



Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Government Information Quarterly 21 (2004) 3–23

**Government
Information
Quarterly**

Designing electronic government information access programs: a holistic approach

Sharon S. Dawes,* Theresa A. Pardo, and Anthony M. Cresswell

Center for Technology in Government, University at Albany, State University of New York, 187 Wolf Road, Albany, NY 12205-1138, USA

Available online 13 December 2003

Abstract

That electronic government information repositories are growing in number, use, and diversity is one manifestation of the emergence of e-government. These information-centered programs both shape and respond to user demand for electronic government information as computer-mediated user access has displaced traditional staff-mediated access. These programs are no longer concentrated in statistical agencies but increasingly are offered by a wide array of mission-driven operating agencies to complement their other services. This study identified the design dimensions of electronic information access programs by examining mature existing programs. These dimensions address users, uses, organizational capabilities, data characteristics, and technology. The study then explored the application and interdependence of these dimensions in three efforts to design and develop new access programs. The study produced an empirically based, testable model of observable dimensions that shape the cost, complexity, and potential performance of these programs. In addition, the article offers government managers some insight into the practical implications they will face in designing and operating electronic information access programs.

© 2003 Elsevier Inc. All rights reserved.

1. Introduction

A look at the e-government agendas of many states and the federal government shows that electronic information access strategies and programs are a consistent feature of e-government initiatives. The strategies and programs of interest here are those intended to provide

* Corresponding author. Fax: +1-518-442-3866.

E-mail address: sdawes@ctg.albany.edu (S.S. Dawes).

new or enhanced electronic means for acquiring information from a government agency. The Bush Administration E-government Agenda, for example, consists of 24 initiatives grouped into four portfolios.¹ Each of the portfolios—Government to Citizen, Government to Business, Government to Government, and Internal Efficiency and Effectiveness—includes initiatives that depend on programs of electronic information access. A notable characteristic of these initiatives is their placement in the operating or mission agencies of government (such as the Social Security Administration and the Environmental Protection Agency [EPA]), rather than in the so-called information or statistical agencies (among them the Bureau of Labor Statistics and the Census Bureau).

This trend is placing new demands on agency leaders for investments in information access programs that complement their service or regulatory programs. This demand for direct access to government information, from both inside and outside government, is influencing the design and management of these information services. They are becoming increasingly focused on electronic data and records as the format desired by users.² They are also shifting from staff-mediated modes of access to user-directed computer-mediated access, now made possible over the Web.³ As programs move away from traditional concepts of centralized control and physical custody of information, public managers discover a need for different skills, more strategic investments in technology, a more distributed or collaborative method of management, and a broader understanding of users, their needs, and their capabilities.⁴ In support of efforts to respond to these new demands, the National Archives and Records Administration (NARA), the National Science Foundation (NSF), and other granting organizations are increasing investments in research and practical guidance for these efforts.⁵

2. Lessons from the field

Three cases reported in the public administration literature highlight the challenges facing those responding to demands for electronic information access programs. The experiences of the Vermont Human Services Agency succinctly summarize the interwoven organizational, technical, political, and data issues.⁶ In creating its “Community Profiles,” the agency set out to gather and publish outcomes and indicators on many aspects of health, education, economic vitality, safety, and welfare. In doing so, it faced six major challenges: devising a working definition of “community” grounded in meaningful local terms, creating an integrated database that bridged both “stovepipe” legacy systems and organizational turf, dealing with the issues surrounding small numbers in statistical presentations, designing a user-friendly presentation format, providing a context for useful comparisons among localities, and fostering informed and continuous use of the data. These six issues present an array of considerations focused not only on the public policy goal of the initiative, but also organizational issues, data management and analysis problems, and user capabilities and needs.

Another example adds economics to the list of considerations. In Washington State, a Public Policy Information Access Task Force considered the conundrum of providing “free” access to government information (i.e., at no cost to the customer) while facing the need for revenue to cover the underlying costs of developing and operating the necessary electronic information services.

Analysts noted that such programs should distinguish between content (which they agreed should be free) and the delivery mechanism (which must be paid for through tax levy or user fees).⁷

The EPA's Toxics Release Inventory (TRI), a result of the first "right-to-know" legislation, was designed to help citizens and communities understand local environmental risks. The TRI suddenly put EPA in the information access business. Initially implemented in 1989, the TRI today offers data, search capabilities, public online dialog about issues, and links to a variety of other publicly accessible EPA databases. The program is a complex and expensive one for EPA to administer, but it is credited with a profound impact on environmental quality and policy in the United States.⁸ In this example, we see the mutually reinforcing interdependence of an information access service with the operational mission of the agency.

3. Current research efforts

These and similar experiences provide the foundation for a comprehensive framework to guide the design of the electronic access programs that are so visible in the goals of e-government. Yet research on electronic information services and repositories is almost entirely focused on deepening our knowledge of the individual components of these complex and interconnected enterprises. A review of the past two years of the *Journal of the American Society of Information Science and Technology* (JASIST), for example, offers a host of articles on related topics. Many are associated with technology, such as search engine performance,⁹ Web media agents,¹⁰ and visualization tools.¹¹ Other articles focus on the characteristics of users including their cognitive styles,¹² motivation,¹³ searching behavior,¹⁴ and individual differences.¹⁵ Still others deal with information content and presentation, such as judgments about Web information quality,¹⁶ credibility,¹⁷ and relevance.¹⁸ All of these are important topics, but taken individually, they do not help us address the combination of factors that a program operator must contend with in the course of daily operations—and that a program designer must consider from the outset of planning.

By contrast, in the research literature devoted to government information policies and programs, we find a rich body of work about how the networked environment affects the principles, policies, and structures of the traditional government programs and organizations whose main purpose is information management and dissemination. These agencies are critical to the government's ongoing commitment to public access to government information.¹⁹ Even so, organizations like the Government Printing Office or the Federal Depository Library Program²⁰ are only a small part of the vast array of information repositories and services embedded within all kinds of public agencies at every level of government—and these others are given almost no attention by either information science or public management researchers.

4. Developing a holistic framework

The research reported here forms a bridge among three areas: the case studies reported in the public administration literature, the conceptual debates in the government information

literature, and the research on individual elements of electronic access prevalent in the information science literature. It takes a holistic view, attending to users, uses, policies, organizational capacity, data characteristics, and technology, in the context of a complete program in which these components interact. Our research suggests that these interactions are perhaps more important to the design of an effective electronic access program than any of the individual elements.

The findings reported here are the result of two complementary streams of research conducted over a period of about six years, from 1995 to 2002. The first stream was embodied in a study funded in part by the Electronic Records Research Program of the National Historical Publications and Records Commission (NHPRC).²¹ The overall objective of the research was to determine whether a standard approach to access program design was possible across a broad range of program purposes and structures.²² This research relied heavily on interviews with the administrators of 22 national and state government agencies, nonprofit and for-profit organizations, and providers of free and fee-paid data repositories in the United States and Europe. The interviews gathered information on effective practices and issues faced in operating programs as well as program policies, management strategies, technology tools, and resource-sharing models.

The investigators categorized the interview data by topic and organized it into clusters of concerns that appeared to be present in all types of repositories regardless of purpose, size, or sponsorship. These concerns played out in different ways in different settings, but they were clearly present in one form or another in each program. We came to call these clusters the “dimensions” of electronic access programs.

The second stream of research includes the development and evaluation of three prototype data repositories built for government agencies in New York State. In each of them, the main goal was to increase the availability and use of information for operations, planning, evaluation, and decision making by providing a Web-based program of access to electronic data and records. These prototypes differed from one another in many respects, as did the programs that were the subject of the earlier interviews. The interview data, however, revealed a consistent set of considerations for the designers and operators of data repositories, despite the differences in program purpose and structure. The prototypes represented an opportunity to test the validity of the interview findings.

- The New York State Geographic Information System (GIS) Clearinghouse, contains selected spatial data sets contributed by state, local, federal, and academic sources for the shared use of the GIS community in the state. It also contains metadata and access instructions about a large number of additional data sets that are maintained in custodial agencies. In addition, the Clearinghouse offers information about GIS tools, specialized training, and cooperative analytic and data development projects.²³
- The Homeless Information Management System (HIMS) prototype, initiated by the New York State Bureau of Shelter Services is an integrated data repository of service information provided by local governments, nonprofit shelter and service providers, and

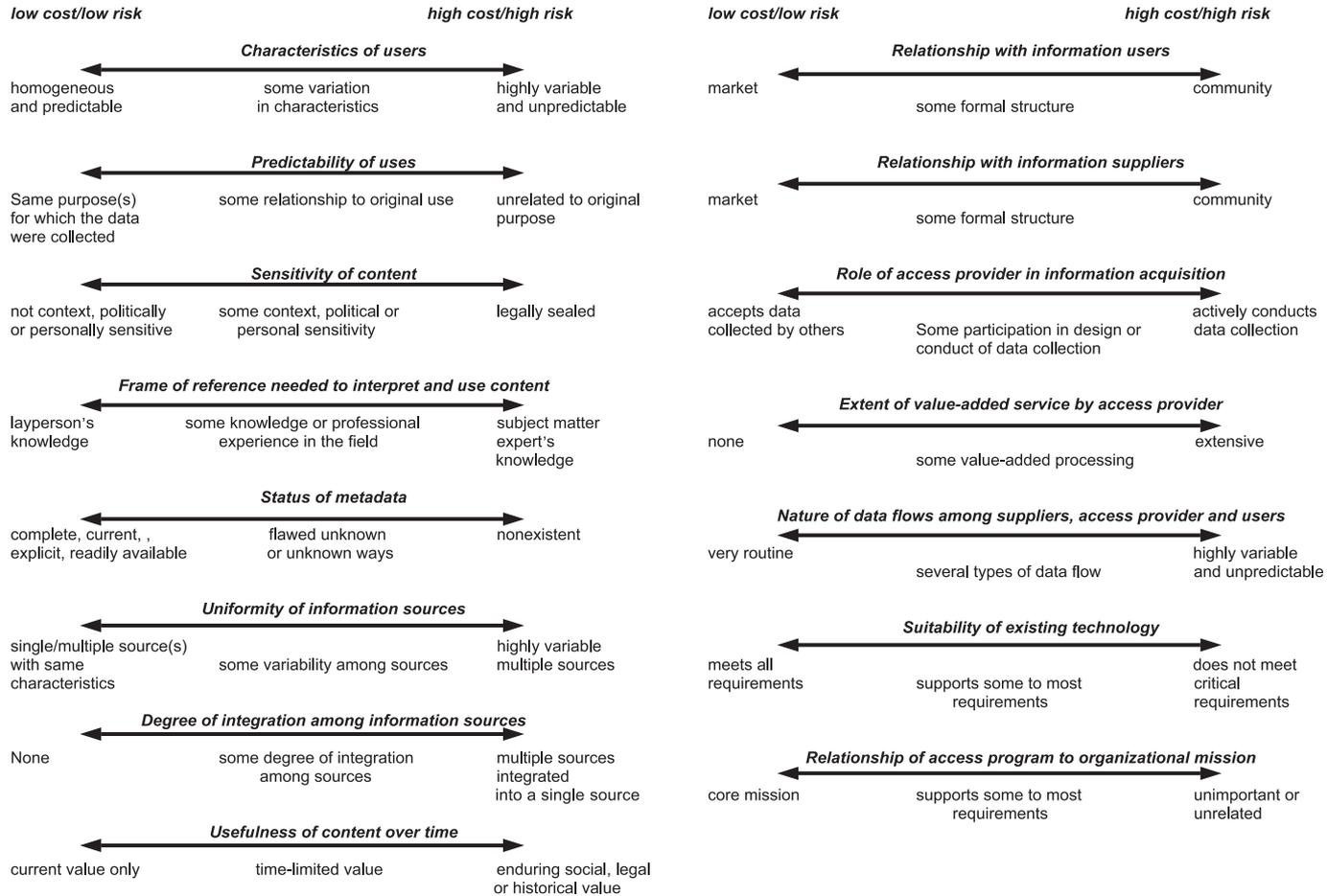


Fig. 1. Diagnostic and planning dimensions of electronic information access programs.

state agencies for their mutual use in program evaluation.²⁴ The prototype repository contained client-oriented data drawn from the case management systems and records of the provider agencies. This was not intended to be a publicly available repository, but one built to support the work of the involved agencies.

- The Kids' Well-being Indicator Clearinghouse (KWIC), a project of the Council on Children and Families, addresses the need for more rapid and timely public access to health and well-being indicator data about children and draws its data sets from the administrative and program statistics of thirteen state agencies.²⁵ The Web-based Clearinghouse replaces an annually published paper book of statistical tables.

The GIS prototype project preceded the NHPRC study; the HIMS and KWIC projects were contemporaneous. All three projects continued to be tracked through 2002 in an NSF-funded study on knowledge networking. These projects provided the practical experiences against which we tested the dimensions that emerged from the interview data.

5. The dimensions of electronic access programs

Fifteen separate but interacting concerns about electronic access programs emerged from the interviews (Fig. 1). Each one addresses a key program design factor. Eight of these dimensions address information users, suppliers, content, or use. Seven additional dimensions consider aspects of the access program and its organizational context. Together, they provide a broad overview of the factors an access provider must take into account in designing a new program or improving an existing one.

The users-uses-suppliers-content dimensions are characteristics of users, predictability of uses, sensitivity of content, frame of reference needed to interpret and use content, status of metadata, uniformity of information sources, degree of integration among information sources, and usefulness of content over time.

The access program dimensions include structure of relationships with information suppliers, structure of relationships with information users, involvement of access provider in original data collection, extent of data analysis or other manipulation conducted by the access provider, nature of data flows, suitability of existing technology, and relationship of the access program to overall organizational mission.

Each dimension is briefly described below. For each dimension, the left anchor end represents the least problematic or least expensive situation that can be expected. The right anchor end represents the most problematic or expensive. The interval between the anchors represents a continuous range of situations between these extremes.

5.1. Dimensions related to users, uses, suppliers, and content

The first set of dimensions addresses the content of a repository as well as the people and organizations that contribute to it or use it.

5.1.1. User characteristics

The first dimension deals with the degree to which user characteristics are consistent and predictable (see Fig. 2). User characteristics include their ability to understand the data content, its limitations, and the conditions under which it was collected. These also include their data-handling and analytic skills, technological capabilities and tools, and interests in the data content and what it can be used for. Registered or licensed users who provide information about themselves would fall on the left side of the continuum. Anonymous members of the general public, whose characteristics vary most and are least predictable, serve as the anchor at the opposite extreme. The middle range includes users whose characteristics can be partially known or predicted. For example, even if they do not register, users of a repository devoted to organic chemistry are likely to be chemists, medical professionals, science teachers, science or medical students, or people with similar interests or knowledge. A repository of popular music might attract users with a much greater variety of characteristics.

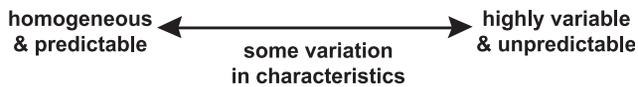


Fig. 2. User characteristics.

5.1.2. Predictability of uses

Electronic records and information may be applied to uses that are very close to or far removed from the uses for which they were created. The degree to which use can be predicted is therefore a key dimension to consider (see Fig. 3). At one end of the continuum are the uses for which the data were originally collected or the records were created. At the other end are uses that have no relationship to the original purpose for data collection. Uses with closer connections to the original purpose lie at points in between. For example, real property records are created to document the history of ownership of land parcels. They are also used as the basis for local property taxes, a related use. The same records could be used to identify high-income neighborhoods for a marketing campaign—a use that is not at all like the original.

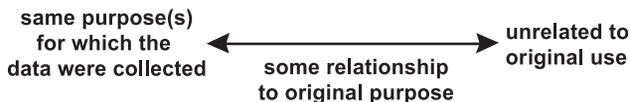


Fig. 3. The predictability of uses.

5.1.3. The sensitivity of content

The subject matter or content of a record will have characteristics that allow it to be placed along a continuum that has, on the one extreme, factual content that is not controversial and, on the other, content that is so sensitive that laws constrain its use (see Fig. 4). Between these

two extremes lies information of varying sensitivity that must be handled by a range of appropriate policies or management tools. Low-sensitivity content might be a daily record of air temperature and wind velocity. Increasingly sensitive content includes campaign contributions, divorce decrees, or adoption records (which are generally sealed).

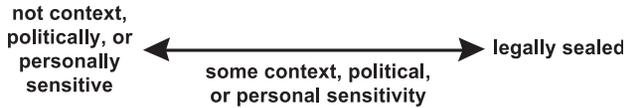


Fig. 4. The sensitivity of content.

5.1.4. The frame of reference needed to interpret and use content

Information content can vary widely in its need for an expert frame of reference. At one end of the continuum lies information that is readily understandable by a lay person (see Fig. 5). At the other is information that cannot be used reliably without the knowledge and experience of a subject matter expert. In between is information of increasing nuance or complexity that requires increasing amounts of contextual knowledge or expertise for use. A lay person generally has the background knowledge to make good use of a library catalog, news stories, or straightforward numerical data in tables or graphs. It takes more background and training to interpret scientific research reports, complex statistical presentations or inferences, or highly technical information pertaining to various professions or scientific disciplines.

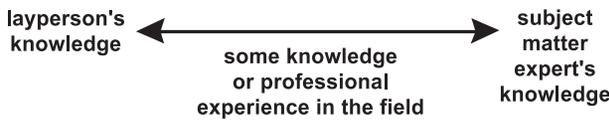


Fig. 5. The frame of reference needed to interpret and use content.

5.1.5. The status of metadata

Variation in the methods and quality of metadata are represented in this dimension. Metadata (i.e., information about the information in an access repository) can be characterized by its completeness, accuracy, explicitness, currency, and availability to users (see Fig. 6). At one end of the continuum is metadata, which has been made explicit, is current and complete, accurate, and readily available to potential users of the data. At the other end of the continuum, metadata is nonexistent. At various points along the continuum, metadata is more or less flawed or incomplete, with different levels of understanding about its shortcomings. For example, metadata may describe when the information was collected, how, and by whom, but it may not provide definitions of key terms or explain how those definitions may have changed over time.

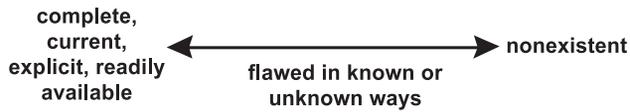


Fig. 6. The status of metadata.

5.1.6. The uniformity of information sources

This dimension refers to the uniformity of important aspects of the information sources being made available for use (see Fig. 7). These characteristics include physical format, original purpose, method of data collection, and the meaning of various data elements. At one end of the continuum are data or records that come from a single source or from multiple sources that are exactly alike in these characteristics. At the other end are records that come from multiple sources with disparate characteristics. For example, taxpayer instructions for withholding income taxes from their wages are collected by every employer in the country, but in a very uniform way, using a standard form provided by the federal government, the W-4. By comparison, case records about social services to individual clients vary widely from one service agency to another.

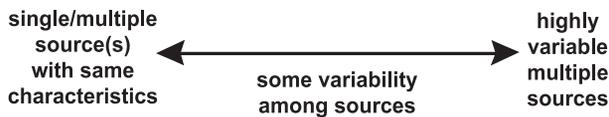


Fig. 7. The uniformity of information sources.

5.1.7. The degree of integration among information sources

If the repository is composed of information from multiple sources, the degree to which that data will be integrated is an important consideration (see Fig. 8). At one end of this dimension are repositories that maintain multiple sources as separate entities. The National Spatial Data Infrastructure, for example, collects or points to a virtual collection of many separate spatial data sets. At the other end are repositories in which information from multiple sources is integrated into a comprehensive single secondary source, such as a data warehouse. For example, our HIMS prototype created a single data repository about homelessness composed of data extracted from scores of government and nonprofit agencies, which were then organized and merged into a single data resource.

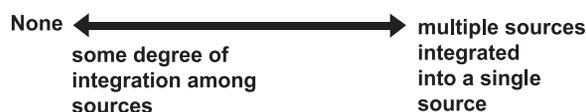


Fig. 8. The degree of integration among information sources.

5.1.8. *The usefulness of content over time*

This dimension represents the variation in the pertinence and value of information content over time (see Fig. 9). Some information has only current or short-lived usefulness; other information may be of enduring social, legal, or historical import, and is thus worthy of indefinite preservation. At points in between are records whose value to users diminishes over a medium- to long-term interval. For instance, correctional institutions maintain various records essential to accomplish their work, including logs and prisoner case files. A log might document the rounds a prison guard makes on a daily basis, and these have value in the short term, proving that the guards fulfilled their responsibilities or providing a way to estimate when an unwitnessed activity (an escape, a fight, a suicide) took place. Shortly after the production of these records, their value begins to decrease steadily until it disappears. On the other hand, prisoner case files might have a good deal of continuing value. These document each prisoner, including dates of incarceration and release, age, ethnicity, offense, behavior in prison, medical condition, and others. The prison uses these records to track prisoners, and these records have value to the prison during the time of incarceration and for a short period afterward. In addition, these records may also have permanent value as a way to document changes in prison conditions, population characteristics, and other historical trends.



Fig. 9. The usefulness of content over time.

5.2. *Dimensions related to organizational structure and context of the access program*

The foregoing dimensions pertain to the information content of an access program and to the use of that content. The following dimensions pertain to the organizational environment in which the program will operate.

5.2.1. *The relationship with information user*

The structure of relationships between the access provider and information users could be described as varying from a simple market kind of mechanism, to a more formal, rules-based arrangement, to a more community-like relationship (see Fig. 10). Market relationships are based on low transaction costs, mutual exchange, short-term involvement, and little or no need for shared identity or values.²⁶ Libraries operate under this model. Bureaucratic, legal, or contractual relationships are based on formal agreements or policies, and characterized by longer-term involvement, and higher costs to establish and maintain. Subscription services to online professional journals are typical examples. Community relationships are based on long-term familiarity and trust, with shared identity, values, and mutual interests. The establishment and maintenance of these relationships are more costly than the others. If a repository has different user groups, different relationship can exist with each of them.

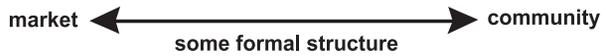


Fig. 10. The relationship with information users.

5.2.2. *The relationship with information suppliers*

The same relationship structures that apply to users also apply to suppliers of information (see Fig. 11). They range from simple market transactions between access providers and information suppliers (such as information brokers who buy state vehicle registration files to serve the information needs of vehicle manufacturers) to formal arrangements (such as the information that the US Securities and Exchange Commission requires of businesses) to complex community structures (such as voluntary data clearinghouses like the New York State GIS Clearinghouse). The most elaborate example of a community in our interviews was the USDA and the Cooperative Extension Service, where many resources are shared, and personnel move back and forth between organizational units and locations. As with users, access providers can have one type of relationship with one set of data suppliers and a different one with another.



Fig. 11. The relationship with information suppliers.

5.2.3. *The role of the access provider in information acquisition*

This dimension shows the extent to which an access provider plays a role in original data collection (see Fig. 12). At one end, the information creation work of the suppliers is independent from the work of the access provider. Examples are data libraries that do no original data collection, but accept the information resources that are created or collected by others. The Inter-university Consortium for Political and Social Research (ICPSR), for example, specifies standards for acceptance of data sets, but is otherwise largely unconnected with the work of the data suppliers. Typically, government archives also fall at this end of the spectrum. At the opposite end, the access provider plays a significant role in data collection or creation. The Central Archive, an institute of the Cologne Association for Social Research at the University of Cologne, collaborates with the research community to design data collection methods and metadata requirements for new data sets. Some access providers, such as the National Center for Educational Statistics (NCES) and the US Census Bureau, are the main or only data collection agencies for their repositories. These programs would also be at the high cost end of this dimension.

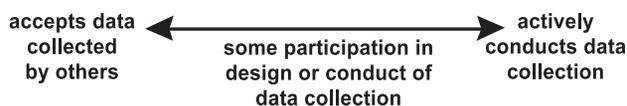


Fig. 12. The role of the access provider in information acquisition.

5.2.4. *The extent of value-added service provided by the access provider*

This dimension captures the extent to which the access provider performs work that changes or adds value to the information supplied by others (see Fig. 13). At the high cost end of this dimension, the access provider transforms and analyzes the data in substantial ways (e.g., aggregation, constructing indicators, statistical analyses, data mining). As a result, what is available to users is significantly different from or enhanced beyond the original sources. The changes in the data may involve format, content, and ways of presentation. Analytic products, such as reports or summaries, may be made available to users. The US Census Bureau, for example, provides not only basic census data, but many different subsets, analytic reports, and analysis tools for users. At the other end of the dimension, the provider does not transform or analyze the data in the repository, providing only access to the original data as received from suppliers. In the middle would fall programs that conduct modest value-added activities, such as providing a search capability or categorizing information according to theme or source. The New York GIS Clearinghouse is an example of this middle ground. It does not process the contributions of information suppliers, but provides several ways to search through the data, including by source, by theme, or by coordinates.

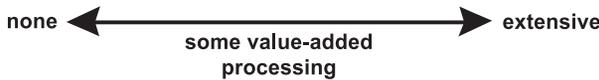


Fig. 13. The extent of value-added service provided by the access provider.

5.2.5. *The nature of data flows among suppliers, access provider, and users*

This dimension addresses the way in which information flows from the suppliers to the access program provider, and on to the end users (see Fig. 14). At one end, the flow of data into the repository and the demand for access to that data are unpredictable and variable in timing, volume, and other characteristics. A general government archive would be an example of midrange variability on this dimension. For a very routine case, both the nature and flow of the data into the repository would be consistent and predictable, as would be the demand for access by users. The Federal Deposit Insurance Corporation (FDIC) is an example of a routine program, with regular, rigid requirements for submission of data by banks and routine reporting and access provisions for a large volume of users.

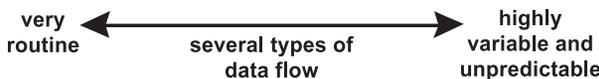


Fig. 14. The nature of data flows among suppliers, access provider, and users.

5.2.6. The suitability of existing technology

Electronic access programs necessarily rest on technology foundations. This dimension represents the degree to which the access provider's existing technology can support the desired access program (see Fig. 15). At one end are programs that already have sufficient infrastructure and technical support to operate with the desired features. On the other are situations in which existing technology does not meet even the most basic requirements. For example, an organization that does not have a Web server cannot offer a Web-based information repository to others without new resource investments. Security is an important factor here as well. Security technologies must be explored and understood to ensure that the appropriate technologies have been employed to provide the desired level of protection. The KWIC prototype is an example of a program that was necessarily built from scratch technologically, since the host agency initially had no technical staff nor suitable hardware or software.



Fig. 15. The suitability of existing technology.

5.2.7. The relationship of access program to the organizational mission

This dimension represents the degree to which the access program is central to the core business or mission of the provider organization (see Fig. 16). If the organization's primary purpose or mission is to provide access to information, it will have a less-problematic rating on this dimension. Archives and data libraries would be examples of such organizations. For the opposite end of the dimension, the access program would be considered a minor, unimportant, or even unrelated part of the overall organizational mission. This program will have to compete with other, higher priority activities for attention and resources. Certain advocacy organizations would fall in the middle of this dimension. They may collect or acquire information and perform analysis primarily to advance the policy agenda of the organization, and only secondarily to provide data to other users. At the more difficult end would be programs that are transitory or of interest only to certain individuals, but are not important to the organization as a whole or to its long-term strategies and goals.



Fig. 16. The relationship of access program to the organizational mission.

6. Interactions among the dimensions—lessons from the prototypes

While each dimension should be assessed independently at first, an understanding of their interaction adds crucial insight into possible priorities and options for program design. When the analysis indicates a problematic situation with one dimension, it is often possible to adjust others to compensate. The discussion below shows how different situations can be addressed by adopting policies or practices, setting limits, or establishing requirements. These brief discussions of the interdependence among dimensions are drawn from research on the prototypes described above. They are not intended to be exhaustive, but to illustrate key relationships and possible actions to reduce or eliminate problem situations.

6.1. Users

The more homogeneous and predictable the user population, the more focused the implementation of the program can be. Issues related to the sensitivity of the data may be more easily addressed when the user population is known and can be asked to agree to behave in certain ways. Predictability of the nature of use is also likely to be greater. Metadata can be developed to meet the known user community's needs rather than incurring the cost of developing metadata that is broad and detailed enough to serve an unidentifiable or general population. The HIMS prototype benefited from a homogenous, clearly identified user population, consisting of licensed shelter providers and the government officials they work with everyday. Because the users of the data were also the suppliers, they were strongly motivated to deal with data definitions and data quality issues. They also shared the same values with respect to confidential treatment of personal information and were able to develop policies and methods to assure confidentiality.

6.2. Uses

When information collected for one purpose is used for a different purpose, there is potential for misuse, misunderstanding, and misinterpretation. The intended use, the nature and skill of the user, and the quality of metadata are therefore highly interdependent. Users should be made aware of the limitations of the data as well as its potential value. The more removed secondary users and uses are from the original purpose for data collection, the more they will need ready access to complete, accurate, and timely metadata and perhaps some expert advice about data use. The GIS and KWIC prototypes had different experiences regarding uses. The GIS program designers focused almost entirely on metadata in their initial implementation. They did not have the expertise, the funding, or the community consensus necessary to gather hundreds or thousands of data sets into a single repository. Instead, they built a metadata repository that gave interested users standard descriptions of data holdings that were retained by their originators. As a result of this key design decision, the development team (which included many potential users) devoted much time and effort to creating standardized and comprehensive data descriptions that would allow users to determine if a particular data set might meet their needs and how it could be responsibly used.

While the GIS program was entirely new, the KWIC program was replacing an existing data resource—a printed book of statistical tables prepared annually by thirteen contributing state agencies and compiled and distributed by the Council on Children and Families. As a statistical publication, little attention had been paid to metadata. However, when the KWIC Clearinghouse was being designed to include the underlying data sets, the participants discovered that the source files were often so poorly documented that only experienced insiders could use them with confidence. Metadata creation became a top priority for the project, but since the hosting agency (the council) was not the data supplier, it needed to build new expectations and working relationships with the data providers to accomplish this crucial task.

6.3. Sensitivity of content

Politically, personally, and context-sensitive content will have a strong impact on design and implementation. More sensitive content will require more stringent governance and access policies, regular review of their effectiveness, and well-trained staff to handle the data appropriately. Sensitive content will also require the use of technical safeguards that ensure security and prevent improper access. The HIMS prototype experienced this issue most acutely. The design team discovered that providers of different kinds of shelter services considered different kinds of data to be highly sensitive. For most shelters, data that would reveal client identity was strongly protected, but for domestic violence shelters, the physical location of the shelter or client was the most sensitive information. Dealing with these nuances was crucial to the success of the program and was only possible when program and policy specialists worked together with system designers. The sensitive content of HIMS also required that it offer secure access and require the users to sign on and be authenticated before they could access the data. Neither the GIS nor KWIC programs needed this kind of security feature.

6.4. Needed frame of reference

When a high level of expertise or a particular frame of reference is needed to interpret and use information in a repository, planners need to consider the capabilities of expected users and enhance the quality and usability of metadata and user support accordingly. Designers could also consider repackaging the content to make it more suitable for less expert users or provide some ready-made analysis for the most common uses or questions. In the KWIC program, the frame of reference issues tended to center on the degree of expertise a user needed to have in data analysis and statistics. When the child indicator data were a book of printed tables, much of that expertise was built into the table presentations. The data analysts in the agencies who prepared the tables knew the places where the data were weak and how to handle the situations where data definitions or data collection methods changed from previous years. Moreover, each table was self-contained, pertaining to a single service or indicator without regard to the others. In the KWIC Clearinghouse, all these would change. Users would have access to the underlying data and would need the same information that the agency analysts had about the data. Moreover, users would have the ability to merge and compare data from different agencies and different services in ways that had not been done before. Some of these

comparisons would simply be invalid due to incompatible periods, geographic breakdowns, or inconsistent data element definitions in the different data sets. Designers had to build cautionary messages into the clearinghouse to advise users about possible comparability problems and even included some routines that would prevent users from conducting certain kinds of comparisons that would give seemingly useful, but invalid, results.

6.5. *Metadata*

The metadata dimension is a critically important one and one over which planners often have the most control. The quality, completeness, and user-friendliness of metadata can be adjusted in many ways to account and compensate for variations in source data, the needed frame of reference for responsible use, or the unpredictability of users and the uses they have in mind. The more a program's analysis tends toward the high or problematic end of each dimension, the greater the importance of good metadata. Metadata was crucial in all three prototypes. The GIS program needed the most extensive metadata for two reasons. First, users could not explore the data itself to determine if it would be worth working with for their purposes. They had to rely on metadata alone to make that first determination, so the descriptions needed to be detailed. Second, spatial data are extremely adaptable for many uses, including uses that have no relationship to the original reason for data collection. To make it possible for the New York Clearinghouse to join a growing national movement, portions of the national standard for geospatial metadata were adopted and used consistently by the data custodians to describe their data holdings.

For KWIC, metadata was essential to the overall strategy of the clearinghouse—giving users access to underlying data rather than to prepared analyses. No metadata standard existed for this effort, and none of the contributing agencies had a standard to offer. The creation of metadata became a major effort involving the Council for Children and Families and all the data suppliers. Because metadata was not deemed essential for the internal purposes of the contributing agencies, it tended to receive a low priority and took considerable time and effort to develop.

In the HIMS program, the preparation of metadata accompanied many arduous discussions among the participants about which data elements were truly essential to their program evaluation goals. Once that key data set was identified, the group then had to harmonize different definitions for the same element or develop processing routines that would take similar data from the different input sources and transform it into a standard format and definition for the shared repository.

6.6. *Uniformity and integration of sources*

These related dimensions have important implications for the design and operation of an access program. Consider these factors:

- Content from multiple sources or in multiple formats increases the overhead associated with managing relationships, handling the information, and making it available for use.

- The larger the number of sources, the larger the number of relationships that will need to be maintained with suppliers.
- The greater variation among sources, the more work needed to describe and maintain them.
- When integrating data from multiple sources, the level of effort rises as uniformity among sources diminishes.

The skill required to integrate multiple disparate data sources into a new resource can be enormous. As we found in the HIMS case, integration demands a fine-tuned understanding of the content and clear expectations about intended users and uses. The KWIC program was designed to allow some user-directed integration (with appropriate cautions about comparability), but made no attempt to integrate the data from different sources as part of the clearinghouse itself.

6.7. Usefulness over time

If the content has enduring social, legal, or historical value, metadata is critical to its long term and effective use. High-quality and complete metadata that addresses context and frame of reference will help ensure that the content remains understandable to future users. A design for information of enduring value must also emphasize standards and include technology choices to ensure migration to new formats and preservation long into the future. The GIS program addressed this need by dividing responsibilities between the central clearinghouse and the agencies who maintained custody of the data. The clearinghouse maintained the most current metadata and left responsibility for preservation of both current and earlier versions of data sets to the custodial agencies. For both HIMS and KWIC, designers discovered that the ability to analyze trends depended not only on longitudinal data, but on an historical record of changes to the way the data was collected, compiled, and managed.

6.8. User and supplier relations

If relationships are marketlike, planners need to pay less attention to administrative activities, which are much more important in formal arrangements that require rules or contracts to guide them. Community relationships demand considerably more staff and leadership attention because they rely on long-term shared activities that build trust for joint efforts. These relationships require a larger commitment of resources and must lie closer to the heart of the organization's mission than a program that operates on a simple transaction basis. The KWIC program faced both extremes. User relationships were marketlike, since any person could access the public Web site without registering or otherwise engaging in a relationship with the host agency. By contrast, the supplier relationships were those of a community in which the host agency and the agencies that contribute the data need to work out many details and agreements about the timing of data collection, formatting and preparation of data, the extent and nature of metadata, ways to identify and enter corrections and updates, and so on. Fortunately, KWIC was very much aligned with the overall mission of the council and received the kind of executive support it needed to invest heavily in this management effort.

6.9. Access provider involvement in data collection

If the access provider plays no role in original data collection, it will be important to require information suppliers to include good metadata with their information sources. This was the case for both the GIS and KWIC programs. If the access provider participates in the original data collection strategy or work, the access program can benefit from staff who have a much deeper understanding of the information resources they are making available to users. However, the cost of the program is likely to rise to accommodate this additional role and the complex relationships with suppliers that it implies. HIMS exemplified this model. The state agency host was intimately involved in regulating and supporting the shelters through licenses, inspections, and educational programs. This common contextual knowledge allowed all the participants to work jointly on the data definitions and the data analysis strategy, as well as to build key features into case management systems that would more easily capture and transmit the needed data between the shelters and HIMS.

6.10. Data transformation and other value-added services

Value-added services can compensate for inexperienced users, highly variable data sources, and the need for an expert frame of reference. By providing indicators, normalized data, analytic reports and summaries, user-oriented tools, instructions, and support services, an access provider makes complex or voluminous data more accessible and useful for more users. These services, however, are costly and demand a broader range of staff skills and technical tools than in a program that does not provide these services. The GIS program dedicated staff to these functions, with an initial emphasis on user education about GIS tools and techniques. As the program matured, it began to offer an online help desk, map products, and downloadable imagery.

7. Value of the research

Three kinds of value emerged from this research. The first is its multidisciplinary contribution to the study of government information strategy and management. We have drawn together concepts from information science, public management, and government information policy into a holistic framework for better understanding the nature, dynamics, and components of electronic access programs. By identifying and defining a consistent set of observable factors that shape the cost, complexity, and potential performance of information access programs, we have provided an empirically based, testable model for assessing the design and predicting the outcomes of such efforts. This model could be used, for example, to examine the implications of policy changes for the design and performance of access programs.

Future research could include a variety of field investigations that test the applicability of the dimensions, refine them, and identify others that impact program design, operation, and performance. For example, such investigations might ask which of the dimensions are more closely associated with cost, with user satisfaction, with sustainability, and so on. The

description and understanding of each dimension could be augmented by research into the range of situations that each one represents. Here we might seek detailed examples of situations at the anchor ends of the dimensions. Field tests and expert evaluations could be combined to develop the elements of an expert system that guide planners not only through the analysis of each dimension, but that predict likely challenges and suggest courses of action that take into account the interaction among dimensions. All of these would enhance our understanding of programs that provide electronic access to government information, and would improve their design and performance.

The second value is in the potential for the study findings to contribute to information policy development. Electronic information access programs address the values represented by the First Amendment, the Freedom of Information Act, and Open Records Laws. Conversely, they present challenges for information stewardship, protection of personal privacy, and information security. As such, they are excellent venues for developing and testing ways to balance these competing policy goals and to evaluate their effects.

The third value of this study is in highlighting practical implications for those making decisions about electronic access programs. By presenting the design factors as scaleable dimensions, we offer practitioners a logical and intuitive method of assessing their current status and initial plans against their desired goals. The interactions among dimensions that we point out show how combinations of factors can have a strong effect on program design and likely program performance. These interactions illustrate how important it is for agencies to involve a variety of specialists as well as users and suppliers in their design strategies. As a result, effective electronic access programs must be context-specific, collaborative ventures. A holistic framework, such as the one presented here, allows for a more complete understanding of the needs to be addressed, the resources available to meet them, and the policy, management, and technology challenges likely to be encountered along the way.

Notes and References

1. Office of Management and Budget (2001). *The President's management agenda, fiscal year 2002* [Online]. Washington, DC. Available at: <http://www.whitehouse.gov/omb/budget/fy2002/mgmt.pdf>. Retrieved on August 15, 2003.
2. Wiederhold, G. (1995). Digital libraries, value, and productivity. *Communications of the ACM*, 38(4), 85–96.
3. Laskowski, M. S. (2000). The impact of electronic access to government information: What users and documents specialists think. *Journal of Government Information*, 27(2), 173–185.
4. Cheverie, J. F. (1999). Federal information in the networked environment: A perspective from the Coalition for Networked Information. *Government Information Quarterly*, 16(3), 261–275.
5. National Science Foundation (2002). *Digital libraries initiative, phase 2* [Online]. Arlington, VA. Available at: <http://www.dli2.nsf.gov/>.

6. Murphey, D. A. (1999). Presenting community level data in “outcomes and indicators” framework: Lessons from Vermont’s experience. *Public Administration Review*, 59(1), 76–82.
7. Danner, D., & Taylor, P. W. P. (1997). Principle and practicality: Funding electronic access to Washington State government information. *Journal of Government Information*, 24(5), 347–359.
8. Jobe, M. M. (1999). The power of information: The example of the US Toxics Release Inventory. *Journal of Government Information*, 26(3), 287–295.
9. Bar-Ilan, J. (2002). Methods for measuring search engine performance. *Journal of the American Society of Information Science and Technology*, 53(4), 308–319.
10. Wenyin, L., Chen, Z., Li, M. J., & Zhang, H. J. (2001). A media agent for automatically building a personalized semantic index of web media objects. *Journal of the American Society of Information Science and Technology*, 52(10), 853–855.
11. Zhang, J., & Wolfram, D. (2001). Visualization of term discrimination analysis. *Journal of the American Society of Information Science and Technology*, 52(8), 615–627.
12. Chen, S. Y., & Macredie, R. D. (2002). Cognitive styles and hypermedia navigation: Development of a learning model. *Journal of the American Society of Information Science and Technology*, 53(1), 3–15.
13. Shapira, B., Kantor, P. B., & Melamed, B. (2001). The effect of extrinsic motivation on user behavior in a collaborative information finding system. *Journal of the American Society of Information Science and Technology*, 52(11), 879–887.
14. Cothey, V. (2002). A longitudinal study of World Wide Web users’ information-searching behavior. *Journal of the American Society of Information Science and Technology*, 53(2), 67–78.
15. Ford, N., Miller, D., & Mass, N. (2001). The role of individual differences in Internet searching: An empirical study. *Journal of the American Society of Information Science and Technology*, 52(12), 1049–1066.
16. Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society of Information Science and Technology*, 53(2), 145–161.
17. Wathen, C. N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society of Information Science and Technology*, 53(2), 134–144.
18. Tang, R., & Solomon, P. (2001). Use of relevance criteria across stages of document evaluation: On the complementarity of experimental and naturalistic studies. *Journal of the American Society of Information Science and Technology*, 52(8), 676–685.
19. O’Connell, B. W. (1999). New wine in old wineskins: US government information in a networked world. *Journal of Government Information*, 23(3), 217–225.
20. Barnum, G. (2002). Availability, access, authenticity, and persistence: Creating the environment for permanent public access to electronic government information. *Government Information Quarterly*, 19(1), 37–43.
21. Ray, J. M. (1998). Search for tomorrow: The electronic records research program of the U.S. National Historical Publications and Records Commission. *Journal of Government Information*, 25(4), 367–373.

22. Pardo, T. A., Dawes, S. S., & Cresswell, A. M. (2002). *Opening gateways: A practical guide for designing electronic records access programs* [Online] (2nd ed.). Albany, NY: Center for Technology in Government, Available at: <http://www.ctg.albany.edu/resources/pdfrwp/gateways.pdf>. Retrieved on October 21, 2003.
23. Kelly, K. L., Pardo, T. A., Dawes, S. S., DiCaterino, A., Foderingham, W. (1995). *Sharing the costs, sharing the benefits: The NYS GIS cooperative project* (Center for Technology in Government Project Report 95-4). Albany, NY. Available at: <http://www.ctg.albany.edu/resources/abstract/gis95-4.html>. Retrieved on October 21, 2003. See also: Dawes, S. S., & Oskam, S. (1995). The Internet, the State Library, and the implementation of statewide information policy: The case of the New York State GIS Clearinghouse. *Journal of Global Information Management*, 7(4), 27–33.
24. Center for Technology in Government (2001). Building trust before building a system: The making of the homeless information management system. *The insider's guide to using information in government*. Albany, NY. Available at: http://www.ctg.albany.edu/guides/usinginfo/Cases/bss_case.htm. Retrieved on October 21, 2003.
25. Center for Technology in Government (2001). Moving from paper to the Web: How the Council on Children and Families is transforming a static information resource into a dynamic one. *The insider's guide to using information in government*. Albany, NY. Available at: http://www.ctg.albany.edu/guides/usinginfo/Cases/kwic_case.htm. Retrieved on October 21, 2003.
26. Ouchi, W. G. (1980). Markets, bureaucracies, and clans. *Administrative Science Quarterly*, 25(March), 129–141.