Minnesota Office of Technology

Minnesota Office of Technology
332 Minnesota Street, Suite E1100
St Paul, MN 55101-1322
612.215.3878
Fax 612.215.3877

A Quest for Data Quality

Eliminating Redundant Storage of Data

A Report to the Governor and the Minnesota Legislature

January 1998

Table of Contents

Per the requirements of Minnesota Statutes Chapter 3.197, the cost of preparing and printing this report is estimated to be: $6,950 for staff time and $161 for printing.

This information can be made available in an alternative format such as Braille, large print, audio tape or computer diskette. It is also available at our web site at http://www.ot.state.mn.us. For TTY communication, contact the Minnesota Relay Service at 612-282-5353 or 1-800-627-3529 and ask them to place a call to the Office of Technology at 612-215-3878.

Executive Summary

Under Minnesota Statutes 16E.04, subd. 2, f, the Office of Technology:

> *"... in consultation with the intergovernmental information systems advisory council and the legislative reference library, shall recommend specific standards and guidelines for each state agency...in regard to the following: 1) establishing methods and systems directed at reducing and ultimately eliminating redundant storage of data..."*

Minnesota Statutes 16E.04, subd. 2, h, states:

> *"The office shall report to the legislature by January 15 of each year on progress in implementing paragraph f..."*

The Office of Technology has chosen to partner with the Information Policy Council (IPC) to prepare its first annual report on the statewide establishment of standards and guidelines and the elimination of redundant storage of data. This report describes a need to refocus critical data management objectives to address:

- inappropriate, unplanned redundancy;
- data reuse as the preferred method of sharing data;
- data management as the key to eliminating redundancy;
- enablers and barriers to achieve statutory data goals;
- a strategy for eliminating unplanned redundancy; and
- statewide progress and suggested future directions for data management (see Appendix C).

Over the past a decade, some legislators and agencies have promoted "data sharing" and the "elimination of redundant storage of data" in statute. During this time, significant progress has been made toward meeting those goals. Many large agencies have implemented IRM (Information Resource Management) strategies and principles to help eliminate systems that serve only single uses or organizations. "Communities" of agencies with common interests have identified sharable data and are working to make sharing a reality. Several standards and guidelines are in place with more under development. Education and training – focused on data modeling and management – have reached a widespread audience within state government. Ad hoc groups have formed to address data issues and share expertise across organizational boundaries.

Barriers, however, remain and work is required before the state realizes its data management goals. Minnesota still needs to develop a statewide picture of its data, clarify the statutory vision, define "successful data reuse and exchange", establish additional data standards, and develop ways to implement and measure data management progress.

The state needs to revisit the legal implications of eliminating redundant storage of data, which might lead to the conclusion that shared databases are required. Minnesota's vision and objectives for data sharing and data redundancy are described in statute; however, some statutes are in conflict with others and may even be mutually exclusive. For example, Minnesota Statutes Chapter 16E.04, subd. 2 promotes data sharing, while Chapter 13.05, subd. 4 and 9 requires statutory authority for sharing certain data. Data-sharing barriers resulting from conflicting statutes, or from different requirements across jurisdictions, must be addressed.

Other legal issues must also be examined. For example, the evolving legal status of electronic records may impact shared databases, as may intellectual property rights or agency cost recovery statutes. Realizing this objective may also require reevaluating funding and/or organizational structures to allow agencies to develop and implement shareable databases.

The state should examine the basic premise that redundant data storage is undesirable. Recognizing there are both pros and cons to redundant data, statutory requirements should specify appropriate elimination of redundancy and avoidance of unplanned redundancy. Use of the term "data sharing" should also be evaluated. Data sharing can be accomplished using computer diskettes, electronic files, or paper reports, each of which duplicates data in the process. This duplication conflicts with eliminating redundant data storage when agencies receiving the shared files and diskettes store what they receive. Assuming the vision of eliminating redundant storage of data is still valid, the term "data sharing" could be replaced with the term "data reuse". "Sharing" can be broadly interpreted to allow data duplication, however "reuse", by definition, avoids duplication.

Background of Data Management

The Information Policy Council (IPC) began seriously addressing data management in 1986 when it adopted its four Information Management Principles (see Appendix A) based on the following conviction:

> *"Management of state government will be greatly enhanced with better management of its information. The gains will not only be in the efficiency of operation but also in taking fuller advantage of information when making critical decisions. This will be accomplished when we consider information as a state resource and cooperate toward a common direction for the state's information facilities, networks, and data.*
>
> *To that end, these general principles represent a foundation of understanding and agreement. These principles will assist agencies in accomplishing their legislatively mandated responsibilities while also contributing effectively to the collective needs of the state."*

In the early 1990s the state adopted an Information Resource Management (IRM) strategy based in part on the Information Management Principles. IRM addresses the management of data, applications and technology resources in support of agency business missions. The state first established Six Critical Success Factors (see Appendix B) as prerequisites for implementing IRM. Along with policies, standards and guidelines for implementing IRM, the state was on the way to achieving IRM success.

Critical Data Management Issues

Several important issues face the state in its quest to eliminate redundant storage of data. Those issues with implications for eliminating redundant data are detailed below.

What Is Redundant Data?

Some believe "redundant data" means data that is out of control, not well managed, stored in several places, or duplicated and out of sync. Years ago redundant data storage was a fact of life. Technology was not powerful enough to enable meeting everyone's needs with centralized data storage. Networks linking people together did not exist, or were prohibitively expensive. Over time technologies changed and today they no longer impose the same limitations. New problems with performance, security, access, privacy and accountability surfaced when central data storage became feasible. For example, security and privacy are at risk when data is stored in a central location and can be accessed by more people.

Technologies must meet business needs in addition to solving technical problems. A data warehouse is an example of technology that meets business needs by enabling information access without compromising security or performance of operational systems. A data warehouse is a set of tools that provides interested parties with access to a specialized collection of data and data analysis tools. Unlike operational databases, warehouses contain integrated, historical data, organized in views that are meaningful to the users (rather than organized for data storage efficiency). Data warehouses may appear to store redundant data, but actually produce variants of data that provide value to information users. Data proliferation is minimized because people can get the data they need without having to create their own (redundant) databases. However, current technologies cannot support warehouse-like access to data from a source database while still meeting operational performance and security requirements. Thus, elimination of data redundancy is not an option in the foreseeable future.

Technology is not capable of addressing all the issues that arise when trying to reduce redundant data. Yet the state needs realistic solutions that meet all of its business needs – not just the one for reducing redundant data. Other requirements that must be considered include balancing access with privacy, protecting record security, assuring accountability for information accuracy about citizens, and providing value to taxpayers.

Recommendation: The efforts to avoid redundant data must be replaced with efforts to avoid inappropriate or unplanned redundant data. This begs the question: When is it appropriate to create redundant data? Criteria should be established to determine when redundancy is appropriate. The criteria should apply to data with short-term uses that are not retained, and for data with long-term uses or retention requirements.

How Does Data Management Eliminate Redundant Data Storage?

"Data management" is relatively new concept within the information-processing industry. Formal data management is still not well defined or commonly understood because of perceptions that organizations already do "manage" their data. When data management is taken for granted or treated

as an afterthought – as it has been for years – little attention is paid to formalizing or documenting its discipline.

Until recently database administrators, as part of database design and physical data storage, handled data management. Data decisions were based on storage efficiency or ease of maintenance by technical staff. Decisions were rarely based on supporting the business, or on user needs for access. The industry now recognizes the need for a complimentary "data administration" function to separate the business and technical aspects of data management. However, people still do not agree on which aspects of managing data belong to each discipline.

The shift toward a business view of data can be attributed, in part, to IRM philosophies. IRM brings key data issues to the forefront, and expands on previous thinking to address voids that existed in the past. IRM – with a "view from 20,000 feet" – recognizes data management is not a technical responsibility, but belongs to business decision makers. Under IRM, data are managed like other resources, such as money or property: protected, tracked and valued. Data provide organizational history, support daily functions and describe the financial wellbeing of an organization. For government, data must also be held in trust for and made accessible to citizens.

Rapidly changing technology adds another dimension of complexity to the data management challenge. Long ignored areas, such as appropriate data access, are now critical for World Wide Web publishing and data warehouse implementations.

Data management applies sound values and principles to the creation of data policies, standards and guidelines that define what to do and how to do it. Several essential aspects of data management must be addressed before agencies can eliminate redundant data storage. Agencies must have a clear understanding of their data and what it means before they can compare it to data from other organizations. They must identify access requirements and restrictions, define terms and establish rules for communicating about common data, goals and objectives. A byproduct of data management is the knowledge needed to comply with public policy.

Key data management components that enable elimination of redundant data include:

- Data Standards and Guidelines
- Common Metadata
- Reusable Data
- People and Skills
- Knowledge Exchange and Collaboration
- Data Access
- Enabling Legal Framework
- Technology Readiness and Availability

Key data management components

Data Standards and Guidelines

Data administration provides standards, rules and language for documenting and describing data to ensure all involved parties can reach a common understanding. Standards provide a consistent framework of what must be done. Guidelines provide how-to methods and implementation information. Standards and guidelines for data management describe what must be in place across organizations to share data, protect data integrity and avoid compromising security.

Data standards help ensure data mean the same thing regardless of the source, and that all needed data can be collected. For example, if location data were always collected the same way (e.g. state, county, and city), data about a place could be located regardless of which agency collected it. Effective standards must be applicable, measurable and achievable regardless of the size or complexity of an organization. In Minnesota, this means small agencies without dedicated information technology staff as well as large agencies with entire information technology departments must be able to comply. Since most shared data are historical, the data are typically collected and stored prior to use. Thus standards, measurement criteria and methods must be in place before data are collected and stored.

Recommendation: Define and develop a comprehensive set of data standards, guidelines and implementation and compliance criteria related to data use that are achievable throughout state government and are responsive to individual needs.

Common Metadata

Metadata is data about data that provides a common link between data from different sources. Labels on cereal boxes illustrate the concept of metadata. Regardless of brand, cereal boxes describe serving size, calories per serving and fat grams per serving. The metadata "fat grams per serving" and "calories per serving" have been standardized for meaning, for the ingredients reported and for measurement process. The metadata on cereal boxes allows people to compare the nutritional value of several products.

Metadata allows us to understand nutritional information on cereal boxes by providing a cross-reference for comparing ingredients. For data stored in a database, metadata cross-references data from one organization or system to data in another. A data model is a commonly used tool for documenting metadata.

Metadata also clarifies the type and meaning of data. The Year 2000 date problem illustrates this point. With standardized metadata, organizations could have searched for all data of the type "date" and greatly simplified the assessment, repair and testing phases. In most systems, however, the lack of metadata made it impossible to consistently find dates even in a single system, much less among several different systems.

Recommendation: Create a statewide data model to define metadata of interest to multiple agencies. Develop a metadata repository or clearinghouse – as was done by the Geographic Information Systems (GIS) community – to document statewide metadata that are available for all agencies.

Reusable Data

The term "data sharing" has been referred to in discussions about eliminating redundant storage of data; however, data sharing can mean photocopying, faxing or sending a file through e-mail. These sharing techniques involve data duplication, which is contrary to the purpose of eliminating redundant storage of data.

When data is shared through duplication the recipient avoids redundant data collection. However recipients may enter the data into another computer, or file it, which creates redundancy. A better strategy for sharing is data reuse. Reuse accomplishes the same thing but, by definition, without duplication. A web page is an example of data reuse because a page containing data can be accessed by anyone. "Data reuse" is like a library where resources can be checked in and out, while "data sharing" is like a copy center.

For the balance of this report, reuse should be viewed as the desired means of achieving "data sharing". "Data sharing", "reuse" and "redundant data storage" will apply only to data stored in databases; not to "records" containing data fixed in time.

Recommendation: Resolve apparent legislative language conflicts regarding state data goals and objectives. Define "successful data reuse and exchange" for Minnesota government. Develop a long-term strategy for building statewide databases that allow data reuse. In the near term, promote community and agency efforts to build common databases.

People and Skills

Knowledgeable people who understand data management and their data, and can participate in establishing frameworks and standards must develop basic data management prerequisites. In this capacity, skilled staff are the backbone of an organization – they build and document an organization's history and operations.

Data management enables an organization to meet business goals in an effective and efficient manner; therefore, it is in the organization's best interests to commit expertise, competence and knowledge to data management. Education and training – to build data skills and establish a common knowledge base – should be supported throughout the organization.

Recommendation: Recognize the unique job requirements for data management positions, such as data administrators and data modelers, within the state system. Establish data management positions in state service such that they are sufficient to attract and retain skilled data professionals. Continue investing in education and training among state staff. Include agency management in training efforts to ensure their participation in effective data decision making.

Knowledge Exchange and Collaboration

The exchange of knowledge and efforts are critical to developing standards, metadata and reusable data for a "community" of agencies. Community participants must bring to the table an understanding of their organization's data and that of other similar organizations. They must have expertise to develop effective standards and define common metadata. Experienced people can also mentor newcomers to enhance a community's overall skills and knowledge.

Several agencies have loaned staff to help other agencies build models. Benefits to recipients include the use of specialized skills that are not needed full time or that do not exist within the agency. Additionally, outside labor frees people to act as subject matter experts in defining their agency's business. For loaned workers, a resource pool can strengthen skills, and appears to boost morale

through career renewal. Loaned workers have been able to offer knowledge of other organization's data, and recognize potentially common data. As a result, some agencies have discovered data commonality that has led to reuse of model fragments and metadata across agencies.

One of the Critical Success Factors is to develop information models. Modeling efforts today result from previous groundwork laid by agencies meeting IRM requirements.

Recently a groups of state employees, ranging from chief information officers to data modelers, formed the Data Issues Group (DIG IT). Many members have participated in past agency modeling efforts. DIG IT plans to formalize knowledge transfer, capitalize on data exchange and facilitate statewide efforts in standards development.

Recommendation: Help DIG IT become a working partner with IPC and the Office of Technology to achieve the state's data goals and objectives. Promote and facilitate the exchange of human and data resources. Develop a means for capturing the information currently being exchanged for future reuse by a wider statewide audience.

Data Access

To achieve reuse, technologies must provide efficient data access without compromising public policy requirements, user needs or security. In practical terms, data reuse cannot be achieved without providing adequate access to data. Some agencies report they recreate data when it is too tedious, slow or difficult to obtain the data from its original source. However, providing access is not that simple. Depending on who needs the data, access could range from a simple solution such as putting data on a web page, to an extremely sophisticated solution such as creating a data warehouse.

Recommendation: Continue North Star II development as the state's focal point for data access. Align standards and guideline development with data access needs. Ensure that access mechanisms, such as warehouses, clearinghouses or web sites being developed throughout the state, are consistent with state goals and objectives.

Enabling Legal Framework

Laws from various jurisdictions can enhance or restrict data management activities. The legal framework under which government operates should be studied for conflicts and loopholes. These areas must be resolved before effective, legal reuse can occur. To balance the sometimes-conflicting legal requirements, public policymakers may need to reexamine the merits of each requirement, including the one to eliminate redundant data. Some of the legal issues discussed here are not new but may be complicated by electronic records. Other issues are new as a direct result of trying to store common data.

Legal framework issues affecting the elimination of redundant data storage include:

- Funding structure
- Lack of a legal entity to manage and administer common data and metadata
- Costs
- Data practices
- Records management

- Lack of consistent definitions
- Legal admissibility of electronic records
- Compliance with external standards
- Ownership, intellectual property and monetary benefits

Funding Structure

The state's funding structure makes it difficult to allocate resources for functions or projects that cross organizational boundaries. Agencies may provide services to community projects without receiving direct benefits or may be unable to contribute funds to the effort. Projects that go beyond the jurisdiction of the Minnesota legislature (such as GIS projects that include the private sector) may not be within the reach of state agencies if funds are required to participate or receive benefit.

The funding structure assumes accountability for project success, data integrity and security. Accountability becomes a fuzzy issue when there is not a designated organization in charge. The state can achieve greater economies of scale, however, if the structural funding issues are resolved and incentives established.

Recommendation: Ensure that funding mechanisms and organizational structures recognize, validate and reward multi-jurisdictional efforts.

Lack of a legal entity to manage and administer common data and metadata

Presently Minnesota government has not designated an organization with authority, responsibility and accountability for managing the state's reusable data. In most organizations, this responsibility belongs to the chief information officer; however, in Minnesota this function has not yet been clearly identified and assigned.

Recommendation: Designate responsibility for the central data management function within the state to a chief information officer and grant this position the authority and responsibility to define and meet the state's data goals and objectives.

Costs

Costs will be incurred in support of data reuse to develop and maintain metadata, data, applications and technology. Metadata are typically stored and accessed within a repository. Repositories require investments in hardware, data analysis and design, implementation, and ongoing maintenance and support. The investment is similar to that of a medium- to large-sized software system. New applications and shareable databases must be designed and developed, and existing systems migrated to the new ones. Statewide data administration and management of the repository databases and applications will also be required.

Recommendation: Encourage funding for central data management to develop metadata and databases for state agency use. Develop a long-term cost recovery plan for statewide data management. (Costs should be eventually recovered as a result of economies of scale and reuse.)

Data Practices

The Data Practices Act plays a significant role in managing data in Minnesota government. The main issues that fall within the scope of this report are:

- the perception that the Act is complicated and confusing; and
- the implications of the Act – many of which were deliberate – for data reuse.

Many agencies and policymakers have difficulty understanding and implementing the Data Practices Act, as evidenced by recent testimony before the legislature's Information Policy Task Force. The subject matter covered by the Act is complex in its breadth and depth. For example, the present Act restricts what data can be collected, for what purpose, and how the data can be used. It requires notifying data providers of how data will be used, defines which data may be released and to whom, and defines which data must be made accessible to the public.

By requiring statutory authority for sharing some types of data, the Data Practices Act inherently discourages reuse (Minn. Stat. Chapter 13.05, subd. 4 and 9). At the same time, other laws (such as Minn. Stat. Chapter 16E.04 subd. 2 on eliminating redundant storage of data) conflict with this philosophy. Each aspect of the Act has implications for the reuse of data. For example, it is a common business practice to collect data together when they are about the same subject. However, the present Act may restrict the ability to collect data this way, if the party collecting it is not the end user of the data.

Data practice issues for which the present Act should be scrutinized, and the implications for data reuse carefully analyzed, include:

- the role of the Act in both restricting and providing access to data;
- the understanding by agencies (whether real or perceived) that the act is difficult to interpret and presents a risk to anyone who misinterprets it;
- aspects of the Act that prohibit agencies from collecting data not directly needed for the purpose that triggered the collection; and
- a close examination of the legal authority to not share data (that are not public) with the goal of getting access where appropriate.

Recommendation: Continue the work of the Information Policy Task Force and the IPC Data Practices subcommittee (see also Lack of Consistent Definitions, p. 15).

Records Management

Effective data management must consider the statutory mandates of the Data Practices Act (Chapter 13), and the laws governing official records (Chapter 15.17) and government records (Chapter 138.17). The laws pertaining to records retention and disposition depend on the existence of metadata that provide definitions of data and records. Redundant metadata, collected to meet differing legal requirements, may contribute to the redundant data storage problem.

In addition to legalities, good business practices should also be considered. For example, the records an agency preserves should be aligned with its business purpose and the importance of its mission to the state.

Recommendation: Address standardization and coordination as applied to data management activities for meeting legal data requirements (see also Lack of Consistent Definitions, p. 15).

Lack of consistent definitions

Past legislative decisions included the use of different terminology and definitions between the Data Practices Act and the Records Management Act. The result may have complicated the interpretation of the Data Practices Act for some organizations. Terms like "record", "data", "information", "information system", "record keeping system" and "trustworthy system", are not defined consistently so agencies have a standard context for their activities.

Recommendation: Revisit the Data Practices Act and Records Management Act in light of data reuse and the elimination of redundant storage of data. Correct and clarify legislation as necessary to allow state agencies to work toward data goals within the context of their missions.

Legal admissibility of electronic records

The primary reason to keep records is accountability to others, including the courts. However, legal admissibility of electronic records is uncharted territory. At present, it is unclear what the courts will accept; thus, agencies can't be sure what records will be recognized as authentic and reliable. Without consistent legal criteria, agencies may feel compelled to maintain both paper and electronic records. In addition, the legal admissibility of electronic records may be based on the "trustworthiness" of systems (hardware, software and procedures) that create and store them. [1]

"Trustworthiness" has data reuse implications. Eliminating redundant storage of data means moving from one system and one set of data to one set of data and unlimited systems. In a reuse environment, it may not be possible to prove which systems and data were accessed or updated at a point in time. Also organizations that use others' data might have to rely on outside documentation to prove authenticity.

A lack of consensus among and between states and other jurisdictions has complicated the creation of standards. In the federal courts or the courts of other states, electronic records might be scrutinized using different, even contradictory, criteria from those of Minnesota. Data with a high legal profile needs to be carefully evaluated to be certain that it can withstand the scrutiny of any court.

Recommendation: Minnesota should be actively involved in external efforts to define the legal admissibility of electronic records.

Compliance with external standards

Minnesota does not have jurisdiction over all organizations with which it exchanges data. External organizations impose standards on agencies that may conflict with state standards, or impose dual requirements. External standards that affect agencies include the Department of Public Safety's reporting to the Federal Bureau of Investigation (FBI), and electronic data interchange (EDI) standards for filing campaign finance reports (Conference of State Government Ethics Regulators) that will affect the Campaign Finance and Public Disclosure Board in the future.

---

[1] For examples of some definitions of "trustworthy systems" see draft standards and guidelines proposed by NIST (National Institute for Standards and Technology) and the COBIT Project (Control Objectives for Information and Related Technology).

Minnesota does not inventory external standards with which state agencies must comply. Yet the state cannot develop and implement effective standards for all state agencies without accounting for the impact of external requirements.

Recommendation: Inventory the external data standards that affect Minnesota government agencies, assess their impact and determine possible actions. Take an active role in cooperating with external partners to develop mutually compatible standards.

Ownership, Intellectual Property and Monetary Benefits

Community databases have multiple data creators and users. Responsibility, accountability and ownership of the community data must be addressed, as well as the roles and responsibilities of data creators and users. Issues related to eliminating redundant storage of data include:

- Who owns the data in terms of receiving monetary benefits? Some agencies copyright and sell their data, and might be unable to retain this benefit with community databases.
- Who owns the data in terms of cost recovery for its creation and storage? Under current law, agencies can recover costs, including development costs, for data provided to the public.
- Who is accountable when liability issues arise?

Recommendation: Identify and pursue resolution of legal issues associated with intellectual property, ownership and receipt of benefits.

Technology Readiness and Availability

Technology is critical to elimination of redundant storage of data. New technology will be required to implement statewide metadata and to enable data access without loss of important protections. The design, development and management of reusable resources will be achieved by organizations that are "technology ready", meaning that they have the leadership, accountability, skills base, planning and management capabilities, and knowledge necessary to implement and manage the new solutions.

Recommendation: Develop a strategy for becoming "technology ready". Continue building on IRM Critical Success Factors within agencies.

Eliminating Data Redundancy: Barriers and Enablers

Even if the state pursues the data management recommendations in this report, success will depend on eliminating the barriers and ensuring enablers are present. This portion of the report identifies enablers and barriers in the quest for eliminating data redundancy, and explains how each effects the state's goal.

Enablers for Eliminating Redundant Data

Major enablers include effective leadership, appropriate use of technologies, collaboration and cooperation, education and a data management infrastructure.

Leadership

Effective leadership is critical to success in the strategic management of information resources. Leadership must come from the top where organizational direction can be established and resource commitments made. Executive leadership is the first Critical Success Factor that agencies have been working to develop for the past several years.

At the agency level, leadership means individuals and/or steering committees with responsibility, authority and accountability for managing information resources. At the state level, leadership comes from the community through IPC and from "special interest" communities such as Criminal Justice. Legislative support and understanding also enable agencies and the state as a whole to achieve its data goals and objectives.

Appropriate Technologies

Technology can be a barrier or an enabler to achieving the state's goals. To enable eliminating redundant data storage, technologies can be used that promote interoperability regardless of platform and allow agencies to open their doors to the incoming flow of data and to those seeking access. Technologies can enable data access, assist with analysis or manage metadata and its linkages to real data. Examples of relevant technologies include decision support systems (DSS), data dictionaries, repositories, data warehouses, and SQL-based data query tools.

Collaboration and Cooperation

Collaboration is the effort of people combining forces to reach common goals. Collaboration efforts are successful when participants have a high level of commitment, new knowledge is created that would not otherwise exist, and clear outcomes are defined.

Collaborations streamline the work of a community and act in its behalf. Collaboration outcomes that can benefit an entire community include data standards, procedures to facilitate data reuse, and knowledge and skills exchange across agencies.

Formal collaborations exist to enable efforts over time, while other collaborations exist only for specific projects. Examples of formal collaborations include the IPC and the Criminal/Juvenile Justice Information Policy Task Force. A collaborative project is the Metro GIS data model project, that defined metadata for Minnesota's geographic community. The Data Issues Group is a collaboration of mutual data interests.

Most collaborations depend on resources of participating agencies; however, some exist in statute or have funding sources. An example of the latter is the Criminal Justice collaboration. The long-term effectiveness of these efforts depends on a continuous flow of resources, people and budgets to accomplish goals.

Education

Education and training are key to realizing goals. When many people acquire common knowledge, the overall capability of collaborations is improved and the experience of individual agencies is enriched. Outcomes such as standards or data models are of higher quality when created by those with a sound base of theory and practitioner-level expertise.

Education and training are critical when implementing an outcome. For example, a successful educational program for implementing data standards was undertaken by the GIS Standards Committee. The committee conducted metadata workshops and training classes, established a metadata clearinghouse where metadata could be placed for reuse, and provided free software to public sector organizations to use when developing metadata. As a result, the GIS community is well positioned to develop standardized metadata and to reuse metadata created by others.

Data Management Infrastructure

A data management infrastructure needs to be in place to achieve broad data goals. People can build community data when they first have their own data "house" in order. They know what data they have, what it means, and are ready, willing and able to collaborate with others in developing new data that meets common needs.

The Critical Success Factors include building capability in several areas. Data models establish data requirements and define metadata; planning provides mechanisms for managing data development and identifies appropriate projects; and policies, standards and guidelines establish boundaries, requirements and best practices.

Barriers to Eliminating Redundant Data

Major barriers to eliminating redundant data include technological constraints, issues of management, an inadequate core infrastructure, organizations not able to collaborate effectively, legal constraints, and the inability to implement or follow through on actions.

Technological Constraints

To some, data access implies "punching holes in the firewall" and, thus, introducing security risks. The ability to consolidate or access large amounts of data raises privacy concerns. Technology has not addressed these concerns adequately to avoid risk to operational databases; therefore, solutions must be found to provide legal access without compromising security or privacy.

Responsibility and accountability for the status and integrity of data could also be at stake. Technology can blur the lines between the creators and the users of data when common databases are used.

Other technological deficiencies that could impede progress toward common databases include the lack of adequate capacity in terms of bandwidth, servers and networks; hardware that is out-of-date; proprietary database technologies; and hardware and software platforms that are not compatible.

Technology constraints result from products acquired over time that are in various stages of obsolescence. Upgrading the technology is not always the answer because migration can be costly and time-consuming. Organizations are also faced with existing data and metadata that will not be easily converted to a community system.

Management Constraints

Agencies may lack an adequate IT work force to support IRM operations. The shortage of skilled IT professionals has also affected the private sector, where large signing bonuses and hiring incentives as generated high rates of staff turnover. According to surveys, staffing concerns are a major barrier for most agencies. Several agencies report having IT positions they cannot fill, especially in specialized areas such as data or database administration. Increasing the state's data management programs could further tax the labor pool, leaving agencies without adequate staff to maintain community data and metadata.

Some agencies report inconsistencies in implementing IRM due to a lack of clear, statewide goals and objectives. Achieving common goals depends on having a common framework and vision. Although some policies, standards and guidelines exist, they may not go far enough when put to the test, and statewide goals and objectives have never been developed.

Managers also need to balance short-term and long-term needs. Data management is a long-term proposition; however, the demands of short-term resources often take precedence.

Inadequate Core Infrastructure

Agencies are unable to reduce redundant storage of data without a good management and technology infrastructure needed to support common databases, such as:

- hardware and software technologies, including data access tools;
- designated management authority to manage and maintain metadata and data;
- shared vision, including agreement about appropriateness of redundancy;
- skilled and adequately staffed workforce;
- methods, standards and processes;
- common language for describing metadata;
- common way to store and access metadata information;
- base of well-defined metadata; and
- accessible databases that provide adequate security, privacy protection, etc.

Some agencies, such as the Minnesota Department of Transportation, and communities, such as Criminal Justice, are addressing infrastructure needs to support common databases. Until Minnesota government addresses the lack of a core infrastructure, however, it will be difficult to support statewide data reuse.

Organizational Effectiveness

Present organizational and policy structures encourage "data territoriality" through allocation of funds, receipt of benefits (e.g., the sale of marketable data), assignment of accountability and evaluation of outcomes. Agency perspectives, with regard to past history, self image and self interests, are reinforced in the process. As long as Minnesota allocates resources, provides rewards and measures outcomes of individual agencies instead of statewide or community efforts, it will be difficult to achieve a shared vision or realize value and benefits beyond the agency level.

Some agencies find it difficult to work towards a future vision for statewide data. Among specific concerns reported by agencies are:

- a lack of statewide data goals and objectives;
- the need to align consultants and outside contractors with statewide data goals and objectives;
- the difficulty migrating from single agency systems to community systems;
- the need for clarification about implementing IRM and the Six Critical Success Factors; and
- the lack of knowledge exchange about standards and the standard setting process.

Legal Constraints

Legal constraints with regard to data privacy, conflicting statutory goals, legal limits on data sharing, existing organizational structures, and narrowly defined assignments of accountabilities and funding continue to make it difficult for agencies to work towards statewide data goals. Legislative support has generally not been present for projects designed to establish community infrastructures or common databases and applications across organizational boundaries. For example, a cross-agency project to develop a common identifier for organizations that do business with the state did not receive funding in either of the last two biennium.

Implementation Constraints

Practical constraints surface when "the rubber meets the road" in implementing cross-organizational projects. Critical requirements for project success tend to fall through the cracks with cross-organizational projects, such as:

- committed leadership;
- accountability;
- availability of skilled staff and funds;
- project management; and
- ongoing maintenance and long-term system operations.

Strategy for Eliminating Unplanned Data Redundancy

The Office of Technology recommends the following strategy for eliminating inappropriate and unplanned redundant storage of data.

Review and Update IRM Requirements

The state needs to revisit and update IRM strategies. The IRM Six Critical Success Factors provide prerequisites for good information resource management in Minnesota government. Most agencies have been able to incorporate these best practices into day-to-day business operations. The Six Critical Success Factors were meant to start agencies along an evolutionary path toward IRM excellence, but were not meant to be the end result. For example, important aspects of managing information resources, such as process management and process improvement, are both absent from the Six Critical Success Factors.

Actions needed include:

- Revisit and expand the Six Critical Success Factors.
- Develop programs for process management and process improvement.
- Establish methods, benchmarks and measurement criteria to compare efforts across organizations and projects.
- Promote "learning organizations" through a strategy of continuous improvement.
- Recommend better funding mechanisms for projects, especially community projects.
- Ensure that projects are successful and agencies can develop common principles and strategies while still meeting IRM and public policy requirements.
- Build the capability and readiness of the IT work across organizations and projects.

More work will occur in this area over the next year. With this level of attention to the management of IT, the environment is established to permit successful use and reuse of data without unacceptable redundancy.

Define State Goals and Objectives for Data Management

Agencies report that a lack of statewide direction with clear goals and objectives is impeding progress towards eliminating redundant data. Activities to rectify this include:

re of the business of state government and its supporting technology. The nology and the IPC created a baseline inventory of agency systems. Other data rts include tracking IT investments and identifying boards and councils involved gy. Each piece helps build a knowledge base about the business of state nd its technology. As more agencies engage in strategic planning, the state can nprove this knowledge. Agency business plans, activities and missions are key to ewide data management requirements.

malized statewide direction, goals, objectives, and measurement criteria for of the ongoing activities of IPC is to blend agency requirements into a unified at sets direction for all. This activity will provide valuable input to the creation of a ewide direction, goals and objectives. Standards and guidelines will be developed cies in meeting the goals and objectives. Important outcomes of this effort must ition for "successful data reuse and exchange" and a measurement process to ress.

nd Enhance Enablers

examine barriers to achieving its data goals to see which can be broken down. e significant progress towards meeting IRM goals; yet barriers have prevented ess. In some cases barriers are distinct, such as laws that prevent exchanging cases barriers may be due to a lack of enabling factors. For example, data ganizations may fail for lack of funding, resources or accountabilities despite good ideas. The state also needs to examine enablers to see which can be enhanced.

nd Grassroots Efforts

path toward eliminating data redundancy is based largely on ad hoc and is time to formalize these groups and recognize their outcomes. Ad hoc groups, nue in spite of barriers and without the resources to effect lasting outcomes,

ork Access Planning (SNAP), a subcommittee of the Information Policy Council

al Information Systems Advisory Council (IISAC)

y Council (IPC)

mation Systems (GIS) community

up, Information Technology (DIG IT)

s Clarification and Future Progress

evaluate laws, organizational structures, funding structures, budgets, and s data-oriented legislation. This assessment should examine whether the als are still appropriate and evaluate their collective implications.

Finally the state should consider that its defined data goals are far-reaching and strategic, representing a long-term commitment and requiring a long-term solution. It may be best to report progress in terms of timelines for achieving targeted milestones rather than on a cyclical basis.

Appendices

Appendix A: Information Management Principles

Data management is effected by the four Information Management Principles:

Management Principle

Information systems should support an organization's mission. Top management is responsible for linking mission and functions to information systems coherently.

Data Principle

State government data is a resource of the state to be managed and shared across organizational lines.

People Principle

Information is a resource that serves to extend human capabilities to better serve the public.

Standards Principle

State information resources and tools must be managed consistently so that the necessary linkages among state agencies and between state and local government are supported.

Appendix B: IRM Six Critical Success Factors

Executive leadership and involvement

Information resource management requires active support of top management to assume responsibility for managing and developing information resources under their control. Because information resources have agency-wide and statewide value, managing them requires authority and accountability that only executive management can provide. Agency leadership is required for projects to achieve their potential.

Information management policies, standards and guidelines

Agencies need a foundation of policies, standards and guidelines that is aligned with agency purpose and directs the strategic management of information resources. Policies represent business rules and boundaries for managing information in support of an organization's business. Standards measure and control quality. Guidelines provide procedures and best practices for developing, maintaining and controlling information resources. The Office of Technology develops statewide policies, standards and guidelines that provide a framework for agencies to use in developing their own.

Planning

Planning allows an agency to carry out its mission and achieve its purpose. Planning must be conducted from statewide, agency and project perspectives. Statewide planning efforts ensure that collaboration happens and the right projects are done in the right sequence. Agency plans for information resources support major business functions and provide strategic priorities and direction for developing and maintaining information resources. Information resource planning is necessary to allocate and align information resources with the organization's strategic business plan. The planning process also allows agencies to allocate resources to future projects.

Models

Information resource models help an agency manage and share data effectively. Models describe and show relationships between data, business processes, events and technology in support of business functions. Models also help identify effects across functions or systems, and facilitate making decisions, planning projects and managing project scopes. Models also provide critical details in support of product and process quality initiatives.

IRM organization

Accountability and responsibility for information resources should be agency-wide. Agencies should be structured so information resources can be managed and shared appropriately. Agencies must also assess and foster their organization's capabilities to develop, operate and maintain information resources.

Skills

Agencies must invest in acquiring, retaining, training and retraining a skilled workforce of professionals to conduct projects and implement information resources. An appropriate skill base is required before agencies or projects can be successful.

Appendix C: Progress and Future Directions

The following table shows progress to date and future directions for each of the data management components. See also the attached graphic following the table.

	Past Efforts	Recent Progress: 1994-1997	Future Directions and Needs
Enabling Legal Framework	• Data Practices Act • Records Management Act	• Minnesota Digital Signature laws passed in 1997 • Mandated data formats (EDI, federal reporting requirements, etc.) • Information Policy Task Force formed in 1997 • IPC subcommittee on Administrative Information Access Policies formed in 1997	• Legal limitations for data exchange and reuse (federal and state) revisited • Conflicts in Minnesota laws resolved (between efficiency and privacy) • Legal admissibility of records and "trustworthy systems" established • Data exchange authorized • Accountability and funding issues resolved • Inconsistent legal definitions resolved • Data Practices complications resolved
Common Metadata	• Attempt to acquire statewide repository failed to gain community acceptance • Community project to provide a uniform business identifier for companies doing business with the state was not funded • Lack of metadata standards in the past has made fixing the Year 2000 problem more difficult, time consuming and the results more error prone • Metadata "Clearinghouse" (GIS community) with free software for the public sector to create common metadata	• 9 enterprise models identified by agencies are complete or in process • 10 agencies identified building project models • 2 community data models complete (Criminal Justice and Metro GIS) • GIS metadata "Clearinghouse" created • IPC metadata project in process • Agency reuse of models (portions of models created by other agencies)	• Expanded awareness of metadata role in data management • Metadata repositories for statewide use • Expansion of community and statewide models • Library of common metadata and models • Expansion of statewide metadata (also see section on Standards and Guidelines)

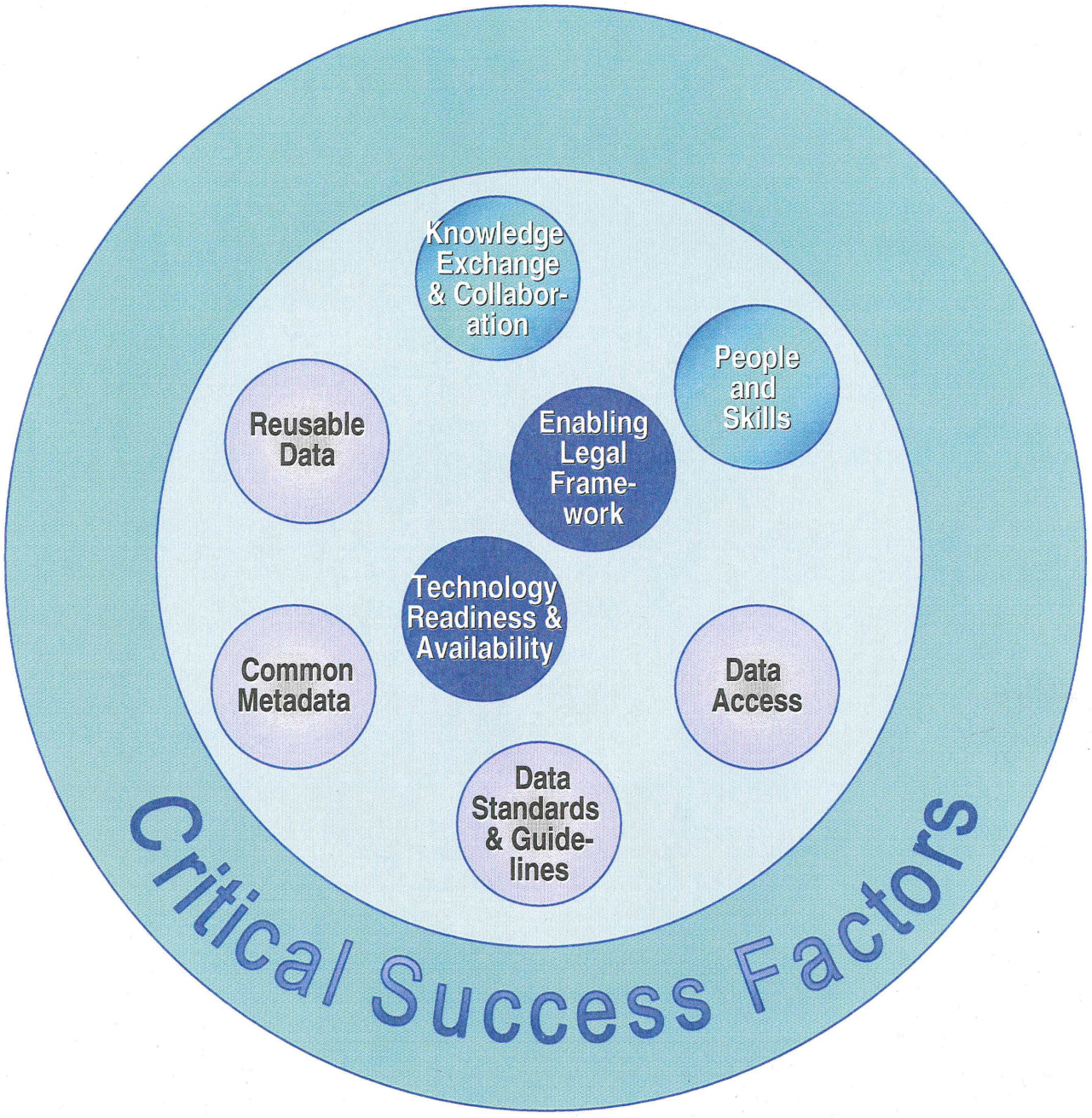
	Past Efforts	Recent Progress: 1994-1997	Future Directions and Needs
Knowledge Exchange and Collaboration	• Only known past efforts to share expertise were through the Computer Symposium: Thriving in the 90s	• Cross-agency participation in "community-wide" models • Agency pool of recording analysts • Data Issues Group (DIG IT) cross-agency data group to exchange knowledge, provide forum for data issues of general interest and leverage work of multiple agencies	• Joint efforts enabled • DIG IT role clarified and strengthened, possibly as IPC subcommittee • Communication vehicles established for information exchange
People and Skills	• Only 1 agency CIO prior to 1993 • No Data Administrators identified prior to 1994 • Few, if any, trained project managers • Few, if any, trained modelers	• 15 agency-identified CIOs • 4 formal data administration positions • 40 IT positions added to address improved IRM and data management • Approximately 500 state agency employees educated and trained in data management, data modeling and project management.	• Mechanism to retain newly trained professionals established (The state can't compete with private sector salaries.) • Continued education in data-related disciplines • Expansion of data administration "community": CIOs and Data Administrators • Increased job recognition for data management professions
Data Access	• Data Practices Act • GIAC	• 4 agency data warehouses built • 2 agency data warehouses planned • 2 agencies anticipate data warehouses in the future (more than 2+ years) • Information Policy Task Force created • IPC subcommittee on Administrative Information Access Policies created • North Star II funded	• Public access requirements clarified • Agency accountability strengthened • Data warehouse guidance and direction • Continued work by Information Policy Task Force and subcommittee • Continued development of North Star II
Technology Readiness and Availability	• Limited availability of enabling technology	• Limited availability of enabling technology (still an issue) • Limited use of repository and warehouse technologies • Limited use of Computer Aided Software Engineering (CASE) and model development tools	• Agency readiness developed: skills, implementation capabilities • Statewide repository with access for all agencies • Statewide data warehouse coordination • Technological capability improved for managing models, data, access, privacy, preservation of electronic records, etc.

	Past Efforts	Recent Progress: 1994-1997	Future Directions and Needs

Data Standards and Guidelines	• Technology standards created by the Department of Administration in the early 1990s	• 7 State of MN Guidelines:
		  ° metadata naming
		  ° IRM modeling
		  ° GIS metadata
		  ° document imaging
		• 5 State of MN Standards created:
		  ° state and county codes
		  ° document imaging
		  ° Year 2000 date data
		• 7 agencies have developed or adopted data standards
		• External standards exist for:
		  ° NCIC (FBI)
		  ° Conference on State Gov't Ethics Regulators (COGEL)
		  ° Federal Information Processing Standard (FIPS)
		  ° HUD
		  ° Post Office
		  ° electronic financial transactions
		  ° electronic commerce in health care: eligibility, enrollment, transfer of records, record privacy (federal)
			• Metadata standards and guidelines developed, compatible with GIS and data archives communities,
			• Standards and guidelines to address data topics (e.g., sharing vs. reuse; appropriate redundancy of data)
			• Statewide architecture
			• Stronger statewide data standards aligned with statewide architecture
			• Strategies for how to comply with, and possibly influence, external standards
			• Statewide data standards aligned with external efforts, such as the Dublin Core (metadata standards), etc.

Reusable Data		• Organizations currently exchanging data with others include:
		  ° Trade and Economic Development
		  ° Criminal Justice community
		  ° GIS community
		  ° Ombudsman for Mental Health
		  ° Economic Security
		  ° Human Services
		• Data exchange may be based on creating redundant copies of data
			• Successful data reuse/exchange defined
			• Measurements for data management progress
			• Mechanisms that promote reuse
			• Legal issues resolved (affecting reuse)
			• Redundant data *collection* addressed
			• Strategies to avoid duplication
			• Statewide data infrastructure – repository, metadata, standards
			• Intersystem communications

A Quest for Data Quality:
Eliminating Redundant Storage of Data



LEGEND:

Significant efforts are underway

Some efforts are underway, but significant work remains

Existing infrastructure in place, but significant work remains