access, and uses of parcel data; value of parcel data to the organization; data management, preservation, and sharing methods; data forms and formats, data flow among organizations; and associated costs; issues and barriers. The data were analyzed using qualitative methods to identify similarities and differences among users and uses of this information, to identify barriers to value creation, and to map the flow of data among the organizations that produce the data and those that access and use it for secondary purposes.

2.1 Definitions, Uses, and Value of Parcel Data

The study began with an attempt to find a common definition of parcel data, however the interviews showed that different users apply quite different definitions. Interviewees usually defined “parcel data” by describing how they use it to support their work. For example, planning departments said that parcel data encompasses the parcel identification number, zoning codes, actual uses, boundaries, and physical characteristics. By contrast, real property tax administration officials referred to parcel data as parcel identification and location, information about structures, the parcel owner, and the assessed value. Taken as a whole, parcel characteristics identified by the interviewees fell into eight categories: identification, location, ownership, occupancy and use, structures and improvements, taxation, physical geography, infrastructure, and taxation. Each category comprises a set of detailed attributes. For example, the category “location” includes ‘parcel address’, which can be descriptive (i.e., 123 Main Street) or spatial data (i.e., a set of coordinates), or both; ownership attributes identify the owner and historical information about ownership.

Public, private, and non-profit organizations, as well as individuals use parcel data for diverse purposes. Although each use is different, all rely in some substantial way on the core parcel data collected at the municipal level. The following highlights illustrate this broad range of primary and secondary uses, each of which delivers a form of economic or social value to the state or community.

Real property assessment and taxation. Parcel data is the foundation for real property tax administration, which is the primary use of this data. Municipal assessors collect specific data elements to establish fair market value for every property. County real property offices use this data to create countywide tax maps and assessment rolls. The state real property agency uses the data to create final tax rolls and equalization rates that place local tax calculations against a statewide standard.

Buying and selling private land. Private land sales occur daily in every locale. Some are single transactions between two parties, while others are more complex and involve multiple individuals and organizations. The most important documentation of a land sale is establishing and recording the ownership rights to a parcel.

Directing emergency response. Emergency response centers use parcel data for almost every incident including attribute categories such as identification, location, and structural and improvement data. This information helps with emergency routing,

can instruct personnel on how to enter a property safely. and identify who should be notified to turn off utilities like gas or water.

Transportation routing. State and local roadways change frequently, affecting several different sectors. Parcel identification, location, and ownership information is used to handle routing of oversized truckloads to avoid low bridges and limited access highways, create and update bus routes for school districts, and determine priority routes for snowplows. Utility companies use parcel data to coordinate maintenance fleets using infrastructure data, structures and improvements, and parcel location data.

Facilities siting. Municipalities use location information to plan for growth and to increase their ability to attract businesses and jobs. Companies wishing to establish or relocate often have very specific site requirements. Using GIS capabilities and parcel identification, ownership, infrastructure, and physical geography data, localities can provide options that help them compete for these businesses.

Planning and prioritizing environmental initiatives. Environmental restoration of old industrial sites, brown fields, or wetlands are a major economic focus for local governments. For these projects, local priorities must be combined with data about parcel identification, location, physical geography, occupancy and use, and surrounding infrastructure, often in the context of a GIS system.

Infrastructure management. Infrastructure, parcel identification, ownership, location, occupancy and use data, as well as structures and improvements data are all used to plan rights of way, and changes, improvements, or additions to utilities such as electricity and water supply.

2.2 Stakeholders and their interests

The study identified a wide range of individuals and public, private, and non-profit organizations that use parcel data. The main stakeholder groups include:

- Assessors – municipal (city or town) officials who collect parcel data.
- County Real Property Tax Services (RPTS) – the hub for real property tax administration at the local level of government.
- State Office of Real Property Tax Services (ORPS) – the state-level office designated to guide tax administration and pursue statewide tax equity
- Other Local Agencies – such as planning departments, emergency response services, public health agencies, and water authorities.
- Other State Agencies – government organizations such as transportation, homeland security, environmental protection, and health and human services.
- Data Re-sellers – private companies that add value to public data and sell it to other entities such as insurance companies, real estate brokers, or consultants.
- Private Sector Users – such as realtors, utility companies, and engineering firms.
- Non Governmental Agencies and Community Groups – provide services such as environmental planning and economic development and civil society organizations such as those that serve senior citizens.

- Property Owners and Occupants – those who own and pay taxes on parcels and people or businesses who live or work on a parcel they may or may not own.

Every interviewee represented an organization that used parcel data to perform particular functions. Each organization's intended use determined the specific data attributes needed. For instance, an emergency response organization must have accurate and up-to-date occupancy and structure information in order to know the number of individuals living on a property and the location and placement of a building; but it may not need to know who owns the parcel or the details of the tax history. Alternatively, an environmental protection organization may need detailed location and physical geography data for a specific region but may not need parcel identification numbers or the sales data. However, the interests of these groups coincided with respect to their mutual desire for accuracy, timeliness, and consistency of parcel information as well as strong consensus among most secondary users for easy one-stop access to authoritative data sources in a variety of formats.. Table 1 shows both the areas of common agreement and the areas of divergence among these key stakeholder groups.

As the table shows, all stakeholder groups have a strong interest in high quality data, which they generally characterized as data that is factually accurate, up-to-date, and consistent from time to time and place to place. These general agreements, however, masked a great deal of variation. The interviews demonstrated that accuracy, timeliness, and consistency were important to everyone, but at different levels and for different reasons. For example, an engineering firm planning a residential subdivision may define "accurate" as highly detailed survey-quality information. By contrast, a town attempting to designate a rough boundary for a new municipal park would say an area bounded by certain streets is "accurate" for this purpose.

Table 1. Agreement and divergence among stakeholder interests								
Stakeholders		Interests						
	Type of user	Consistency	Accuracy	Timeliness	Low or no cost retrieval	Online access	Choice of format	One authoritative source
Assessors	primary	X	X	X	X			
County RPTS	primary	X	X	X				
State ORPS	primary	X	X	X	X	X		
Other Local Agencies	secondary	X	X	X	X	X	X	X
Other State Agencies	secondary	X	X	X	X	X	X	X
Data Resellers	secondary	X	X	X		X	X	
Private Sector Users	secondary	X	X	X	X	X	X	X
NGOs & Community Groups	secondary	X	X	X	X	X	X	X
Property Owners & Occupants	secondary	X	X	X	X	X	X	X

Most secondary users also strongly preferred that parcel data be available in electronic form, online, from one authoritative or trusted source, and in a variety of formats which they can select from to meet their particular needs. These features add convenience, flexibility, and efficiency to information search, access, and use. They also add confidence that the data is authentic and well-documented so it can be used in appropriate ways. However, these interests are typically less important or not shared by the primary users, the same organizations that collect the basic data – assessors and county and state real property tax offices – because their needs are met by their own internal activities, organized in a way that supports their main missions.

2.3 Issues associated with public access to and use of parcel data

Almost every person interviewed expressed a desire for better quality data. Most users spent considerable resources obtaining, improving, and standardizing parcel data before they were able to use it for their own needs. Much of this cost is associated with a need to supplement, verify, correct, or integrate basic data collected by others. Even when the data they begin with is of high quality, however, it may not be sufficiently detailed or readily comparable with other sources, or derived from systems that are technically compatible.

Interviewees also mentioned incompatible technologies used by different counties as a barrier to more effective data use. For example, not all parcel maps use the same mapping projection, which is a mathematical model for converting locations on the

earth's surface in a way that allows flat maps to depict three dimensional features. Although some technologies convert files originating from different mapping projections easily, there are still others that do not preserve the integrity of shape, or the accuracy of area, distance, or direction. To the extent that comparable data cannot be compiled for the whole state or for regions larger than towns or counties, potential value from data use by both government and private users is diminished or lost.

Data management issues were also prominent. Data management associated with the primary purpose of real property tax functions works in a relatively standard way all over the state. However, data management outside of tax functions varies widely and the typical arrangement is best described as ad hoc. Users of parcel data often make many individual data requests to different data sources. A few local governments have data management strategies in place that address the needs of external users, but in most places requests are handled on a one-by-one basis. At the same time almost no feedback mechanisms exist between data users and data collectors, so that the investments that users make in data improvements are not fed back into improvements in the original data sources. There is a notable absence of update and feedback mechanisms in the typical data flow which nearly always goes in only one direction – from the data source to a requester. Many users who obtain data from municipal, county, and state sources often find and correct errors as they use it, but these users are neither expected nor allowed to return data corrections, enhancements, or other improvements to the data sources. Consequently one clear overall benefit of use, data improvement for everyone, is never realized. Instead, when users obtain periodic updates from their data sources, they can actually make the situation worse because the data they have improved for their own use could be overwritten by some of the same old errors that still exist in the source files. Their difficult choice is then to forgo the updates in order to keep their own corrections, lose some of their corrections in order to obtain updated files for other records, or engage in very costly and time consuming matching and integration processes .

3 Discussion

The current emphasis on opening US government information to the public plus the evolving capability of technological tools for doing so, offer many opportunities to satisfy the value-creation goal of greater government transparency. Public and political expectations for the success of information-based open government strategies are high but as the case above illustrates, significant challenges remain for making government data suitable for uses that generate social and economic value. The parcel data case reveals the complexities of these initiatives by taking a comprehensive view of a single data resource and all of its primary and secondary uses and users. Multiply the findings in the case by thousands of data sets, and the magnitude of the challenge embodied in initiatives like Data.gov becomes more visible. Policies, governance mechanisms, data management protocols, data and technology standards, and a variety of skills and capabilities both inside and outside government are needed if

these information-based initiatives are to succeed in creating social and economic value beyond the government itself.

These issues have been explored in the research literature at least since the 1990s. This body of work (which comes from research in informatics, management information systems, and public management), offers a set of considerations and guidelines that can help these newer initiatives achieve better results. .

First, while clearly *more* information is being made available in recent efforts such as Data.gov, the case study above amply demonstrates that the problems of diverse user needs and capabilities [6, 14], the limitations of internally-oriented data management techniques [2, 9,], untested assumptions about information content and accuracy [1, 13], and issues associated with information quality and fitness for use [14]. All of these remain barriers to value creation.

Within mainstream public administration research, the impacts of information and knowledge have received relatively little attention [8] particularly when considering the salience of information for open government. However, research findings from other fields indicate that information-based initiatives are fraught with challenges. Political science, information science, and digital government research all shed some light particularly in the areas of information quality, system design, management, and information sharing. This body of work has shown that high-quality data should be not only intrinsically good, but also contextually appropriate for the task, clearly represented, and accessible to users. In other words, it needs to be “fit for use” [15]. The same information may be fit for some uses, but completely inappropriate for others that have different temporal, security, granularity, or other requirements. Users may need to make choices or trade-offs among these characteristics [1, 12], but they need good data descriptions to help them decide. In order for users to assess data quality, they need to understand the nature of the data and because data producers cannot anticipate all users and uses, the provision of good quality metadata is as important as the quality of the data itself [6]. Moreover, e-government interoperability and information integration research demonstrates high sensitivity to the nature of information and quality factors such as comprehensiveness, authoritativeness, trustworthiness, and perceived value as determined by the information seeker [9].

Research on performance measurement, shows how unrealistic assumptions and popular rhetoric about the nature of information can mask the difficulties inherent in information-based strategies for open government. These assumptions include the commonly-stated beliefs that more information leads to better governance and that information is objective, neutral, and readily available [13]. From a political science perspective, Meijer (2009), for example, asserts that computer-mediated transparency has several characteristics that can actually threaten public trust such as its unidirectionality (i.e., it is not interactive), decontextualization (i.e., it is removed from shared social experience), and its overly structured form (i.e., it is predominately quantitative).

Likewise, research on making government data available to public users has identified the need to understand the processes that produce the data and the development of new skills and services to support data users. Research on designing public access programs pulls together many of these findings into a complex, multi-

dimensional framework of information access considerations [6]. All of the foregoing research is useful in understanding the nature of public information resources and the challenges of making them available for use outside the government. The case study above goes a step deeper by tracing the path of a specific information resource, land records and associated parcel data, from initial data collection, to its primary use in tax administration, to a wide variety of secondary public and private uses by many external stakeholders. As such, it gives us a picture of the issues associated with the nature, use, and value of the government information over time and from multiple perspectives.

4 Practical recommendations and future research

Some of the challenges of information-based open government strategies, can be understood as technical problems addressing information storage, access, inquiry, and display. Another way to understand the challenges are as management problems such as defining the rationale and internal processes of data collection, analysis, management, preservation, and access. The challenges also represent policy problems including examining the balance and priority of internal government needs versus the needs of secondary users, the resources allocated to serve both kinds of uses, and the criteria for assessing their effectiveness and public value [5]. Consequently if open access to government data is to generate its potential value for society, then government information policies and practices need to be better aligned with the needs of secondary users.

For most government agencies, providing information for public use is an extra responsibility that may compete for resources with the demands of mission-focused operations. As our case study illustrated, vast amounts of useful information are contained in government data systems, but the systems themselves are seldom designed for use beyond the collecting agency's own needs. With few exceptions, making data holdings available to the public in a meaningful and useable way is a new responsibility of government agencies that will need thoughtful investments in skills, tools, and policies, as well as some changes in processes and practices. One needed practice improvement is the creation of formal feedback mechanisms that connect data users to data sources. Feedback from users could lead to ongoing data improvement as users discover and correct errors in the data. By providing the opportunity and a formal mechanism to communicate data errors and enhancements back to the data sources, improvements in the overall quality and integrity of the data can benefit all future users, including the government itself..

Within government organizations, professionals will need to develop several kinds of skills to support public information access and use. Technical skills for information management, display, and integration need to be augmented by communication skills that serve the needs of a wide user community. New roles may need needed as well to coordinate agency-level and government-wide programs of information dissemination and user support services.

Thinking about government information as a public resource reveals not only its potential public value, but also its vast complexity. Understanding the public value proposition of information-based transparency strategies requires us to look at them from multiple perspectives. The government's own perspective brings into focus information stewardship and management considerations that assure good quality data and metadata that is accessible and usable by people with different intentions and different capabilities [5]. It also encompasses policy considerations, such as giving data access programs appropriate priority for funding and determining how much the government will engage in value-added services itself and how much it will leave to private providers. From a technology perspective, semantic interoperability and data presentation and visualization tools are just two of many topics that need research and which could be studied in the context of research-practice partnerships.

In addition, because the community of information users is clearly not homogeneous, their different needs and capabilities cannot all be served by the same kind of information or the same forms of information [6]. More complete stakeholder analysis could lead to better understanding of users' needs for interfaces, services, and analytical tools. Careful stakeholder analysis and engagement could generate broader appreciation for the many ways different people think about, use, and benefit from government data. Such an appreciation could set the stage for active collaboration and joint investments. This is not to say that every use can be predicted in advance, but that a thorough assessment of the needs and capabilities of a wide variety of users could lead to well-articulated standards for data description and formats, high-quality data, and good data management practices that serve many different needs.

Finally, information-based open government strategies constitute a natural experiment in value creation. Evaluation research could help identify which kinds of information content or formats generate different kinds of value for different communities of interest as well as whether information-based strategies as a whole actually lead to not only economic and social benefits, but also to a stronger democracy.

References

1. Ballou, D., & Pazer, H. (1995). Designing Information Systems to Optimize the accuracy-timeliness tradeoff. *Information Systems Research*, 6(1), 51-72.
2. Ballou, D. & Tayi, G. (1999). Enhancing data quality in data warehouse environments. *Communications of the ACM*, 42(1), 73.
3. Bellamy, C., & Taylor, J.(1998). *Governing in the information age*. Philadelphia: Open University Press.
4. Cook, M., Dawes, S., Helbig, N., and Lishnoff, R. (2005). Use of Parcel Data in New York State: A Reconnaissance Study. Available at http://www.ctg.albany.edu/publications/reports/use_of_parcel_data.
5. Dawes, S. (2010). Information policy meta-principles: stewardship and usefulness. *Proceedings of the 43rd Hawaii International Conference on System Sciences*.

6. Dawes, S., Pardo, T.A., & Cresswell, A. M. (2004). Designing electronic government information access programs: A holistic approach. *Government Information Quarterly*. Vol. 21, pp. 3-23.
7. Dawes, S., Helbig, N., & Cook, M. (working paper). Governance of Public Information Resources: Lessons from a field study of land parcel data. Available upon request to the authors.
8. Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2005). New public management is dead: Long live digital-era governance. *Journal of Public Administration Research and Theory*, 16, 467-494.
9. Klischewski, R. & Scholl, H. J. (2006). Information quality as a common ground for key players in e-government integration and interoperability. *Presented at 39th Hawaii International Conference on System Sciences*, Big Island, Hawaii.
10. Meijer, A. (2009). Understanding modern transparency. *International Review of Administrative Sciences*. Vol. 75, pp. 255-269.
11. OMB (US Office of Management and Budget). (2009). Open Government Directive. Available at <http://www.whitehouse.gov/open/documents/open-government-directive>.
12. Pipino, L., Lee, Y., & Wang, R. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211-218.
13. Radin, B. (2006). *Challenging the performance movement: Accountability, complexity, and democratic values*. Washington, D.C.: Georgetown University Press.
14. Wang, R., & Strong, D. (1996) Beyond accuracy: What data quality means to data Consumers. *Journal of Management Information Systems*, 12(4), 5-34.