IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA ITB9467
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

Chapter XV

Data Warehousing and the Organization of Governmental Databases

Franklin Maxwell Harper
National Civic League, USA

ABSTRACT

Data warehousing is a technology architecture designed to organize disparate data sources into a single repository of information. As such, it represents a strategy for creating the architecture necessary to support the vision of e-government. Data warehousing enables a new type of "decision intelligence" by providing access to historical trend data, typically difficult to retrieve through operational database systems. Government data warehousing is complex, expensive, and often fraught with data privacy and security issues. E-government goals may be met through a successful data-warehousing project, be it in the form of a more efficient, informed government or as a result of increased public access to information. But given the substantial barriers to success, a thorough planning and investigation process is necessary.

INTRODUCTION

Governments today are inundated by rapidly increasing volumes of data. The source of this data explosion lies in the proliferation of data sources (Hoss, 2001), largely stemming from the recent rise in e-government initiatives (Norris, Fletcher, & Holden, 2001). This could turn out to be either a boon to decision makers or a bane, as the ever-growing store of data holds the potential to improve government decision making, but

could just as easily stagnate and continue to add cost to governmental operations. It is important, therefore, to somehow intelligently manage this data and create an environment where it may be analyzed quickly and cheaply. Optimally, the relevant portions of data could be shared with the public through the Internet. One potential technological approach to this situation, and the approach that this chapter will concern itself with, is known as data warehousing. Data warehousing is a technology architecture designed to organize disparate data sources into a single repository of information. It is intended to leverage data to promote organizational knowledge. For government agencies that are "drowning in data and dying for information," a data warehouse can provide an environment where powerful information comes cheaply and quickly (Singh, 1998).

More specifically, data warehousing is the practice of intelligently managing historical "secondhand" data (Simon, 1997) by periodically copying data from multiple sources into a large database optimized for the extraction of information. A data warehouse enables its users to retrieve useful data regardless of "platform, application, organizational, and other barriers" (Simon, 1997).

To understand more clearly the uniqueness of data warehousing, one must understand the distinction between an *operational* database and an *informational* database (Singh, 1998). Operational databases are far more common — they are read/write sources of current data, and are typically used to support online systems. Informational databases, on the other hand, are typically read-only and contain historical, subject-oriented data optimized for information gathering and analysis. A data warehouse is a type of informational database, intended to enable a discovery-oriented approach to database querying (Inmon, 2002). In this way, data warehousing enables a new type of "decision intelligence" by providing access to historical trend data, typically difficult to retrieve through operational database systems. It can serve as an "unambiguous source of informational truth within the organization" (Poe, 1998).

Data warehouses can be put online via the World Wide Web to enable public access to government information stores. Certainly, some government agencies will have a strong interest in releasing their data to the public in this way. In a survey conducted by the International City/County Management Association (Norris & Demeter, 1999), over 70 percent of all governments responded that they use their web sites for the purpose of information dissemination or citizen education. Citizens most commonly visit government web sites for the purpose of retrieving information, often searching through online databases (Larsen & Rainie, 2002). They believe that the most important potential benefits of e-government have to do with government accountability and greater public access to information (Hart-Teeter, 2000). Government data warehouses provide a service in line with this demand, and extend the potential availability of information well beyond that provided in conventional web-enabled databases.

This chapter is an examination of the practical considerations relevant to government data warehousing, as they relate to improving decision-making, accountability, and public access to information. It will begin with a discussion of some applications of data warehousing relevant to the government sector, along with several warnings about potential impediments to project success. This will be followed be a review of how best practices in the evaluation, planning, implementation, and growth phases of a project can positively impact the success of a project. The chapter will then examine Iowa State's recent data warehousing project as a model for successful implementation through agency collaboration.