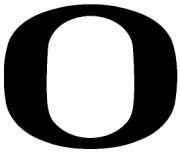

UNIVERSITY OF OREGON
APPLIED INFORMATION MANAGEMENT

Presented to the Interdisciplinary Studies Program:
Applied Information Management and the Graduate School of the University of Oregon
in partial fulfillment of the requirement for the degree of Master of Science

Steps to Data Warehouse Development in K-12 Public Education: A Guide for IT Directors

CAPSTONE REPORT

James E. Rose
Director of Technology Services
Multnomah Education Service District

University of Oregon
Applied Information Management Program

722 SW Second Avenue
Suite 230
Portland, OR 97204
(800) 824-2714

June 2007

Approved by

_____

Dr. Linda F. Ettinger
Academic Director, AIM Program

Abstract

Steps to Data Warehouse Development in K-12 Public Education: A Guide for IT Directors

This study explicates data collection and reporting steps when designing a data warehouse for public education. Literature published from 2002 to 2006 in education-related periodicals concerning data warehouse design and implementation is analyzed. Content analysis is utilized to identify information related to eight steps of data warehouse design for public school districts, proposed by Pare & Elovitz (2005). The outcome is a data warehouse development guide that includes practical examples from the field.

Table of Contents

List of Figures and Tables

Figures

Tables

Chapter I – Purpose of Study

Brief Purpose

The purpose of this study is to explicate the steps to take concerning data collection and reporting when designing a student data warehouse (McIntire, 2002) for use in a kindergarten through 12th grade (K-12) public education context. Literature is analyzed for the inclusion of information that can be categorized within the eight basic steps to establishing a data warehouse as suggested by Pare and Elovitz (2005).

McIntire (2004) explains that "many schools, districts, and states have responded to the No Child Left Behind Act (NCLB) of 2001 by implementing data management systems, including data warehouses, to help them efficiently organize and analyze the enormous amount of student and school performance data they generate each year" (McIntire, 2004, Introduction section, para. 1). Han and Kamber (2001) define a data warehouse as a "repository of information collected from multiple sources, stored under a unified scheme, and which usually resides at a single site" (cited in Pare & Elovitz, 2005 Introduction section, para. 2). Pare and Elovitz (2005) further specify that, "in educational terms all past information available in electronic format about a school or district such as budget, payroll, student achievement and demographics is stored in one location where it can be accessed using a single set of inquiry tools" (Introduction section, para. 2). This one location is defined as a data warehouse.

Levine (2002) explains that "by understanding the basics of data warehousing and steps you can take to enhance data collection and reporting, you can greatly improve your chances of making data-driven decision making a reality." However, the problem identified by this researcher is how to best operationalize enhancement of data collection

and reporting steps. The assumption underlying this study is that the structure suggested by Pare and Elovitz (2005) for eight appropriate steps to data warehouse design (including find appropriate software and hardware, select data sources, determine data importance, convert the information, cleanse the data, import data, determine access, and determine inquiry/analysis toolset) can provide an important procedural framework to the data warehouse design process.

This paper is written for information technology directors working in K-12 public school districts who need to collect data that will be used to build a data warehouse. An Information Technology Director is defined as the professional in charge of planning, organizing and developing an agency's data and data systems (County of Solano, 2006). Small to medium sized school districts, that do not have a strong education service agency to develop a data warehouse for them, often lack the resources to effectively develop and manage one (Borja, 2006, May 4c). As an information technology director in an education service district in Oregon, this researcher observes that for major technology projects like data warehouses, each school district or education service agency determines their need, reviews the available options, and selects a product to use without much input from people outside of their district. Levine (2002) suggests that "everyone's praising data-driven decision making these days, but turning this concept into reality remains a distant dream for many school districts" (Introduction section, para. 1). This paper provides a resource, concerning data collection related to student data, for school district information technology directors to consider as they begin the development of a data warehouse.

This study is designed as a literature review (Taylor & Procter, 2006). Literature published between 2002 and 2006 that describe approaches taken to design a data warehouse in public K-12 schools is collected. The searched materials relate to data warehouse design and K-12 data collection policies. Most collected literature is in the form of articles in periodic journals since no case studies are currently published. The Pare and Elovitz article (which forms the basis of the data analysis process) is published in the April 2005 issue of T.H.E. Journal as part of the "Creating a Framework for High Tech Assessment" theme.

Content analysis (Palmquist, et al., 2004) is utilized to identify information in selected literature that explicates steps used in the design of student data warehouses for public school districts. The content analysis process is guided by application of an approach to data warehouse design developed for use in the education context (Pare & Elovitz, 2005). Steps include such things as selecting data sources to be used and cleansing the data (Pare & Elovitz, 2005). Results of the content analysis are presented in the form of lists that report details of each step, as these are identified in the literature. Lists are categorized in relation to the pre-selected data warehouse design framework provided by Pare & Elovitz (2005).

The outcome of the study is presented in the form of a table, which reframes and presents the explicated eight basic steps to data warehouse design suggested by Pare and Elovitz (2005) in relation to the typical stages of a data warehouse planning process (Mercurius et al.,2004; McIntire, 2002). The intent of the table is to provide technology directors with a guide for the development of a student data warehouse information

collection and reporting plan – one that will better support the decision-making needs of teachers and administrators.

Full Purpose

McIntire (2002) states "With the right information at the right time, school leaders can evaluate the effectiveness of educational programs and create targeted services to meet student and staff needs" (Introduction section, para. 1). The approach in which data is collected and reported can affect the usefulness of the information to the district (Levine, 2002). The infrastructure needed to support data-driven decision-making does not exist in many schools, leaving the potential to affect student performance and other improvement goals likely to be unrealized (Hoff, 2006). The assumption underlying this study is that the development of well designed data warehouses begins to address the infrastructure needs, while at the same time it begins to solve the data quality problem.

Data warehouses are defined in many ways. McIntire (2002) defines a data warehouse as "a consolidated version of data from multiple transactional databases. It contains no unique information of its own. Instead, a data warehouse pulls information from various systems, cleanses the data to eliminate incompatibilities, and then combines it in a uniform format." Similarly, Han and Kamber (2001) characterize data warehouse as a "repository of information collected from multiple sources, stored under a unified scheme, and which usually resides at a single site" (cited in Pare & Elovitz, 2005). In this study the term is used to refer to "a strategic tool that enables staff to look at many types of data to guide decision-making" (Montgomery County Public Schools, 2006, Description, para. 1).

The purpose of this study is to explicate the steps in the design of student data warehouses in public education. As a way to better understand how a data warehouse is designed in K-12 education today, selected literature is examined in relation to the eight

basic steps to establishing a data warehouse as suggested by Pare and Elovitz (2005). While there are other systems that might be used to develop a data warehouse design, the one by Pare and Elovitz (2005) is selected because it is defined explicitly for use within the educational context and the process is published in a major education periodical. T.H.E. Journal is the oldest magazine focused on education and technology (T.H.E. Journal, 2007) and has the largest readership according to its competitor (Readership, Technology and Learning, 2007). The following summarized eight steps to data warehouse design (Pare and Elovitz, 2005), within an educational context, frame the approach taken to data analysis in this study. Fuller descriptions of each step are presented in Appendix A. Steps include:

1. Find the appropriate software, hardware and people to convert existing historical data into a useful format.
2. Select the data sources to be used.
3. Determine what data is important.
4. Convert the information in stand-alone systems to a format that can be queried by OLAP (online analytical processing).
5. Cleanse the data.
6. Import data into the warehouse.
7. Determine who will have access to the data.
8. Determine the toolset (software) that you will use for inquiries and analysis of this data.

As noted by Brown (2006), the amount of work in each phase varies but as a whole the development of a data warehouse is a significant investment for districts. “School leaders

who are considering a data warehouse solution should expect to spend significant time planning prior to jumping in" (Brown, 2006, Planning is key, para. 1).

The steps suggested by Pare and Elovitz (2005) are common across many of the source documents. For example, Pare and Elovitz (2005) suggest in step four that it is necessary to pull data from various sources into a single format. McIntire (2004) further clarifies this by saying "a good data warehousing solution will allow you to integrate data from any source, whether it's your district-wide student information system or informal databases housed on spreadsheets" (Does the solution integrate the information systems you want to explore, para. 1). McIntire (2004) suggests that the additional data allows decision-makers to query the data warehouse to assist in planning.

This study is designed as a literature review as defined by Taylor and Procter (2006). According to Taylor and Procter (2006) a literature review is an account of what has been published on a given topic. Taylor and Procter (2006) further clarify by suggesting that "a literature review is a piece of discursive prose, not a list describing or summarizing one piece of literature after another" (Final notes, para. 1). In this case, literature is collected that explains the definition of a data warehouse, its purpose, suggested design or development advice. The selected literature is primarily from weekly and monthly periodicals that are intended for K-12 administrators and IT directors. Periodical titles include: Education Week, T.H.E. Journal, Technology and Learning, School Administrator, Teacher Magazine, and Scholastic Administrator.

Content analysis is utilized to identify information related to any one of the eight steps used in the design of student data warehouses for public school districts. In particular, conceptual analysis, as defined by Palmquist et.al. (2004), is employed to

establish a process with which to frame and conduct the content analysis. The data set selected for use during conceptual analysis consists of the literature published from 2002 to 2006 in education-related periodicals referring to data warehouse design and implementation. The selected literature is coded for inclusion of information pertaining to the concepts outlined in the eight steps to the design of a data warehouse in a public education setting, provided by Pare and Elovitz (2005). Steps are translated into key coding concepts below (see Figure 1: Eight Key Coding Concepts).

1. Information about selecting software, hardware and people to convert data into a useful format.
2. Information about selecting the most useful data sources.
3. Information about determining the most important data.
4. Information about selecting a common format for information conversion.
5. Information about cleansing the data.
6. Information about importing data into the warehouse.
7. Information about determining access to the data.
8. Information about determining the toolset (software) for data analysis.

Figure 1: Eight Key Coding Concepts

Results of the content analysis are presented in the form of lists, categorized in relation to the pre-selected data warehouse development framework (Pare & Elovitz, 2005). The lists indicate the types of information reported in the literature, related to each data warehouse design step. Themes, key questions, and problem areas are identified within the literature for each concept (Pare and Elovitz, 2005). The lists and themes are presented in tables 9 - 16 in Appendix C. A review of the content of each list should enable IT directors to better understand the data collection and reporting options available to them, as noted in the literature.

The summaries of the themes are focused on improving the IT director's ability to develop a data warehouse that uses data effectively to improve student learning (Bernhardt, 2005). By organizing the summaries of the themes according to typical stages

of a data warehouse planning project, the table is designed to serve as a check-off record when planning a data warehouse development project. The assumption is that by using this table as a guide, school district IT directors will be better able to shorten their planning process by learning from the experiences of their counterparts.

Limitations to the Research

Kreitl (2005) asserts that the term "data warehouse" was unknown in the field of education prior to 2000. Furthermore, Salpeter (2004) suggests that for some districts, the current obsession with data grows out of the need to meet the requirements of the No Child Left Behind Act of 2001 in addition to other accountability-related mandates. For these reasons the study is limited to literature published in the five years from 2002 through 2006. Although outside of the selected time frame, one article, McCright's (2000) "Warehouses Go to School," is included for historical value, because it offers the first chronological reference to school districts' needs to utilize data warehousing tools for reporting on student grades, test scores, attendance, and health records.

The literature selected for review is collected from the ERIC database, online databases through the University of Oregon, the World Wide Web, and public K-12 education-related periodicals. These resources would generally be available to the typical information technology director. The exception is the databases accessible through the University of Oregon library (2007), which according to current borrowing policies limits who can access materials.

The development steps suggested by Pare and Elovitz (2005) are used as initial coding concepts during content analysis because they provide a robust step-by-step framework for developing a data warehouse, presented within the educational context.

Since the use of data warehouses in the field of public education is still relatively new, this researcher believes that it is important to use a framework designed for educators.

Problem Area

Many states are using data-driven decision-making initiatives to improve student performance (Wolf, 2006). States such as Florida are beginning statewide data-driven decision-making efforts to set education policy and practices (Hoff, 2006). Pare and Elovitz (2005, Introduction, para. 6) suggest that:

> The concept of data-driven decision-making is as simple as it sounds. Collecting data, refining the data into a usable format, and basing decisions on the information is the essential concept. Historically, collecting data has not been hard to accomplish in schools. Interpreting and processing data so it can be considered in the decision-making process has been the difficult part. When computers first became available, individuals were needed to write custom programs to pull the data from student record systems. Unfortunately, this data extraction did not generally yield information in a useful format. Another program or set of programs was then needed to manipulate the data so that it could be useful.

New Federal legislation in 2001 increased that need for data collection and manipulation (Snow-Renner & Torrence, 2002) within the public schools. Trotter (2006) suggests that the No Child Left Behind Act has motivated school districts to use data in new ways to increase student achievement. "At the local level, the No Child Left Behind Act is inspiring a small fraction of the country's nearly 15,000 school districts to seek

technological solutions to meet the law's ambitious goals for student learning" (Hoff, 2006, Introduction section, para. 7). Technological solutions are empowering administrators to make decisions based on hard data. According to Trotter (2006), Federal officials indicate that the additional benefit of data-driven decision-making is that resources are used more efficiently, management is better organized, and instruction can be tailored to individual students.

"In education, we are constantly looking for the magic bullet – that intervention which will miraculously result in higher achievement scores for our students, happier and more productive teachers, etc." (Pare & Elovitz, 2005, Introduction section, para. 1). This is not a simple task but as Levine (2002) points out:

> The journey of 1,000 questions about academic success begins with a single "How are we doing?" The desire to integrate all meaningful and available information into the decision-making process is sound. However, the path to collecting meaningful information and reporting that data is not as clear. (The bottom line, para. 1)

In an Education Commission of the States policy brief, Snow-Renner and Torrence (2002) explain that in the last 10 years new means of measuring student outcomes have been developed along with new systems that allow data to be collected and stored more efficiently. Unfortunately, districts often have a "mishmash of formal and informal systems that add up to an ineffective and invalid collection of data" (McIntire, 2002, The data was kept in one of several parallel systems, para. 3). "The goal of a data warehouse is to allow users to explore relationships between information bits that were previously disconnected" (McIntire, 2004, 2. Does the data model fit your

needs?, para 1). “While a few districts have been successful in implementing data warehousing, many others are either reluctant to proceed or have made mistakes that impede their ability to use this information” (Levine, 2002, The bottom line, para. 2). Since data warehouses are still relatively new, further development of the infrastructure, applications, and training are needed (Hoff, 2006).

Many districts continue to suffer from disparate data collection systems that do not share data in the desired manner. As McIntire (2002) reported of his own experiences,

> When we began our evaluation of data management in the district, we thought that most of our information systems worked reasonably well. Our investigation revealed that a lack of clear and comprehensive data management standards had allowed us to devolve into a mess of incompatible parallel systems that gradually lost sync with each other. As a result, our ability to use data to make well-informed decisions was severely compromised. (The data was kept in one of several parallel systems, para. 2)

Since data warehouse use in public education is a relatively new occurrence, there are few case studies indicating best practices. As Levine (2002) explains “how data warehousing works can be complicated and intimidating” (Introduction, para. 2). Tying meaningful student data to student performance remains a challenge for school administrators. As McIntire (2004) suggests, key decision-makers cannot make test hypotheses without the ability to “tease the nuances” from the information in the data warehouse. Pare and Elovitz (2005) suggest “the answer lies in the ability of educational leaders to collect, store, analyze and effectively utilize large amounts of data to inform

their decision-making. This is where data warehousing can help" (Introduction section, para. 1). The need of district administrators to make decisions based on data drives the adoption of data warehousing.

However, as Brown (2006) suggests, adapting data warehouses to a school district can be difficult. In response, "technology companies are scrambling to create or revamp web-based products in a bid to provide fast efficient, user friendly data tools that help educators bolster the bottom line for schools: student academic performance" (Borja, 2006c, Introduction, para. 1). According to McIntire (2004), this expansion benefits school and district leaders by providing options and creating competition. However, McIntire explains it also makes choosing the right solution more difficult. Borja (2006c) suggests that once a district selects an application, it needs trained staff to manage the data collection and presentation application.

"For technology leaders committed to supporting data-driven decision making in their districts and schools, there are a variety of technical challenges to overcome" (Salpeter, 2004). Although the technical challenges exist, Pare and Elovitz (2005) indicate that improved hardware and software have made it easier to extract and analyze data and present it in a useful format. The challenge to find the right data collection solution is often left to the IT staff (Wolf, 2006). This study is developed to assist school district information technology directors in developing the deployment plan for their own data warehouse. Ultimately, "The goal is to use the information to adjust classroom instruction and to allocate resources more effectively to bolster learning" (Borja, 2006b).

Chapter II – Review of References

This Review of References encompasses the principal references used to form the content of this research paper, including documents that are considered most important in framing the purpose and need. This chapter is presented in the form of an annotated bibliography. Each article is summarized with the key topics used, the portions of this paper shaped by the article, and why this researcher chose to use this resource.

Borja, R. R. (2006, May 4). Risk and reward. Education Week. 25. 42-44, 46-47.

This article by Rhea Borja provides a broad view of the current state of technology use in public education including discussions of data warehouses, student information systems, and assessment systems. The article explains the current state of many school districts as not utilizing technology fully and describes how technology companies are trying to fill the void with products. Borja's knowledge of the current state of data use in public education serves as the basis for inclusion in the *Brief Purpose* explanation of need for this study and for describing the target audience.

The author is a business and technology staff writer for Education Week and is published in USA Today and on eSchoolNews online. As a senior writer for Technology Counts, Education Week's annual technology trends issue, Borja is often cited for her work writing about technology trends in education.

Levine, E. (2002, November). Building a data warehouse. American School Board Journal. Retrieved March 23, 2007, from the World Wide Web: http://www.asbj.com/2002/11/1102technologyfocus2.html

This article by Elliot Levine explains what a data warehouse is, what the Schools Interoperability Framework (SIF) does and the steps to building a successful data

warehouse. As an article published in the American School Board Journal, it is written to provide non-technical people with a basic idea of how to support a data warehouse. Levine's work is included in this literature review to explain the difficulties faced by school districts in implementing a data warehouse. His steps to success also offer an alternative to the eight basic steps suggested by Pare and Elovitz and is included in the *Analysis of Data* chapter.

This article is included because of the national prominence of the author. Levine has led multiple successful education-related technology and communication companies. He has worked for companies such Hewlett-Packard, SonicWall, and the JASON Foundation. This experience provides a strong foundation for understanding the pitfalls that school districts can encounter when implementing technology projects.

McCright, J.S. (2000, July 3). Warehouses go to school. eWeek. Retrieved April 17, 2007, from the World Wide Web: http://content.epnet.com/ContentServer.asp?T=P&P=AN&K=3378778&EbscoContent=dGJyMNXb4kSeqLM4v%2BbwOLCmrk%2Bep7ZSs6%2B4Sa%2BWxWXS&ContentCustomer=dGJyMPGuslGwrbVNuePfgeyx%2BEu3q64A&D=aph

This article addresses the installation of a data warehouse in Broward County Public Schools in Fort Lauderdale, Florida. The focus on Broward County Public Schools includes a look at the vendors and applications used as well as the steps taken by the IT director in building the data warehouse. McCright's article serves as the beginning point for data warehouse references found and is noted as an exception to the timeframe included in the *Limitations to the Research* section.

John McCright serves as a department editor for the online publication eWeek and is widely quoted in that capacity. His technology writing career includes writing about both business- and education-related technology topics.

McIntire, T. (2002, June 15). The administrator's guide to data-driven decision making. Technology and Learning. Retrieved March 26, 2007, from the World Wide Web: http://www.techlearning.com/db_area/archives/TL/2002/06/guide.php

This article by Todd McIntire focuses on the needs of school administrators to make decisions based on data. As a three-part article it looks at the administrator's needs, explains what a data warehouse is, and provides a glossary of data mining terms. This article is used to explain the need for this study and as a comparison to the eight steps process of Pare and Elovitz introduced in the *Brief Purpose*. Many of the data mining terms are included in the *Definition of Terms*.

This article is selected because the author, Todd McIntire, is frequently published in trade magazines and serves as a vice-president of Edison Schools, Inc., a for-profit school management company. His experience in public education and the private sector provide a foundation for his work using technology to improve student achievement. His work has been included in numerous publications including Technology and Learning and in SchoolCIO.

McIntire, T. (2004, August 16). Eight buying tips: Data warehouses. Technology and Learning. Retrieved March 26, 2007, from the World Wide Web: http://www.techlearning.com/showArticle.php?articleID=26806926

This second article by Todd McIntire provides practical advice on purchasing and implementing a data warehouse in a public education environment. The article includes seven attributes that a data warehouse should offer and a listing of names of major data warehouse providers. The buying tips are used for comparison purposes with the Pare and Elovitz basic steps in the *Full Purpose* section. This article is used in the *Problem Area* to define the goal of a data warehouse in public education.

The author's experience in both public and private management of public education institutions (noted in the entry above) is a strong background for providing the practical advice in the eight buying tips.

Mercurius, N., Burton, G., Hopkins, B., & Larsen, H. (2004, January). Putting the school interoperability framework to the test. T.H.E. Journal. Retrieved April 17, 2007, from the World Wide Web: http://www.thejournal.com/articles/16570

This article explains how the Schools Interoperability Framework was utilized by the Jurupa Unified School District in Riverside, California. Sub-topics included phases to success, project implementation, software communication and lessons learned. The article explains why the development of the data warehouse was slow. This article is used for comparison purposes in this study in the *Data Presentation* section.

Dr. Mercurius is the Director of Information/Education Technology and Assessment for the Jurupa Unified School District. In addition to his work for the school district, Mercurius has reviewed books with titles ranging from The Virtual High School: Teaching Generation V to Strengths Quest: Discover and Develop Your Strengths in Academics, Career, and Beyond. The hands on experiences of Mercurius and his co-authors/co-workers are valuable to the target audience of this document.

Palmquist, M., Busch, C., De Maret, P., Flynn, T., Kellum, R., Le, S., Meyers, B., Saunders, M., White, R. (2005). Content Analysis. Retrieved April 9, 2007, from Colorado State University, Department of English Web site: http://writing.colostate.edu/guides/research/content/.

This document describes the process for conducting content analysis in academic research. The document discusses types of content analysis, methods for conducting content analysis and the specific steps a researcher should use in planning and executing

a content analysis. The eight suggested steps for conducting a content analysis are used by this researcher as the foundation for the method section of this literature review.

This work is selected based on it being a recommended resource of the UO Applied Information Management program.

Pare, R. & Elovitz, L. (2005, April). Data warehousing: An aid to decision-making. T.H.E. Journal. Retrieved March 21, 2007, from the World Wide Web: http://www.thejournal.com/articles/17217/

The eight steps to establishing a data warehouse presented in this article serve as the basis for this study. The article explains what a data warehouse is and how it is used. In addition, the article provides details to an eight step process to establishing a data warehouse. As the central work for comparison in this study, the Pare and Elovitz article is used throughout this document.

This work is selected because it provides the most specific list of steps to establish a data warehouse of the recent available documents. The authors combined experience indicates an expertise in the education technology field. Dr. Roland Pare is the Director of Information Systems for Hunterdon Central Regional High School in New Jersey. The school is nationally recognized for its use of technology and data to improve student achievement. Dr. Pare also writes software reviews for Multimedia Schools Magazine.

Dr. Leonard Elovitz is the chair of the Educational Leadership Department at Kean University in New Jersey. He has authored more than a dozen articles in professional journals. His work has been published in American Association of School Administrators (AASA) publications and T.H.E. Journal. His teaching and administrative experience extend back three decades.

Chapter III - Method

This study is designed as a literature review (Taylor & Procter, 2006). According to Taylor and Procter (2006) a literature review is an account of what has been published on a given topic and should synthesize and evaluate the material according to the concept of the research question. Taylor and Procter (2006) additionally suggest that a literature review should identify themes present in the body of work. In this study, the researcher is looking for themes regarding common steps in establishing a data warehouse. The goal is to identify categories of information that are addressed in relation to the eight basic steps as identified by Pare and Elovitz (2005).

Currently there are no case studies, per se, indicating best practices of establishing a data warehouse. However, there are many published articles explaining the approaches taken to development of data warehouses in the public K-12 education system. Therefore, the literature included in this review is pulled from education-related periodicals. Since data warehousing is relatively new to education (Dougherty, 2001), many of the articles are intended to first introduce decision-makers to the topic (Borja, 2006b) and then address the steps to follow.

Literature Collection

Four resources are used to collect applicable texts. They include the ERIC database, online databases through the UO, the World Wide Web, and trade periodicals. These sources are characterized by this researcher in the following manner:

- ERIC Database articles are prescreened for applicability to the education community and include journals, periodicals, and abstracts.

- Online databases through the University of Oregon include journals, periodicals, and abstracts. Selection is based on direct relevance to public K-12 education organizations.
- World Wide Web resources are used when referenced by other articles. Articles are included if the resource is recognized by ERIC or another education-related databases available through the University of Oregon library.
- Public K-12 education-related periodicals received by the Multnomah Education Service District, Oregon School Board Association, or the Confederation of Oregon School Administrators. Titles include Technology and Learning, T.H.E. Journal, and Education Week.

Relying on a controlled vocabulary as listed in Figure 3.1, each of the previously identified sources is reviewed for applicable articles.

Data warehouse	AND	Education		
Dashboard	AND	Education		
Data warehousing	AND	Education		
Data warehouse	AND	School		
Dashboard	AND	School		
Data warehousing	AND	School		
School	AND	Data-driven	AND	Decision-making

Figure 3.1 – Terms used in literature search

Within the search results of online databases the researcher reviews the total of returned responses or the top 100 responses if the total exceeded 100. This allows for the most relevant articles to be reviewed.

For trade periodicals, the available issues between 2002 and 2006, the years included in this literature review, are searched for articles relating to the controlled

vocabulary. The following titles are searched in hardcopy form: American School Board Journal, School Administrator, Technology and Learning, and Education Week. After finding articles within ERIC from these periodicals; T.H.E. Journal, Teacher Magazine, and Scholastic Administrator, the online versions are searched for the same time period to ensure all relevant articles are retrieved.

The search of the various sources yielded 31 articles. The majority of pertinent literature are from the ERIC database and EBSCOhost. Fifteen of the articles are selected for informational and background purposes. The remaining 16 articles serve as the data set for the content analysis.

Data Analysis

To help investigate the selected texts, content analysis is utilized to identify the types of information typically used in the design of student data warehouses for public school districts. Krippendorff (2004) suggests that "content analysis entails a systematic reading of a body of texts, images, and symbolic matter, not necessary from an author's or user's perspective" (p. 3). Krippendorff (2004) further defines content analysis as "a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use" (p. 18).

In particular, conceptual analysis as defined by Palmquist et.al. (2004) is employed to establish a process with which to frame and conduct the content analysis. The data set selected for coding includes 20 articles published between 2002 and 2006 with one exception. These articles address K-12 public school data warehouse development needs. Articles range from data warehouse buying tips (McIntire, 2004) to explanations of the strategic value to providing more data to teachers (Bernhardt, 2005).

This literature is examined according to the following plan in order to identify which of the eight steps to designing data warehouse, defined by Pare and Elovitz (2005), are seen described in the selected literature.

This researcher follows the eight steps for conducting conceptual analysis as described by Palmquist et.al. (2004). The coding steps include:

1. Decide the level of analysis – Since the language of data warehouse is evolving this study examines selected literature at the concept level. Concepts are framed by the eight steps to designing a data warehouse, as defined by Pare and Elovitz (2005).

2. Decide how many concepts to code for – All eight concepts provided by Pare and Elovitz (2005) are coded for (see Appendix A for full descriptions). These concepts serve as a pre-determined set of coding concepts.

3. Decide whether to code for existence or frequency of a concept – The coding process is designed to identify the existence of the defined concepts.

4. Decide on how to distinguish among the concepts – The researcher allows for variations of concepts and does not require identical descriptions of the steps. For example, while Pare and Elovitz suggest in step 7 that you "determine who will have access to the data," Brown (2006) advises to know your audience. Both are speaking of the need to provide appropriate access to the end users from parents to teachers to administrators.

5. Develop the translation rules for coding the selected texts - Texts are coded for concepts framed on the eight steps (Pare & Elovitz, 2005). Variations are examined in relation to a set of operational definitions provided in this study (see Appendix A) and in relation to context within each selected source.

6. Decide what to do with "Irrelevant" information – Text that does not address one of the eight steps to the design of a data warehouse is considered irrelevant text. For example, text that addresses what states are doing or how to use a data warehouse to improve instruction is considered irrelevant. Irrelevant text is ignored.

7. Code the texts – Texts are read and coded by hand by the researcher, using a spreadsheet to tally results. The spreadsheet allows documentation of the coding concepts in columns and the selected literature in rows. A separate spreadsheet is maintained for each concept. The spreadsheet columns include article referenced, concept description, and location in article.

Data Presentation

8. Analyze the results – Results of the content analysis are presented in the form of tables, categorized in relation to the eight pre-selected set of data warehousing development steps as defined by Pare and Elovitz (2005) in Data warehousing: An aid to decision-making. A quick review of these tables should indicate not only how each concept is addressed in the literature, but also whether or not information related to each of the eight concepts is found in the selected literature and how the focus is distributed among the eight steps. The coding results are presented in Chapter Four – Analysis of Data. Figure 3.2 illustrates the spreadsheet layout for concept one. The spreadsheet for each of the concepts is included as Tables 1 – 8 in Appendix B.

Concept #1	Concept Description	Article Author	Location in Article
Information about selecting software, hardware and people to convert data into a useful format.			

Figure 3.2 – Format of data collection spreadsheet

The outcome of the study is presented in the form of a series of tables, which re-frame the explicated eight basic steps to data warehouse design presented by Pare and Elovitz (2005) in two different ways. First, data in the set of results tables (Tables 1 – 8 in Appendix B) is redesigned in a way that presents the information in relation to a set of overarching themes. The themes observed include topics covered, questions asked, and known problem areas. Tables 9-16 in Appendix C form the basis for discussion in Chapter Five – Conclusions. Figure 3.3 illustrates the spreadsheet layout for concept one.

Planning Step #1	Topics	Key Questions	Problem Areas
#1. Information about selecting software, hardware and people to convert data into a useful format.			

Figure 3.3 – Format of spreadsheet for theme examination

The Conclusion section of the study (see Chapter Five) is a discussion of the results designed to provide IT Directors with a brief explanation of the themes framed in relation to the suggested steps of a data warehouse planning process (Pare & Elovitz, 2005). When combined with the set of tables provided in Appendix C, the final outcome of the study presents IT directors with a guide for the development of a student data warehouse information collection and reporting plan.

Data warehouse planning processes vary by district although some share similar components. Mercurius et. al. (2004) suggests that school districts follow a three phase planning process. The phases include phase one; detailed assessment and documentation of data collection processes, phase two; development of common data elements, and phase three; building a robust reporting platform. These phases are similar to the four

stages for organizing information systems to support data-driven decision making proposed by McIntire (2002). McIntire expands his stages to four by including a revision and new strategy stage. The Mercurius et. al. (2004) phases roughly translate to the eight steps suggested by Pare and Elovitz (2005). The belief is that by using this guide, as the outcome of this study, school district IT directors will be better able to shorten and improve their planning process by learning from the experiences of other districts.

Chapter IV – Analysis of Data

The data analysis process is described in the Data Analysis section of the Method chapter, and relates to steps seven and eight of the steps for conducting conceptual analysis as described by Palmquist et.al. (2004). The intent during data analysis is to see how a set of eight concepts concerning establishment of a data warehouse are addressed in the selected literature.

Sixteen selected pieces of literature (published by fourteen different authors) made up the set for coding. These were reviewed as one body of information. The researcher first reviewed the primary source in order to clearly understand each of the "eight steps to establishing a data warehouse," as described by Pare and Elovitz (2005). The remaining documents were then coded in alphabetical order by the last name of the first listed author. Multiple documents from a single author were coded chronologically. Coded texts included:

Pare, R. & Elovitz, L. (2005, April). Data warehousing: An aid to decision-making. T.H.E. Journal. Retrieved March 21, 2007, from the World Wide Web: http://www.thejournal.com/articles/17217/

Bernhardt, V. L. (2005, February). Data tools for school improvement. Educational Leadership. Retrieved April 17, 2007, from the World Wide Web: ftp://eff.csuchico.edu/Downloads/ArticlesOfNote/DataTools.pdf

Borja, R. R. (2006, May 4a). Aware of all students. Education Week. 25. 32.

Borja, R. R. (2006, May 4b). District initiative. Education Week. 25. 24-26, 28, 30-31.

Borja, R. R. (2006, May 4c). Risk and reward. Education Week. 25. 42-44, 46-47.

Brown, J. (2006, October). Too much Information. T.H.E. Journal. Retrieved April 17, 2007, from EbscoHost on-line database.

Flaherty, W. H. (2004, July). Data warehouse helps Hanover County Public Schools raise student achievement. T.H.E. Journal. Retrieved April 17, 2007, from the World Wide Web: http://thejournal.com/articles/16838

Hoff, D. J. (2006, May 4b). Keeping track. Education Week. 25. 16-18.

Kreitl, P. (2005, April). Building our data warehouse. School Administrator. Retrieved April 17, 2007, from the World Wide Web: http://www.aasa.org/publications/content.cfm?ItemNumber=1287

Levine, E. (2002, November). Building a data warehouse. American School Board Journal. Retrieved March 23, 2007, from the World Wide Web: http://www.asbj.com/2002/11/1102technologyfocus2.html

McCright, J.S. (2000, July 3). Warehouses go to school. eWeek. Retrieved April 17, 2007, from the World Wide Web: http://content.epnet.com/ContentServer.asp?T=P&P=AN&K=3378778&EbscoContent=dGJyMNXb4kSeqLM4v%2BbwOLCmrk%2Bep7ZSs6%2B4Sa%2BWxWXS&ContentCustomer=dGJyMPGuslGwrbVNuePfgeyx%2BEu3q64A&D=aph

McIntire, T. (2002, June 15). The administrator's guide to data-driven decision making. Technology and Learning. Retrieved March 26, 2007, from the World Wide Web: http://www.techlearning.com/db_area/archives/TL/2002/06/guide.php

McIntire, T. (2004, August 16). Eight buying tips: Data warehouses. Technology and Learning. Retrieved March 26, 2007, from the World Wide Web: http://www.techlearning.com/showArticle.php?articleID=26806926

Olson, L. (2007, February 14). Data-wise school systems seen as sharing key traits. Education Week. Retrieved March 25, 2007, from the World Wide Web: http://www.edweek.org/ew/articles/2007/02/14/23data.h26.html

Rudner, L. M. & Boston, C. (2003, February). Data warehousing: Beyond disaggregation. Educational Leadership. Retrieved April 17, 2007, from the World Wide Web: http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=9029507&loginpage=Login.asp&site=ehost-live

Salpeter, J. (2004, March 15). Data: Mining with a mission. Technology and Learning. Retrieved March 26, 2007, from the World Wide Web: http://www.techlearning.com/shared/printableArticle.php?articleID=18311595

Raw Results

The results of this analysis are presented in Appendixes B and C. Appendix B includes Tables 1-8, which present the findings of the search for information related to the eight steps to establishing a data warehouse suggested by Pare and Elovitz (2005). These tables show which authors covered a particular step in their article and what they had to say about it.

Each concept is found in at least three articles. Most concepts are found in eight or nine articles. Concept six, information about importing data into the warehouse, has the fewest references with only three, while concept eight, information about determining the toolset for data analysis is the most prolific with 13 references.

The distribution of concepts across the articles varies across articles, ranging from the inclusion of one (of a potential total of eight) identified concept to seven concepts. Nine articles include three or fewer concepts. Of the remaining articles, Levine (2002) addresses the most concepts with seven, but did not include concept seven, information about determining access to the data.

When addressing concept one, ***information about selecting software, hardware and people to convert data into a useful format***, most authors suggest choosing software carefully (Kreitl, 2005; Hoff, 2006). In concept two; ***information about selecting the most useful data sources***, a common theme is to conduct evaluations of current systems (McIntire, 2002). Borja (2006b) suggests for the third concept; ***information about determining the most important data***, it is a critical and time consuming process to compile the data from various sources. From the four authors that address concept four; ***information about selecting a common format for information conversion***, the common

suggestion is that the data needs to put in a compatible and non-redundant format (Salpeter, 2004). Found in nine of the articles, concept five; ***information about cleansing the data***, is consistently described as part of the planning process to ensure good results (Levine, 2002). Although only found in three articles, McIntire (2004) and Levine (2002) find concept six; ***information about importing data into the warehouse***, to be important. For the authors that include concept seven; ***information about determining access to the data***, the need to know your audience is seen as vital to the data warehouse (Brown, 2006). The final concept, number eight; ***information about determining the toolset (software) for data analysis***, is the most frequently assessed concept, as most authors agree that an easy-to-use online analytical tool for reporting is needed (McIntire, 2004).

The final outcome of the study is found in Appendix C and is designed as a guide for IT Directors. Appendix C includes Tables 9-16, which presents a pattern analysis of the concepts identified in Appendix B. Patterns are revealed for the following three categories of information concerning the explication of steps to designing a data warehouse: (1) topics covered, (2) questions asked, and (3) known problem areas. The first pattern set, topics covered, is developed by reducing the elements identified in Appendix B and collapsing them into a final set. The second pattern set, questions asked, frames each topic in the form of a key question that should be asked by the information technology director. The third pattern set, problem areas, lists difficulties experienced by school districts as reported in the analyzed articles. A full discussion of Appendix C is presented in the Conclusion chapter of this paper.

Chapter V - Conclusions

This study is intended to provide IT Directors a guide to planning a data warehouse project. The purpose of this guide is to explicate the necessary steps to designing a data warehouse for use in K-12 public education. Literature is analyzed for the inclusion of the eight basic steps to establishing a data warehouse as suggested by Pare and Elovitz (2005).

The final outcome of the study is represented in Appendix C, as a guide for IT directors. Eight tables, that explicate the primary steps to planning a data warehouse development project, are presented and include presentation of the following patterns of information: (1) the topics related to each concept derived from Pare and Elovitz (2005), (2) the key questions that IT directors should ask in the planning process related to these topics, and (3) the common problem areas faced by IT directors, related to each step.

In step one of the data warehouse planning process, IT directors need to address the need for trained staff and sufficient infrastructure to support the data warehouse. Topics to examine in this step include: what software and hardware are available or needed to house a data warehouse, what staff is trained to support a data warehouse and how do they convert data into a useful format. The key goal in this step is selecting the right software that meets the districts' needs and training staff to use it properly. As a way to reach the goal, IT directors should ask questions related to current staff experience with data warehouses, server and network capacity, and how the software should be acquired; purchase or develop in-house. Problem areas include on-going staff development needs, adequate stress testing of infrastructure, and choosing the right software application the first time.

In step two of the planning process, IT Directors must determine the most useful data sources. For the second step, a common and known difficulty is simply finding the right data sources. Some authors suggest that schools store information in many locations and various media types which makes the consolidation process difficult. IT Directors need to determine if the data is available and accurate. Where and how the data is stored becomes a concern also. Common problems include variations in how the data was originally input by end users and data that is stored in non-digital formats.

In a related step, number three, determining what data is critical is the key question. IT Directors need to know what data is needed and if it is worth the time and effort to collect it. In some cases it may not be. Step three problem areas are primarily related to excessive data and the time consuming nature of compiling the data.

In step four of the planning process, IT Directors must determine a common data schema to put all data in. The common format issue raised by step four needs to be addressed by a data warehouse development plan. Some authors, including Borja and Brown, advocate the use of the Schools Interoperability Framework (SIF) as a means of addressing the common format issue. IT directors are warned that not all "SIF Compliant" software actually is compliant.

Of the eight steps, step five about data cleansing has the most closely worded references across the articles. It is clear, based on analysis of this body of literature, that taking the time to cleanse data is critical to the accuracy and value of subsequent reports. Key questions facing IT Directors include how to cleanse the data to prevent errors in the results, how much cleansing is needed and appropriate, and to ensure data integrity in the cleansing process. Like step three, the time it takes to cleanse data can be problematic.

In step six of the planning process, IT Directors must extract information from databases and import it into the data warehouse. Step six appeared in the fewest articles yet is the crucial step in building a data warehouse. If you do not populate the database, there is nothing to analyze. IT directors need to know how their data sources (student information systems, financial systems) will facilitate the importing process. Another key importing question is how frequently the data will be uploaded. The data must be current to have the most relevance to decision-makers.

Since security is such a concern for student records, step seven, which is focused on determining who has access to the data, is important for IT directors to review. The key points are who needs access and what level of access do they require to do their job. It is important to decide who, in the education community, will have access. Do teachers and parents need access? At what level? The access levels need to match the intended audience. A problem facing IT Directors is to maintain security while providing needed access.

In step eight of the planning process, IT Directors must select a toolset for data analysis. The most common concept among the articles was the final step, number eight. A data warehouse is only as effective as the data and knowledge that can be pulled from it. Key questions include how will the data be used and how should results be displayed? The tools need to be easy and planning is imperative to selecting the right tool.

Pare and Elovitzs' eight steps to data warehouse design (2005) serve as a good framework for IT directors beginning the data warehouse development planning process. The experiences of the cited authors provide practical advice and guidance that should help school districts save time and resources by avoiding known problems.

Appendix A – Eight Basic Steps to Establishing a Data Warehouse

Pare and Elovitz (2005) describe eight basic steps to establishing a data warehouse as the following:

1. Find the appropriate software, hardware and people to convert existing historical data into a useful format where, according to current terminology, data can be "mined."
2. Select the data sources to be used. These are usually stand-alone databases containing information in areas such as budget, student achievement and standardized-test results.
3. Determine what data is important. While conceptually there is no limit to the amount of information that can be included, there are practical limits to what information decision-makers will want to examine and how much effort they want to put into the most difficult phase, which is data cleansing.
4. Convert the information in stand-alone systems to a format that can be queried by OLAP (online analytical processing).
5. Cleanse the data. This is perhaps the most time-consuming part of the process. It is amazing how much bad information can exist in a database. For example, when incomplete grades are not made up, the student record may still contain an "I" even though the extracted data is expected to be in a numerical format. The board policy may indicate that all incomplete grades be changed to 59 if they are not made up in the required time, but many never get updated. An error rate of 0.5% means a lot of corrections when dealing with millions of pieces of data. Even worse is that each error must be individually researched to resolve the problem.

6. Bring the data into the warehouse. This is a very important step because it determines what links will exist between data fields. For example, if individual scores on standardized tests are brought in as part of a student's record (unless the test date was also tied to this record), one would be able to run a correlation of student class rank and standardized test score, but would not be able to compute a simple average test score by year. The desired relationships should be considered carefully, and the original data should never be destroyed lest it be discovered that something was missed.
7. Determine who will have access to the data. Much of the data is confidential and cannot be made available universally to all members of the school community. If planned for in the beginning, levels of access can be controlled by password to protect the rights of individuals.
8. Determine the toolset (software) that you will use for inquiries and analysis of this data. Ease of use and standardization are the keys to proper selection.

Appendix B – Tables 1-8 Coding Results:

Explication of the 8 Steps by Pare & Elovitz (2005)

Table 1 – Data collection spreadsheet for concept 1

Concept #1	Concept Description	Article Author	Location in Article
Information about selecting software, hardware and people to convert data into a useful format.	Find the appropriate software, hardware and people to convert existing historical data into a useful format where, according to current terminology, data can be "mined."	Pare & Elovitz (2005)	8 Basic steps to Establishing a Data Warehouse (para. 1)
	Districts need at least a small trained staff to manage a data warehouse.	Borja (2006c)	Cost Can Be an Obstacle (para.3)
	Make sure infrastructure is up to the task.	Brown (2006)	A Smooth Transition (para. 8)
	Determine primary data warehouse software.	Flaherty (2004)	Timely Test Results (para. 3)
	Software selection is important to refining and improving the tool in subsequent years.	Hoff (2006)	"Connections" Coming (para. 10)
	Choose software carefully and consider creating a solution in house.	Kreitl (2005)	Desktop Ease (para. 1)
	Carefully review software/data warehouse providers for benefits associated with their size.	Levine (2002)	Steps to Success (para. 6)

	First consolidate appropriate Information Technology resources in one department including hardware and personnel.	McCright (2000)	New Bundles help organize, disseminate data (para. 13)

Table 2 – Data collection spreadsheet for concept 2

Concept #2	Concept Description	Article Author	Location in Article
Information about selecting the most useful data sources.	Select the data sources to be used. These are usually stand-alone databases containing information in areas such as budget, student achievement and standardized-test results.	Pare & Elovitz (2005)	8 Basic steps to Establishing a Data Warehouse (para. 1)
	It is not easy to gather the information which is contained in various databases, filing cabinets and folders.	Borja (2006c)	Benchmark-Test Market Blossoms (para. 7)
	To be effective the data must be present so selecting sources is important.	Brown (2006)	The Quality of Your Data is Critical to the Processes of Storing It (para. 3)
	Teachers currently collect data in individual spreadsheets and need a common data storage system.	Hoff (2006)	"Connections" Coming (para. 5)
	Determine value of each current data collection and evaluate need for a data warehouse based on the collected data.	Levine (2002)	Steps to Success (para. 2)
	Conduct evaluations of current database systems.	McIntire (2002)	Introduction (para. 2)

	Look for a solution that will allow integration of data from any source (databases, spreadsheets or SIS).	McIntire (2004)	Does the Solution integrate the information systems you want to explore? (para. 1)
	Conduct an information inventory	Rudner & Boston (2003)	Prerequisites

Table 3 – Data collection spreadsheet for concept 3

Concept #3	Concept Description	Article Author	Location in Article
Information about determining the most important data.	Determine what data is important. While conceptually there is no limit to the amount of information that can be included, there are practical limits to what information decision-makers will want to examine and how much effort they want to put into the most difficult phase, which is data cleansing.	Pare & Elovitz (2005)	8 Basic steps to Establishing a Data Warehouse (para. 1)
	Compiling the district's data from various source databases is critical and time consuming	Borja (2006b)	Actionable Information (para. 5.)
	To be effective the right data must be included so districts need to spend time on getting the right data elements selected.	Brown (2006)	The Quality of Your Data is Critical to the Processes of Storing It (para. 3)
	As a first step the district Superintendent wanted to track specific data elements.	Kreitl (2005)	Imagined Uses (para. 2)
	List important data elements currently collected and add new fields as needed prior to data warehouse implementation.	Levine (2002)	Steps to Success (para. 3)

	Determine what data is currently being collected and by whom.	McIntire (2002)	Introduction (para. 2)
	Every piece of information related to a student's performance should be available for analysis.	McIntire (2004)	What data domains are supported? (para. 1)
	School systems need to devote time to selecting the right data to collect aligned with the data warehouse goals.	Olson (2007)	Building a Culture (para. 2)
	Conduct and information inventory	Rudner & Boston (2003)	Prerequisites

Table 4 – Data collection spreadsheet for concept 4

Concept #4	Concept Description	Article Author	Location in Article
Information about selecting a common format for information conversion.	Convert the information in stand-alone systems to a format that can be queried by OLAP (online analytical processing).	Pare & Elovitz (2005)	8 Basic steps to Establishing a Data Warehouse (para. 1)
	Data needs to be put in a common language for the data warehouse. Districts should consider using a SIF (Schools Interoperability Framework) compliant application.	Levine (2002)	The SIF Approach (para. 3)
	Find a common format that existing disparate databases can be translated into.	McIntire (2002)	Introduction (para. 4)
	Standardize the management of data.	Rudner & Boston (2003)	Prerequisites
	Data needs to be put in a compatible and nonredundant format.	Salpeter (2004)	Clean and Accurate Data (para. 1)

Table 5 – Data collection spreadsheet for concept 5

Concept #5	Concept Description	Article Author	Location in Article
Information about cleansing the data.	Cleanse the data. This is perhaps the most time-consuming part of the process. It is amazing how much bad information can exist in a database.	Pare & Elovitz (2005)	8 Basic steps to Establishing a Data Warehouse (para. 1)
	District partnered with Executive Intelligence to clean up district and student-test data.	Borja (2006a)	Introduction (para. 4)
	Cleaning the district's data is critical and time consuming	Borja (2006b)	Actionable Information (para. 5)
	Districts need to plan for a system that can clean up data from various databases.	Borja (2006c)	Benchmark-Test Market Blossoms (para. 6)
	Planning should include make decisions about data cleansing and refresh frequency	Brown (2006)	Planning is Key (para. 1)
	Ensure data cleanliness to prevent problems with false reporting and inaccurate results.	Levine (2002)	The SIF Approach (para. 2)
	Look for comprehensive data cleansing tools.	McIntire (2004)	Buyers Checklist (para. 1)
	Cleanse the data. Ensure the data is of high quality and complete	Rudner & Boston (2003)	Prerequisites

	Technical challenges that need to be addressed include cleaning up data from multiple sources.	Salpeter (2004)	Clean and Accurate Data (para. 1)

Table 6 – Data collection spreadsheet for concept 6

Concept #6	Concept Description	Article Author	Location in Article
Information about importing data into the warehouse.	Bring the data into the warehouse. This is a very important step because it determines what links will exist between data fields.	Pare & Elovitz (2005)	8 Basic steps to Establishing a Data Warehouse (para. 1)
	Disparate information needs to be extracted from individual databases and imported into the data warehouse on a scheduled basis.	Levine (2002)	What is Data Warehousing? (para. 5)
	Look for integration tools for key systems (SIS, informal databases, etc).	McIntire (2004)	Buyers Checklist (para. 1)

Table 7 – Data collection spreadsheet for concept 7

Concept #7	Concept Description	Article Author	Location in Article
Information about determining access to the data.	Determine who will have access to the data. Much of the data is confidential and cannot be made available universally to all members of the school community.	Pare & Elovitz (2005)	8 Basic steps to Establishing a Data Warehouse (para. 1)
	The district selected three levels of data-system access. By administrator, school principal or teacher.	Borja (2006a)	Reading Off the Same Page (para. 5 & 6)
	Know you audience and determine who will have access to the data warehouse.	Brown (2006)	A Smooth Transition (para. 4)
	Decide who gets access to the data warehouse including parents and community members.	McCright (2000)	New Bundles help organize, disseminate data (para. 16)
	Now that the disparate databases are consolidated, determine who should access.	McIntire (2002)	Introduction (para. 5)

Table 8 – Data collection spreadsheet for concept 8

Concept #8	Concept Description	Article Author	Location in Article
Information about determining the toolset (software) for data analysis.	Determine the toolset (software) that you will use for inquiries and analysis of this data. Ease of use and standardization are the keys to proper selection.	Pare & Elovitz (2005)	8 Basic steps to Establishing a Data Warehouse (para. 1)
	Using data effectively requires data tools - particularly student information systems, data warehouse and instructional management systems.	Bernhardt (2005)	Which Tools Do Schools Need?
	District teamed with Executive Intelligence to provide an easy-to-use web interface to the district data warehouse.	Borja (2006a)	Introduction (para. 4)
	Districts need to look for a system that offers some analytical and presentation tools.	Borja (2006c)	Benchmark-Test Market Blossoms (para. 6)
	Planning must consider analytical issues including what tools to deploy	Brown (2006)	Planning is Key (para. 1)
	Determine key software for delivering content and creating reports.	Flaherty (2004)	Timely Test Results (para. 3)

	Teachers need a tool that will help them individualize instruction for students based on collected data.	Hoff (2006)	20-Year Effort (para. 5)
	Administrators can run reports from traditional reporting software or using built in report generators.	Levine (2002)	What is Data Warehousing? (para. 6)
	Select appropriate software for data analysis that allows teachers to create their own queries.	McCright (2000)	New Bundles help organize, disseminate data (para. 14)
	Analyze the data	McIntire (2002)	Data-driven Decision Making, A Step-by-Step Approach (para. 1)
	Look for easy-to-use online analytical processing tools and ad hoc query reporting tools	McIntire (2004)	Buyers Checklist (para. 1)
	Schools need to give explicit data-analysis protocols and reports to administrators and teachers.	Olson (2007)	Building a Culture (para. 7)
	Analyze the data	Rudner & Boston (2003)	Prerequisites

Appendix C – Tables 9-16 Pattern Analysis in Identified Concepts:

A Data Warehouse Planning Guide for Public School IT Directors

Table 9 – Theme examination spreadsheet for concept 1

Planning Step #1	Topics	Key Questions	Problem Areas
Information about selecting software, hardware and people to convert data into a useful format.	Staff training	Is the current staff trained on data warehouse support?	Staff development is an ongoing need, not a one-time solution.
	Infrastructure needs	Is the server and network capacity sufficient to support the data warehouse?	Stress testing is necessary to avoid infrastructure problems.
	Determine software	Would it be best to outsource the software creation?	Choose carefully or risk purchasing software that is inappropriate for a school or district.

Table 10 – Theme examination spreadsheet for concept 2

Planning Step #2	Topics	Key Questions	Problem Areas
Information about selecting the most useful data sources.	Select appropriate data sources	Is the data present and accurate?	Data input can vary by user thus cause problems in historic databases.
	Find data sources	Where and in what media is the data currently stored?	Schools store information in computer databases, filing cabinets, and three-ring binders in various locations.

Table 11 – Theme examination spreadsheet for concept 3

Planning Step #3	Topics	Key Questions	Problem Areas
Information about determining the most important data.	Determine what data is important	What data is critical?	There is potentially too much data that can be collected and the result can be an overwhelmed system or end user.
	Time involved in data selection	Is the end result (the reports) worth the time needed to select the data?	Compiling critical data is time consuming.
	New fields	What data is not currently collected that should be?	Some data needs are currently unknown.

Table 12 – Theme examination spreadsheet for concept 4

Planning Step #4	Topics	Key Questions	Problem Areas
Information about selecting a common format for information conversion.	Data needs to be in a common language/format for the data warehouse	Should schools use products that are SIF (Schools Interoperability Framework) compliant?	Not all products that claim SIF compliance are.

Table 13 – Theme examination spreadsheet for concept 5

Planning Step #5	Topics	Key Questions	Problem Areas
Information about cleansing the data.	Cleansing data	How do you prevent problems with false reporting and inaccurate results through data cleansing?	Inaccurate data can cause problems and false reports.
	Cleansing data from multiple sources	What level of cleaning is needed for each individual data source?	Data cleansing is very time consuming.

Table 14 – Theme examination spreadsheet for concept 6

Planning Step #6	**Topics**	**Key Questions**	**Problem Areas**
Information about importing data into the warehouse.	Imports need to be scheduled	How frequently does the data warehouse need to be updated?	If the data warehouse does not have current data the effectiveness of decisions made using it is compromised.
	Verify key systems have integration tools	Do the SIS and informal systems allow data to be easily imported into the data warehouse?	Lack of common integration tools can limit the value of a data warehouse product for a school.

Table 15 – Theme examination spreadsheet for concept 7

Planning Step #7	**Topics**	**Key Questions**	**Problem Areas**
Information about determining access to the data.	Determine access needs	Who will have access? Do community members and parents need access?	
	Determine level of access	How many levels of access are needed?	The access levels need to match the intended audience.

Table 16 – Theme examination spreadsheet for concept 8

Planning Step #8	**Topics**	**Key Questions**	**Problem Areas**
Information about determining the toolset (software) for data analysis.	Need for effective data analysis tools	How will the data be used?	Planning is important to getting the right toolset implemented.
	Need for data presentation tools	How should the information be displayed?	
	Tool must be easy enough for teachers to run their own queries	What training is needed to allow teachers to create their own student analysis processes?	Some tools are difficult to pull information from.

Definition of Terms

"Analytical Data is consolidated or summarized data used by decision-makers to develop and implement new strategies. Examples of analytical data in schools include average daily student attendance, a longitudinal sequence of test scores, and the distribution of disciplinary incidents by hallway in a school building. Analytical data is based on information gleaned from transactional system data" (McIntire, 2002, A glossary of terms).

A **Dashboard** serves as a user interface that both presents information and enables a user to access or compile new data by means of a series of 'gauges and dial'" (National Forum on Education Statistics, 2000, User dashboard, para. 1).

"**Data Analysis Tools** includes a wide array of technology-based tools for statistical analysis, forecasting, graphing, and highlighting trends. In addition to special-purpose tools, spreadsheets and other common applications already in use in the district can play an important role in data analysis" (Salpeter, 2004, Elements of a data-driven system).

Data Collection is **"**the mechanism for collecting and maintaining high quality education data through consistent application of standard terminology and business rules throughout an organization" (National Forum on Education Statistics, 2000, Data Collection, para. 2).

Data Cubes are "vendor built databases optimized by using a limited number of commonly used data elements" (McIntire, 2005, Does the data model fit your needs?, para. 2).

Data-driven Decision-making is "collecting data, refining the data into a usable format, and basing decisions on the information" (Pare & Elovitz, 2005, Introduction section, para. 6).

Data Reporting is the "presentation of information in multiple formats (as a blend of text, tables, and graphics) and in multiple dimensions (changing an axis to present information more clearly) to clarify the meaning of the data" (National Forum on Education Statistics, 2000, Reporting tools, para. 1).

"Data Warehouse is a consolidated version of data from multiple transactional databases. It contains no unique information of its own. Instead, a data warehouse pulls information from various systems, cleanses the data to eliminate incompatibilities, and then combines it in a uniform format. Because a data warehouse is the single information source for multiple databases, it's the most comprehensive, complex foundation for data analysis and reporting" (McIntire, 2002, A glossary of terms).

"**Decision Support Systems** are those tools and technologies that help administrators make efficient and informed decisions about critical issues such as student and employee performance or financial resource allocation. While one common decision support system is the ordinary spreadsheet, more specialized information management tools can help educators easily make multidimensional queries (asking questions that span several variables, such as grade level, ethnicity, economic status, and test scores) with just a few clicks on a graph" (McIntire, 2002, A glossary of terms).

"**Homegrown Solutions** employ data solutions that use Access, Excel, or other common tools to pull directly from the source databases, skipping the warehouse all together" (Salpeter, 2004, Elements of a data-driven system).

"**Individual Databases** typically include student information systems, human resource records, financial databases, and assessment data from sources such as state tests, benchmark assessments, and instructional management software. In addition, specialized databases with information on individual education plans for special education, disciplinary referrals, professional development and teacher certification, technology support help line calls, community survey results, and library circulation can all play an important role in data-driven decision making" (Salpeter, 2004, Elements of a data-driven system).

An **Information Technology (IT) Director** "plans, organizes and directs the development, installation, implementation, and maintenance of departmental information technology/data; to plan, organize, direct and participate in supporting, training, and assisting departmental users in utilizing a variety of hardware and software applications for data systems" (County of Solano, 2003, Definition, para. 1).

"**Longitudinal Data Systems** provide information about student growth over time that can then be linked with the teachers, programs and schools that have served those students" (Snow-Renner & Torrence, 2002, p. 4).

"**Report Writers** provide the ability to create customized, formatted reports of various sorts is built into most of the data tools, including the data warehouse itself" (Salpeter, 2004, Elements of a data-driven system).

"Student-Information Systems compile and manage demographic data, attendance, and grades, as well as other daily operations. These systems can also do limited data analysis and reporting" (Borja, 2006, System types blend together, para. 3).

"Support Services are supplemental internal resources either consultants or other service providers who can offer help with professional development, needs assessment, data analysis, and system planning. In addition to the companies in this arena, many nonprofit associations and consortia also offer benchmarks, advice, downloadable tools, and other data-related resources" (Salpeter, 2004, Elements of a data-driven system).

“Transactional Data is the day-to-day operational data of an organization that helps it run efficiently. School transactional data, for example, tells us what a student's home address is, what a teacher's highest level of education is, or what a student in sixth grade is averaging in math class. Transactional or operational databases include the student information system, the financial management system, and the library automation system” (McIntire, 2002, A glossary of terms).

Bibliography

About T.H.E. Journal. (2007). T.H.E. Journal. Retrieved April 22, 2007, from http://www.thejournal.com/the/marketing/aboutus/

Apodaca-Tucker, M. T. & Slate, J. R. (2002, April 28). School-Based management: Views from public and private elementary school Principals. Education Policy Analysis Archives, 10(23). Retrieved March 23, 2007, from the World Wide Web: http://epaa.asu.edu/epaa/v10n23.html

Bernhardt, V. L. (2005, February). Data tools for school improvement. Educational Leadership. Retrieved April 17, 2007, from the World Wide Web: ftp://eff.csuchico.edu/Downloads/ArticlesOfNote/DataTools.pdf

Berry, M. J. A. & Linoff, G. S. (2004). Data mining techniques: For marketing, sales, and customer relationship management, second edition. Indianapolis: Wiley.

Borja, R. R. (2006, May 4a). Aware of all students. Education Week. 25. 32.

Borja, R. R. (2006, May 4b). District initiative. Education Week. 25. 24-26, 28, 30-31.

Borja, R. R. (2006, May 4c). Risk and reward. Education Week. 25. 42-44, 46-47.

Brown, J. (2006, October). Too much Information. T.H.E. Journal. Retrieved April 17, 2007, from EbscoHost on-line database.

County of Solano, D. o. H. R. (2003, June 30). "County of Solano, Information Technology Manager." Retrieved April 15, 2007, from http://www.co.solano.ca.us/resources/HumanResources/Information%20Technology%20Manager.pdf.

Dougherty, C. (2001, May 2). More than a snapshot. Education Week. 20. 39, 42.

Flaherty, W. H. (2004, July). Data warehouse helps Hanover County Public Schools raise student achievement. T.H.E. Journal. Retrieved April 17, 2007, from the World Wide Web: http://thejournal.com/articles/16838

Hoff, D. J. (2006, May 4a). Delving into data. Education Week. 25. 12-14, 20-22.

Hoff, D. J. (2006, May 4b). Keeping track. Education Week. 25. 16-18.

Holzberg, C. S. (2006, September 1). Crunching the numbers. Technology & Learning. Retrieved April 17, 2007, from the World Wide Web: http://www.techlearning.com/showArticle.php?articleID=192700195

Honawar, V. (2006, October 1). Data wise. Teacher Magazine. Retrieved March 26, 2007, from the World Wide Web: http://www.teachermagazine.org//tm/articles/2006/10/01/02research.h18.html

Honawar, V. (2007, January 30). Tip of their fingers. Education Week. 25. 38-39.

Kreitl, P. (2005, April). Building our data warehouse. School Administrator. Retrieved April 17, 2007, from the World Wide Web: http://www.aasa.org/publications/content.cfm?ItemNumber=1287

Krippendorff, K. (2004). Content analysis: An introduction to its methodology. Thousand Oaks: Sage.

Levine, E. (2002, November). Building a data warehouse. American School Board Journal. Retrieved March 23, 2007, from the World Wide Web: http://www.asbj.com/2002/11/1102technologyfocus2.html

McCright, J.S. (2000, July 3). Warehouses go to school. eWeek. Retrieved April 17, 2007, from the World Wide Web: http://content.epnet.com/ContentServer.asp?T=P&P=AN&K=3378778&EbscoContent=dGJyMNXb4kSeqLM4v%2BbwOLCmrk%2Bep7ZSs6%2B4Sa%2BWxWXS&ContentCustomer=dGJyMPGuslGwrbVNuePfgeyx%2BEu3q64A&D=aph

McIntire, T. (2002, June 15). The administrator's guide to data-driven decision making. Technology and Learning. Retrieved March 26, 2007, from the World Wide Web: http://www.techlearning.com/db_area/archives/TL/2002/06/guide.php

McIntire, T. (2004, August 16). Eight buying tips: Data warehouses. Technology and Learning. Retrieved March 26, 2007, from the World Wide Web: http://www.techlearning.com/showArticle.php?articleID=26806926

Mercurius, N., Burton, G., Hopkins, B., & Larsen, H. (2004, January). Putting the school interoperability framework to the test. T.H.E. Journal. Retrieved April 17, 2007, from the World Wide Web: http://www.thejournal.com/articles/16570

Mieles, T. & Foley, E. (2005, June 15). School staff and the data warehouse. Technology and Learning. Retrieved March 26, 2007, from the World Wide Web: http://www.techlearning.com/shared/printableArticle.php?articleID=163703320

Montgomery County Public Schools (2006, September 22). Data warehouse. Retrieved April 8, 2007, from the World Wide Web: http://www.mcps.k12.md.us/departments/technology/datawarehouse.shtm

National Forum on Education Statistics. (2000).Building an automated student record system: A step-by-step guide for local and state education agencies (NCES 2000-

324). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

National Forum on Education Statistics. (2006). Forum guide to decision support systems: A Resource for educators (NFES 2006-807). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Olson, L. (2007, February 14). Data-wise school systems seen as sharing key traits. Education Week. Retrieved March 25, 2007, from the World Wide Web: http://www.edweek.org/ew/articles/2007/02/14/23data.h26.html

Olson, L. (2006, June 21). States striving to build data systems across education levels. Education Week. Retrieved March 25, 2007, from the World Wide Web: http://www.edweek.org/ew/articles/2006/06/21/41align.h25.html

Palmquist, M., Busch, C., De Maret, P., Flynn, T., Kellum, R., Le, S., Meyers, B., Saunders, M., White, R. (2005). Content Analysis. Retrieved April 9, 2007, from Colorado State University, Department of English Web site: http://writing.colostate.edu/guides/research/content/.

Pare, R. & Elovitz, L. (2005, April). Data warehousing: An aid to decision-making. T.H.E. Journal. Retrieved March 21, 2007, from the World Wide Web: http://www.thejournal.com/articles/17217/

Prat, N., Akoka, J., Comyn-Wattiau, I. (2006, January 23). A UML-based data warehouse design method. Decision Support Systems. Retrieved April 17, 2007, from the World Wide Web: http://0-www.sciencedirect.com.janus.uoregon.edu/science?_ob=ArticleURL&_udi=B6V8S-4J3NXXK-1&_user=2148430&_coverDate=12%2F31%2F2006&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000056308&_version=1&_urlVersion=0&_userid=2148430&md5=0960b06bbd4a090c4f20df93062e8504

Readership. (2007). Technology and Learning. Retrieved April 22, 2007, from the World Wide Web: http://www.techlearning.com/techlearning/mediakit/Readership2007.pdf

Reeves, P. L. & Burt, W.L. (2006). Challenges in data-based decision-making: Voices from principals. Educational Horizons. Retrieved March 21, 2007, from the World Wide Web: http://www.eric.ed.gov/ERICDocs/data/ericdocs2/content_storage_01/0000000b/80/36/0a/66.pdf

Rudner, L. M. & Boston, C. (2003, February). Data warehousing: Beyond disaggregation. Educational Leadership. Retrieved April 17, 2007, from the World Wide Web:

http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=9029507&loginpage=Login.asp&site=ehost-live

Salpeter, J. (2004, March 15). Data: Mining with a mission. Technology and Learning. Retrieved March 26, 2007, from the World Wide Web: http://www.techlearning.com/shared/printableArticle.php?articleID=18311595

Snow-Renner, R. & Torrence, M. (2002). No child left behind policy brief: State information systems. Education Commission of the States. Retrieved March 21, 2007, from the World Wide Web: http://www.ecs.org/clearinghouse/34/61/3461.pdf

Taylor, D. & Procter, M. (2006). The literature review: A few tips on conducting it. Retrieved April 8, 2007, from the World Wide Web: http://www.utoronto.ca/writing/litrev.html

Trotter, A. (2006, May 4). National effort. Education Week, 25. 48-49. Retrieved March 25, 2007, from the World Wide Web: http://www.edweek.org/ew/articles/2006/05/04/35federal.h25.html

University of Oregon, Library. (2007). Borrowing from UO libraries and beyond. Retrieved April 18, 2007, from the World Wide Web: http://libweb.uoregon.edu/borrowing/

Viadero, D. (2006, July 26). Research center to scour states' data troves. Education Week. 25. 12.

Weiser, C. (2007, February). Dashboard software. Scholastic Administrator. pp. 18, 19.

Wolf, M.A. (2006, July). Inside the states. T.H.E. Journal. Retrieved April 17, 2007, from the World Wide Web: http://www.thejournal.com/articles/18837_4