Illinois State Board of Education

Feasibility Study and Functional Requirements Analysis for Development and Implementation of the ISBE Data Warehouse

Prepared By

MTW Solutions, LLC
3425 Constitution Court, Suite 201
Jefferson City, Missouri 65109

www.mtwsolutions.com
March 6, 2006

This page intentionally left blank

INTRODUCTION

Over the past four months MTW Solutions and ISBE have worked closely to examine the business context, needs, challenges, and processes needed to realize a comprehensive data warehousing solution for Illinois's State Board of Education. This Feasibility Study represents the efforts of a combined team of ISBE technical and program staff whose time was generously shared with our consultants.

MTW would in particular like to thank the following individuals:

Rebecca Watts	**Chief of Staff**
Ginger Reynolds	**User Project Sponsor**
Linda Mitchell	**Technical Project Sponsor**
Connie Wise	**User Project Manager**
Terry Chamberlain	**Data Systems Administrator**
Dennis Powell	**Technical Project Manager**
Warren Summers	**Technical Coordinator**
Candy Taylor	**Technical Coordinator**

As described in this Feasibility Study's findings and recommendations, MTW Solutions believes that ISBE has the resources, infrastructure and drive to create a "best in class" data warehouse solution that will bring far ranging benefits to the department, staff, and ultimately children of Illinois education system.

Revision History

Version	Status	Date	By	Summary of Changes
v.1	Submitted	02/24/06	EL/SM	
v.2.1	Submitted	03/17/06	EL/SM	

Sign-Off

Version	Date	Name	Signature

Copyright Statement

Table of Contents

CHAPTER 1. EXECUTIVE SUMMARY

The Illinois State Board of Education (ISBE) has recognized the necessity to broaden its current data collection and analysis systems to include a data warehouse. To initiate the process of building an integrated data system ISBE contracted for the conduct of a comprehensive feasibility study. Through the Request for Sealed Proposal process, ISBE selected MTW Solutions, LLC and subcontractor eScholar, LLC to author this feasibility study and functional requirements analysis.

Following a kick-off meeting in late November and ISBE's approval of the project governance documents, MTW Solutions began an in depth data collection and analysis phase that included interviews with the IT Steering Committee and the business and technical owners of the major systems targeted for inclusion in the data warehouse, state board member, designated district superintendents, a regional superintendent and two state legislators. The data collection phase also included interviews with representatives from five state education agencies that have implemented data warehouse or related systems to identify industry best practices and gather details about their systems. Simultaneously, the MTW Solutions project team completed additional research to define potential data warehouse system solutions and began to construct a high-level architecture, a data flow diagram, and data model for an ISBE data warehouse.

Through the data collection phase, a vision of an ISBE data warehouse - who would use it, how it should function, what data it should include - began to emerge. (Refer to Chapter 4 *ISBE's Data Warehouse Vision.*) The anticipated users include ISBE staff and board members, District staff, the media, research organizations, legislators, and the general public. Each of the user groups will access the data warehouse for purposes ranging from a parent seeking general information about a school district they hope to enroll their children in, to an ISBE staff member preparing a detailed report for EDEN reporting. To meet the needs of the broadest spectrum of ISBE user needs the data warehouse could function as:

- A strategic decision making tool
- A clearinghouse for data
- A tool for unifying the fragmented data maintained by ISBE
- An instrument for facilitating communication throughout the agency
- A mechanism for reducing support call volume, and
- A tool for generating State and Federal Reports

A data warehouse functions by pulling data from currently existing systems within ISBE, as well as incorporating external data, and housing all data in separate but inter-related databases. After the data are extracted from the source systems and transformed and loaded into the data warehouse, users can access the information by either accessing pre-defined reports, completing queries from the data warehouse or extracting the data to complete more detailed analysis using other software solutions. ISBE has over 200 data systems in place with 100 of these systems

operating online. To successfully build the data warehouse with data that will maximize the operation of the data warehouse, the MTW project team along with ISBE system owners narrowed the number of data collections to 20 systems that will be primary systems for inclusion in the data warehouse. (Profiles of these systems can be found in Chapter 5 *ISBE Systems.*) Each of these source systems functions and stores data differently, and requires an evaluation of how to best extract the data from the system and transform and load it to the data warehouse. This process is illustrated through the use of data flow diagrams, data models and a data warehouse solution architecture diagram. These illustrations can be found in Chapter 9 *Data Model, DFD & Solution Architecture*.

A data warehouse will have significant impact on the operations within ISBE. As an example, ISBE management will need to increase their focus on defining data collection practices and communicating the data quality standards through improved training programs both inside and outside the agency. This improvement in training and related policy development will facilitate the loading of high quality data into the data warehouse and have significant impact on the success and use of the data warehouse. Other risks and benefits impacting the ISBE are addressed in Chapter 6 *The Impact of the Data Warehouse*.

In determining how to best implement a data warehouse ISBE can draw upon a wealth of knowledge and experience gained by other education agencies as they navigated through the tangible and intangible issues of constructing a longitudinal data system. Through general research and interviews with representatives of state education agencies from Nebraska, Georgia, New Jersey, New York and Wyoming, the MTW Solutions project team identified best practices used in education data warehouse projects, as well as identified those events that can derail a project similar to ISBE's data warehouse project. (A summary of best practices and profiles of the five interviewed states can be found in Chapter 7 *Best Practices*.) ISBE also benefited from the expertise and contribution of Shawn Bay, president of eScholar and one of the leading experts in education data warehouses. He is recognized as one of the pioneers of data warehousing and has played a significant role in the development of tools and processes used in data warehousing both in corporate America and throughout the education industry.

Drawing upon all the resources and information gathered through the data collection phase, the MTW Solutions Project Team structured a recommended approach, identified next steps and developed a cost analysis for the ISBE data warehouse initiative. The MTW project team identified the tools required to build and maintain the data warehouse and the criteria that should be used in evaluating vendors' solutions. The overall approach to building a data warehouse includes the recommendation of using a phased implementation for building the data warehouse that will enable ISBE to spread the cost and impact on staffing resources over a period of several years. Before moving forward with the project and clarifying the requirements of the data warehouse, ISBE technical, program and management leaders will need to consider and resolve a large number of issues impacting the data warehouse project. Details of these issues with recommendations for addressing them can be found in Chapter 8 *Recommended Approach* and

Chapter 11 *Next Steps*.

Costs for the implementation of a data warehouse fall into six main areas: hardware, software, development, project management, training and annual maintenance. Software and development will be the two largest cost areas. Cost estimates, discussed in Chapter 10 *Cost Analysis,* outline both a low and high range of pricing in each of these areas and assume all services and products will be purchased by ISBE specific to the data warehouse effort. Should ISBE's ultimate solution include existing hardware, software and the inclusion of ISBE development and managerial staff then these estimates will need to be revised accordingly.

CHAPTER 2. GENERAL INFORMATION

2.1. Purpose

In November 2005, the Illinois State Board of Education made a decision to assess the feasibility of building and implementing a data warehouse and identify the requirements of such a system. MTW Solutions, LLC was contracted to undertake a review of currently available data, identify data gaps, and suggest steps for moving forward. This document is in response to that charge.

2.2. Background

In September 2005, as an initial step in this effort, ISBE issued an RFSP for the conduct and completion of a Feasibility Study and Functional Requirements Analysis for the Development and Implementation of an ISBE Data Warehouse. On October 27th ISBE notified MTW Solutions that they, and their partner eScholar, were the selected vendors for performing this work.

Starting in mid-November with an initial on-site kickoff meeting, MTW began conducting a series of interviews with the IT Steering Committee and the business and technical owners of the 20 major systems targeted for inclusion in the data warehouse, designated district superintendents, and two state legislators. From these interviews, MTW Solutions began documenting both the business context of the data warehouse and potential solution areas.

Based on the business requirements, user group profiles, and existing system infrastructure documented during the discovery process, MTW began an evaluation and analysis of the issues and challenges uniquely facing ISBE in the realization of an enterprise data warehouse solution.

2.3. Scope

In keeping with the requirements of the RFSP for the Feasibility Study, the scope of this study includes the following:

1. Identifying existing ISBE data collection systems that can be discarded or that can be modified or developed to draw data from the data warehouse system thereby reducing the data burden placed on districts.
2. Clarifying ISBE's vision for Data Warehouse by defining the business context and impact on the ISBE organization.
3. Identifying privacy and confidentiality issues related to development and implementation of a data warehouse. Develop standards and procedures that will guarantee the security and confidentiality of the data stored and maintained in the data warehouse system under the provisions of the federal Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 CFR Part 99), and the Illinois School Student Records Act (105 ILCS 20/).

These standards and procedures will include who has access to the system at both the state, regional, and local levels, how authorized users gain access to the system, and how data will be entered or imported into the system.

4. Developing user-group profiles that include reporting needs by data area, requirements and how they will access the system.
5. Identifying and discuss the impact, risk and issues of developing a data warehouse system.
6. Investigating and document requirements for disaster recovery for the data warehouse system.
7. Defining potential data warehouse systems by identifying best practices for data warehouse systems in state education agencies.
8. Defining potential data warehouse system solution areas and provide recommendations in each area. Identify infrastructure, hardware, and software requirements for the data warehouse system. This documentation will include criteria for selecting report creation and distribution software, OLAP tools and data extraction, load and transformation (ETL) tools that can be utilized by personnel at the state, regional, and school district levels.
9. Developing a data model, data flow diagram and solution architecture for a data warehouse.
10. Formulating a cost analysis for the data warehouse system
11. Providing direction for next steps that identify activities and tasks that need to be completed before the data warehouse can be constructed and implemented.

2.4. Targeted Audience and Participants

This Feasibility Study document is targeted to the Data Warehouse Feasibility Study project team and those individuals internal and external to the agency who will play a part in the subsequent review of the data warehouse project and those who will play key roles in the decision to move forward with the design and implementation of the data warehouse. This document assumes a basic understanding of ISBE data collection and system processes. Realizing that the concepts and terms of data warehousing may be unfamiliar to readers, this document includes an introduction to data warehouses and also includes a glossary of terms at the end of the document. Those users of the document with little or no knowledge about data warehouses are encouraged to review the Introduction to Data Warehouses chapter and glossary before reviewing the rest of the document.

The following groups and individuals generously committed their time and energy in assisting MTW Solutions in the discovery, analysis and management of the Feasibility Study:

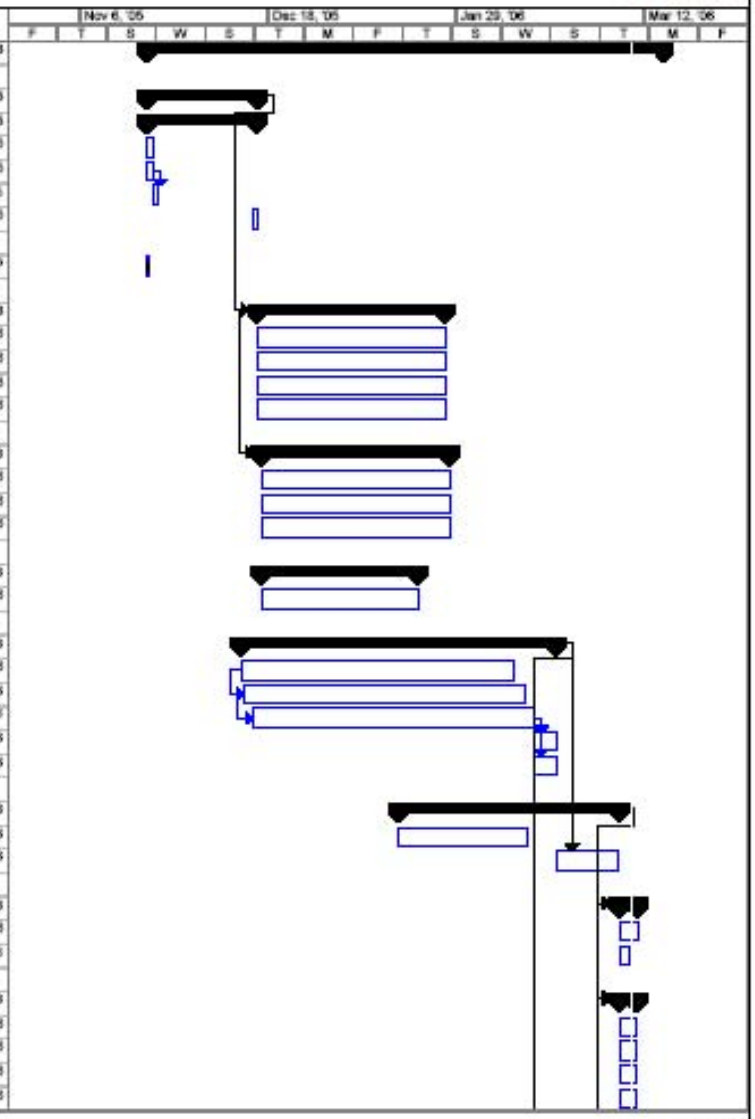
Participants	Area
Donna Luallen, Henri Fonville, Eric Thatcher, Rich Loman, Rich Dehart, Sheryl Bradley	Accountability
Steve Rothenberg, Rhonda Robinson, Pat Brennan	Administrative Services
Lilibeth Gumia	Bilingual, Regional Safe Schools
Andrea Brown	Board Member
Mike Kelly	Carlinville Public Schools
Dennis Williams, Lynn Rhoades, Brenda Stonecipher, Sharon Battles, Marti Woelfle	Certification and Professional Preparation
Jeff Aranowski	Chicago ROE Services
Pat Folland, Ann Horton	Chicago Special Education Compliance Monitoring
Warren Summers	Data Systems
Robin Lisboa, Beth Robinson, Feng Naolho	English Language Learning
Robert Wolfe	External Assurance
Myron Mason	Federal Grants and Programs
Gail Steinhour	Federal proposals and reporting
Linda Mitchell	Financial Services
Melissa Oller	Fiscal Services
Tim Imler	Funding and Disbursement Services
Jonathan Furr	General Counsel
Deborah Scheiter	Internal Audit
Dennis Powell	ISBE SIS
Chris Schmitt	Nutrition Programs and Support Services
Ginger Reynolds	Project Sponsor
Dr. Kuzneweski	Rockford Public Schools
Deborah Vespa	School Business and Support Services
Beth Hanselman	Special Education Services – Springfield
Becky McCabe	Student Assessment
Scott Norton	Technology Support

2.5. Method of Study

Working with the IT Steering Committee, MTW Solutions worked to establish project governance documents and standards. These governance documents included a description of major tasks and deliverables, roles and responsibilities, specific methods of communication, and status reporting. A project plan was established detailing a work breakdown structure, milestones and deliverables, roles and responsibilities, and change and issue management. The following project plan was used to guide the Data Warehouse Feasibility Study to completion.

ISBE Data Warehouse Feasibiity Study Schedule - Final

ID	Task Name	Duration	Start	Finish
1	ISBE Data Warehouse Feasibility Study	665 hrs?	Mon 11/21/05	Thu 3/16/06
2				
3	Establish Project Governance	151 hrs	Mon 11/21/05	Thu 12/15/05
4	Develop Project Standards and Procedures	151 hrs	Mon 11/21/05	Thu 12/15/05
5	Identify ISBE Steering Committee & Advisory Committee Members	1.75 days	Mon 11/21/05	Tue 11/22/05
6	Customize Project Management Procedures	1.75 days	Mon 11/21/05	Tue 11/22/05
7	Develop and establish Project Scope and Change Management Procedures	0.88 days	Tue 11/22/05	Wed 11/23/05
8	Obtain ISBE signoff on Project Management Procedures	0.88 days	Thu 12/15/05	Thu 12/15/05
9				
10	Project Kickoff Meeting	2 hrs	Mon 11/21/05	Mon 11/21/05
11				
12	Define Business Context for Data Warehouse	240 hrs	Thu 12/15/05	Thu 1/26/06
13	Review ISBE motivators for a Data Warehouse System	30 days	Thu 12/15/05	Thu 1/26/06
14	Review Organizational Issues / Costs	30 days	Thu 12/15/05	Thu 1/26/06
15	Review End User "wants and needs"	30 days	Thu 12/15/05	Thu 1/26/06
16	Review Organizational Environment	30 days	Thu 12/15/05	Thu 1/26/06
17				
18	Determine Potential Solution Areas	240 hrs	Fri 12/16/05	Fri 1/27/06
19	Determine Capability Enablers	30 days	Fri 12/16/05	Fri 1/27/06
20	Identify Capability Gaps / Deficiencies	30 days	Fri 12/16/05	Fri 1/27/06
21	Identify Solutions Areas	30 days	Fri 12/16/05	Fri 1/27/06
22				
23	User Group Profiles	200 hrs	Fri 12/16/05	Fri 1/20/06
24	Define User Groups	25 days	Fri 12/16/05	Fri 1/20/06
25				
26	Detailed Requirements	402.04 hrs	Mon 12/12/05	Mon 2/20/06
27	Conduct Requirements Gathering Meetings	45 days	Mon 12/12/05	Fri 2/10/06
28	Distribute Minutes from Requirements Gathering Meetings for review	45 days	Mon 12/12/05	Mon 2/13/06
29	Categorize Requirements across meetings	45 days	Wed 12/14/05	Wed 2/15/06
30	Create list of Functional Requirements by Category	2.63 days	Wed 2/15/06	Mon 2/20/06
31	Create list of Non-Functional Requirements by Category	2.63 days	Wed 2/15/06	Mon 2/20/06
32				
33	Develop Data Model	282.04 hrs	Mon 1/16/06	Mon 3/6/06
34	Research Best Practices from other State Level Data Warehouses	21 days	Mon 1/16/06	Mon 2/13/06
35	Create Data Model Diagrams	10 days	Mon 2/20/06	Mon 3/6/06
36				
37	Propose Vendor-Neutral Solution Architecture	35 hrs	Mon 3/6/06	Fri 3/10/06
38	Create Architecture Diagrams	4.38 days	Mon 3/6/06	Fri 3/10/06
39	Create sizing estimates	2.63 days	Mon 3/6/06	Wed 3/8/06
40				
41	Identify Physical Requirements for Data Warehouse System	32 hrs	Mon 3/6/06	Fri 3/10/06
42	Document existing ISBE Infrastructure / Hardware Direction & Preferences	4 days	Mon 3/6/06	Fri 3/10/06
43	Describe criteria for selecting Report Generation / Distribution	4 days	Mon 3/6/06	Fri 3/10/06
44	Describe criteria for selecting OLAP tools	4 days	Mon 3/6/06	Fri 3/10/06
45	Describe criteria for selecting ETL tools	4 days	Mon 3/6/06	Fri 3/10/06



During November and December, MTW Solution's analysis team conducted onsite interviews with ISBE project sponsors, IT staff, business users, and representatives of the state, regional and local level users. The information gathered in the user group profile sessions has been analyzed and summarized to describe the identified business context of the data warehouse, the composition of data to be captured, and the types of issues and answers the system is to address.

Concurrent with this process, MTW Solution's technical staff worked with ISBE's IT staff to gain an understanding of ISBE's technical and application infrastructure. These findings have been used to compile this feasibility study's analysis of how existing systems will interface with the data warehouse, as well as where systems and infrastructure will need to be strengthened and enhanced to support a robust data warehousing solution.

CHAPTER 3. INTRODUCTION TO DATA WAREHOUSES

A data warehouse is "a copy of transaction data specifically structured for query and analysis." Ralph Kimball, *The Data Warehouse Toolkit.*

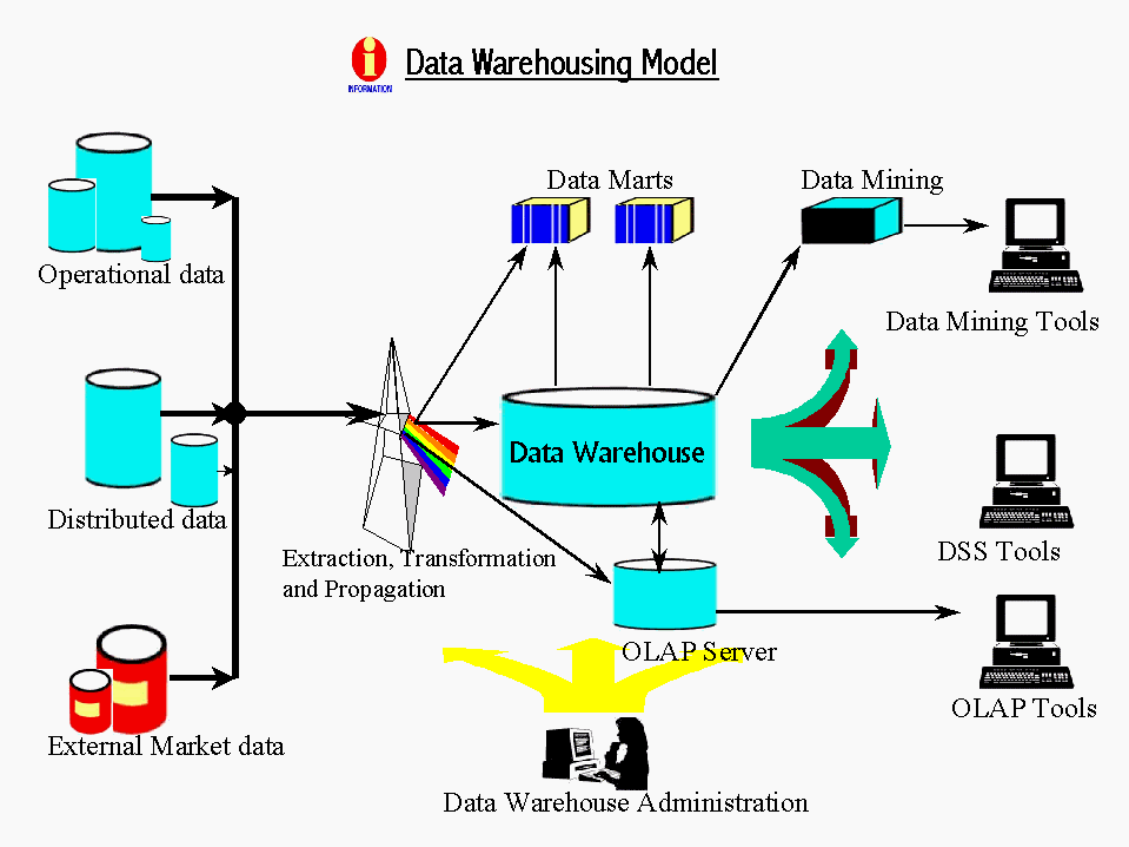
From the earliest development of mainframe systems, the systems rarely stored historical data and the data was often stored in complex formats and was difficult to access. As the development of computer applications expanded, the complexity and differences of how data was collected and stored exploded, often creating silos of information within an organization, with little or no way for the organization to use the collected data from these systems for analysis and/or decision-making. In the late 1980's, the concept of bringing together data in a common environment evolved into the use of the term "data warehouse". (Project Team Consultant Shawn Bay is considered one of the pioneers of data warehousing and has been active in data warehousing development since the mid 1980's while working with Procter & Gamble and Unilever.) According to Bill Inmon in a Byte.com article entitled Wherefore Warehouse, "*Data warehouses squarely address these inadequacies of the operational environment by integrating data, providing historical data, and providing detailed as well as summary data.*"

A data warehouse provides a means to use compiled data as a decision support system or knowledge-based applications architecture. A **decision support system** is a term that refers to an interactive computer application that can be used to gather and present data from a variety of sources. The results of the gathered information can be used for decision-making purposes. **Knowledge-based applications architecture** is process of using two components, a knowledge base and inference engine to solve problems or make decisions by using knowledge and analytical rules defined by experts in a particular field.

The data or pieces of information in the data warehouse can come from disparate systems including mainframe legacy systems, Web-based operational or transactional systems or even a simple Microsoft Excel Spreadsheet. The data warehouse ISBE is considering will bring together a variety of information including student, teacher, staff, financial and performance data. Currently, ISBE has over 200 systems in place with 100 of these systems operating online.

Bringing together data from various sources requires considerable planning. At a minimum, it is necessary to identify the types of data that need to be collected, and define the meaning of the data elements, so that there is consistency throughout the organization of how the organization interprets particular types of data. Also factoring into the identification of types of data to include in the data warehouse is how the organization will use the data warehouse as a decision support system. This includes defining the type of reports and queries that the organization needs to obtain from a data warehouse.

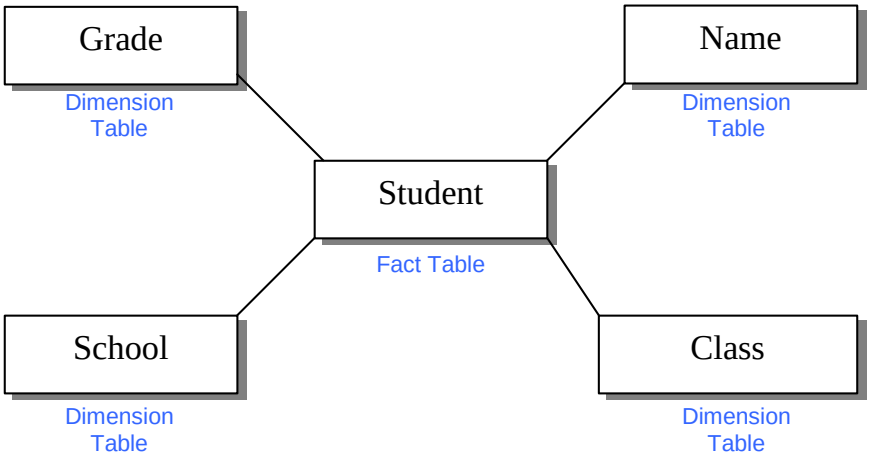
One of the earliest steps in the development of a data warehouse is to define the data warehouse architecture. This architecture defines the elements and services of the data warehouse. For example, it will illustrate how the data will come together, how it will be transformed and how and where it will be stored. Below is an example of a data warehouse architecture diagram.[1]



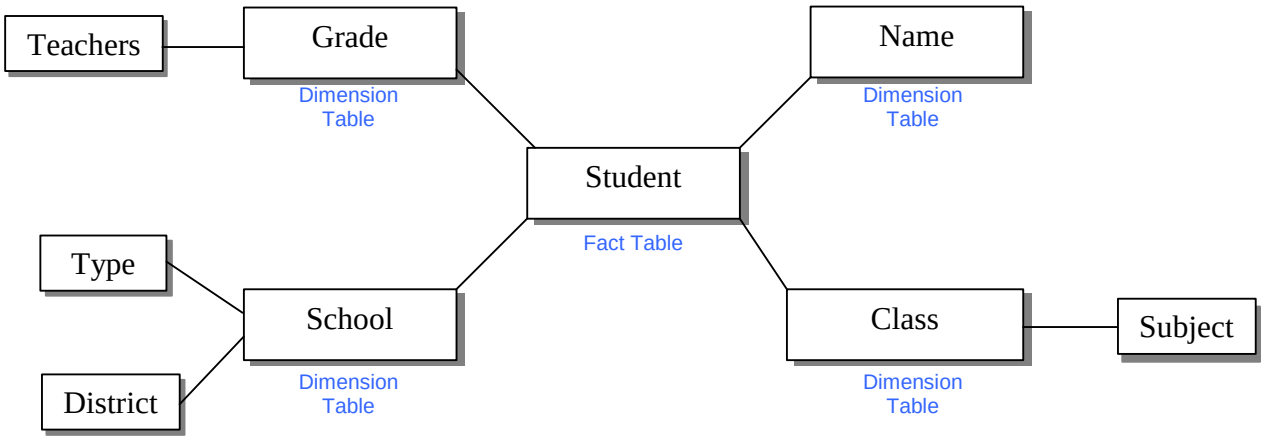
The architecture of the system can vary. A construct known as the **operational data store** can be used to create an additional source of information that is separate from the data warehouse. Data passes through the Extract, Transform, Load (ETL) process into the data store and then passes an additional ETL process before being loaded into the data warehouse. An operational data store is a "subject-oriented, integrated volatile, current-valued, detailed-only collection of data in support of an organization's need for up-to-the-second operational, integrated, collective information."[2] Because the operational data store is loaded with data more frequently than the data warehouse, it can provide more current reporting information than the data warehouse. For example, the data store may be updated daily with data from the source systems, whereas the

[1] Intellibusiness.com: "Data Warehousing – For Better Business Decisions, Anjaneyulu Marempudi.
[2] "The Operational Data Store", Bill Inmon. InfoDB, 1995.

data warehouse may updated weekly or monthly. The data store and the data warehouse are designed to store multi-dimensional data through a system of related tables. The related tables are known as a **Relational Database Management System (RDBMS).** To understand the system of related tables Database Architects use a graphic depiction to illustrate how the data are grouped. Creating a Dimensional Design or **Star Schema** provides the means of structuring data based on a set of known dimensions. It allows the data warehouse to store multi-dimensional data in a relational database management system. The Star schema includes a central fact table surrounded by a group of dimension tables. The **fact table** illustrates the primary information in the data warehouse and the smaller **dimension** or **lookup tables** contain data element details about a particular data element listed in the fact table. **Star Schemas** provide easy to understand structures for the data warehouse. The information within these groupings are referred to as Stars because diagrams of these groupings resemble star patterns. Below is an example of a Star Schema:



An extension of the star schema is the **Snowflake Schema**, which is the means of applying additional dimensions to the dimensions of a star schema in a relational environment. Below is an example of a Snowflake Schema:



Because of the differences between systems and how systems define and store data, it is necessary to process the data so that the data from multiple source systems can come together in a new environment. This process is called **ETL,** an abbreviation for **E**xtract, **T**ransform and **L**oad.

- **Extract** is the process of reading the data from the source database. This can be done using batch, historic or real-time processes. Prior to the extraction process it may be necessary to clean the data. **Data cleansing** is the process of removing errors and inconsistencies with the source data.
- **Transform** is the process of converting the extracted data from its previous form into the form it needs to be in so that it can be placed into another database. Transformation occurs by using rules or lookup tables or by combining the data with other data.
- **Load** is the process of writing the data into the target database.

One approach to the architecture of a data warehouse is to use data marts to organize the data. A single **Data Mart** is data structure that is optimized for access. This approach supports the notion of simplicity by allowing the user to focus on one or just a few subject areas at a time. It is very important, that while designing data marts, that mart-spanning relationships be maintained globally. For example, Students relate to the facts contained in many marts, including Attendance, Assessment Results and more. Using the ISBE data warehouse as example, there may be a data mart for student information, one for staff/teacher information, etc. Using the concepts of data marts and star and snowflake schemas allows the data warehouse to be scalable and also provides a simple tool for visualizing the location of the data. Data warehouse solutions take differing approaches to handling data marts. Some use data marts as a transitional step in migrating data to the data warehouse; others create data marts from the data warehouse to provide a reporting data source to users not requiring data falling within the greater scope of the data warehouse; and other solutions either do not employ data marts or create limited views against the data warehouse to provide “virtual” data mart capabilities.

After the data warehouse has data loaded, it is possible to begin to utilize the data for analysis. The **multi-dimensional analysis** process is completed using a variety of software products and provides the ability to manipulate the data by a variety of categories. The multi-dimensional analysis process is also referred to as “slicing and dicing” or “drilling-down”. Products that allow users to complete multi-dimensional analysis are referred to as **OLAP tools.** OLAP is an acronym for On-Line Analytical Processing. OLAP tools can incorporate data acquisition, data access, and/or data manipulation. These tools can produce pre-defined or canned reports or software to provide users the ability to execute queries directly against the data warehouse to produce any type of report. Finally, end users can access OLAP Services by directly using the OLAP tools or they may access pre-defined reports via a Web browser.

CHAPTER 4. ISBE'S DATA WAREHOUSE VISION

4.1. Overview

Over the past five years, ISBE has seen a growing demand to provide analysis across multiple years and multiple education functions within the agency. While ISBE has extensive system resources with as many as 200 separate systems in house, there are limitations to generating reports, providing analysis and making data supported decisions within the agency. This shortfall within business operations resulted in identifying the need to "*Create a comprehensive data warehouse to foster and support more highly informed decision-making by all ISBE constituents*." – from ISBE's 2005 Comprehensive Strategic Plan for Elementary and Secondary Education.

Realizing the complexity and financial investment of building a data warehouse system, ISBE decided that it was necessary to conduct a feasibility study and define business requirements before proceeding with the project. The feasibility study and supporting business requirements gathering provided further direction to ISBE's vision for a data warehouse system. This section provides a closer examination of the key issues.

The project team identified a number of key questions that need to be answered when considering the implementation of a data warehouse.

1. Who will be the users of the data warehouse?
2. What are the business requirements of the system?
3. What types of data will be stored in the data warehouse and how will the data warehouse be used?
4. What is the impact of creating a data warehouse and what are the corresponding risks and issues to consider?

These questions are inter-related and the answers will help ISBE to further define their vision for a data warehouse. This chapter and Chapter 6 will address these key areas.

4.2. User Group Profiles

Who will be the users of the data warehouse?

Based on onsite analysis, the following populations have been identified as potential users for the data warehouse:

- ISBE
- Districts and Regions
- The Public (Parents, as well as the general public)
- Legislators
- Media
- Research organizations

A measure of success of a data warehouse implementation is the size of its user base. ISBE has a strong desire to have high utilization of the data warehouse by as broad a user base as possible. The following sections describes each of the types of users (*User Group Profile*), how each group will access the data warehouse, the types of information and detail to which they will have access, and the potential uses they have for this information.

The general public and other entities (including the legislature and the media) that do not have access to ISBE's IWAS security portal will have access through ISBE's public web interface to predefined data warehouse reports and queries. Organizations requiring data not available through this public interface (e.g. research organizations desiring student-level data) will need to work with ISBE to develop data sharing agreements within the constraints of federal and state privacy mandates. ISBE may wish to explore the idea of creating a data warehouse registration process where entities desiring access to data warehouse data not available via the public portal may request access online.

ISBE, regions and district users will access the data warehouse through a Web-based interface accessed through ISBE's IWAS portal (which will provide all necessary authentication and authorization security functionality). Depending upon individual security levels, ISBE and district users will have the option to view pre-formatted ("canned") reports, view public, non-confidential (as defined by the Freedom of Information Act) information, run Web-based interactive ("ad hoc") queries, and create data extracts that would support local use of statistical and analytical tools (ranging from Excel to SPSS).

ISBE

The data warehouse will be serving a variety of "constituents" within ISBE including division analysts, compliance specialists, board members, and ISBE committees/subcommittees. Levels

of access to the data warehouse will vary based on determinations of a user's security rights and needs but, as with the districts and regions, Web-based access will be through IWAS. Non-Web based access to the underlying database (primarily for the purpose of conducting complex analysis and reporting activities) will have to be set at a policy level and determined (in many cases) on a case-by-case basis.

Major tasks leveraging the data warehouse will include:

Major Task	Explanation of Tasks
Policy Support	Provide data and analysis results to Board Members, upper management, and legislators to support broad policy making decisions
Analysis	Provide the data and tools necessary to answer "what-if" questions, support trend analysis, support drill-down capability to understand aggregate results, and facilitate the use of sophisticated data mining tools. Additionally, the wider availability of data and analysis tools should reduce the analysis tasks requested of the Data Systems division.
Federal and State Reporting	Support the creation and generation of recurring reports required by the state and federal government. Reports may include the Annual School Report Card, Federal Career and Technical Education, Bilingual, Regional Safe Schools and Truants Alternative & Optional Education reporting., Child Nutrition Services, Federal Special Education, the Non-Fiscal Common Core of Data Report (CCD), the biennial EEO-5 (a report containing race/ethnicity and gender information about certified and non-certified staff), and EDEN.
Compliance Monitoring	ISBE divisions and staff are responsible for a multitude of compliance monitoring activity including special education (including LRE issues), building/facility safety, audits, NCLB program monitoring. Use of the data warehouse will assist in more effectively scheduling compliance-monitoring activities with onsite district activities covering a greater range of areas in a single visit, as well as highlighting districts requiring attention against a greater range of compliance issues.

Major Task	Explanation of Tasks
Technical Support	The comprehensive "at your fingers" nature of the data warehouse interface(s) means that individuals responsible for providing compliance and technical support to districts will be able to do so much more thoroughly and quickly. Additionally, the analysis functions available to this staff should allow them to proactively identify areas needing technical assistance.
Information Request Fulfillment	Divisions within ISBE are regularly asked to provide information to a variety of requestors including legislators, the media, parents and the general public. Information should be more easily accessible to those receiving the requests. Additionally, those receiving the requests may be able to direct the requestors directly to the data warehouse Website in order to access the data themselves.

Districts/Regions

Based on interviews with Project Sponsors and ISBE Board members, management believes that the data warehouse should be constructed to be of maximum use to districts and regions. In combination with ISBE's Student Information System (SIS), formulation of the data warehouse should reduce the amount of data entry required by the districts. By eliminating duplicate requests for data from the districts and by no longer requesting information from the districts that ISBE is already gathering through other mechanisms, the districts should see significant efficiencies in its data collection and reporting activities.

The district profile function of the data warehouse will assist districts in evaluating themselves from a "big picture" perspective. At a glance, the district and regional superintendents should be able to discern the overall health of a district and pinpoint areas that are both above and below average (e.g. program participation, funding, student performance, teacher certification and compliance issues). Because districts are the original providers of district data, each district will have access to the lowest level of detail contained within the data warehouse for their district. This level of access will provide the districts with powerful drill-down capabilities and the ability to conduct in depth analysis. Details of how districts will have access to the low-level data will be discussed in the Privacy and Confidentiality section of the feasibility study. The data warehouse will also provide districts with the ability to conduct cross-district comparisons (especially against districts of similar composition and size); respond to district-level requests for information by the media, legislature and other organizations; and provide assistance to parents requesting

additional information for specialized services available within the district.

Parents and General Public

Determining the business needs and requirements of a data warehouse for the general public, given the breadth and depth of the population, is not possible within the scope of this feasibility study. However, based on the most common types of inquiries ISBE receives, the data warehouse can be designed to address many widespread business requirements. Much like the districts and regions, the public has an interest in understanding how specific schools and districts are performing, how funding is being distributed, and the types of services available to students with special needs.

The general public is already accessing a variety of information from ISBE via Web interfaces (e.g. the Interactive Report Card, FRIS Inquiry, ILEAR (Annual Financial Report information). Providing a public "face" to the data warehouse, as well as publicizing its availability, will allow more inquiries to be funneled to a common and consistent source of data. Parents will be able to more easily access information that directly bears on their children's education including school and district performance, services available within their school district, funding and expenditure information, teacher qualifications and experience, and special and technical education available within their school or district.

Two areas that need to be addressed when providing a public interface to the data warehouse are the varying levels of user computer/Web proficiency and the ability of individuals to understand the information they access. Screens must be simple and intuitive. Complex querying facilities will serve more to confuse most users than to empower them. Online help must not only be available to guide users through the screens they are using, but must also act as a dictionary for explaining how data are derived and what it means. Definitions for calculated and aggregated data available to the public must be clearly visible.

State-level Legislators

As state education agencies are increasingly being held accountable for the results of state supported programs and are equally held responsible for the management of state appropriations, legislators are frequently seeking more quantifiable and qualitative information to evaluate success of existing programs, as well as seeking the opportunity to use data as a decision-making support tool for new programs. The project team met with two Illinois state legislators, State Senator David Luechtefeld and Representative Roger Eddy, to solicit their thoughts and identify their needs for a decision support tool like the ISBE data warehouse. Both said that the reports and queries that could be generated from the data warehouse would benefit legislators in decision-making processes and provide a means to do more effective planning. Legislators and their staff could access the data warehouse by the same Web-based access used by the general public user groups. If they require data that is not available through public access, they could

submit a request directly to the State Board of Education. Analysts within the agency would have a much quicker response time by accessing the data warehouse. Both Senator Luechtefeld and Representative Eddy expressed concern of student-level data being available and stressed the importance of maintaining confidentiality and meeting the requirements of FERPA and other laws that regulate privacy and confidentiality. They were highly supportive of aggregated data being available to the general public, media and districts.

Media

The media is a major consumer of education data and generates a large percentage of the data requests ISBE receives. Much like the public, varying degrees of computer proficiency and understanding of education related data are major factors when presenting data to the media.

Requests from the media typically fall in several categories:

- Student assessment data
- Funding/spending
- Program participation
- Information about specific student populations (e.g. race, special education, gender, income)
- Educator certification/qualification data
- Educator mailing labels by position and subject

The data warehouse will provide a unique opportunity for ISBE to present data to the media in a form and content best suited for their needs. This can be accomplished via several methods:

- Report data warehouse information in pre-formatted ("canned") reports that map to the most common types of media requests.
- Provide simple, intuitive querying capabilities.
- Clearly describe how aggregations and calculations are derived.
- Embed definitions for education terms directly within the interfaces of the data warehouse (e.g. minimize the use of acronyms). This will help define terms that may have multiple or ambiguous meanings (e.g. poverty or low performing).
- Ensure support is available to readily assist media organizations.

Research organizations and institutions

Research organizations and institutions represent a different user group than those discussed in the above sections. Some research organizations and institutions will find needed data results from the public access tools the data warehouse will provide. However, by the nature of the detailed data requirements needed for research projects, most researchers will be seeking access to data that may be restricted from general public user groups. Historically, access to student-level data and other protected levels of data for research purposes has been approved on a case-by-case basis. Similarly, access to protected data for research purposes via the data warehouse will have to be decided on a case-by-case basis. States are working on developing written policies and procedures governing the access to data for research purposes. These policies are being crafted to include data sharing agreements between the state education agency and the research facility detailing the use and publication of the protected data.

4.3. Business Requirements of the Data Warehouse

What are the business requirements of the system?

At its essence, the data warehouse serves as a consolidated repository for information gathered from different sources and presented to users in a consistent and integrated structure. While each user group will have differing (and overlapping) business needs, an ISBE data warehouse will need to function and meet the following business requirements:

- Be a strategic decision making tool
- Be a clearinghouse for data
- Be a tool for unifying the fragmented nature of data maintained by ISBE
- Be an instrument for facilitating communication throughout the agency
- Be a mechanism for reducing support call volume
- Be a tool for generating State and Federal Reports

The following sections describe each of these business requirements.

Strategic Decision Making Tool

A strategic decision making tool assists users in arriving at more informed decisions by enabling them to ask “big picture questions”. The data warehouse as a strategic decision making tool is one of the primary reasons for building a data warehouse. Having the ability to generate reports and query the database will have substantial positive impact on districts, regions, ISBE and the state.

The success of a data warehouse as a strategic decision making tool is determined by three factors:

- The breadth, timeliness and quality of data contained within the data warehouse,
- The availability and usability (access) to the data warehouse by decision makers, and
- The power and scope of the data warehouse's reporting and querying capabilities.*

The quality of the data entering the data warehouse will have significant impact on the success of the data warehouse. A focus on quality of data needs to extend from the districts through the agency. Often times it is necessary to shift and change current approaches to quality of data.

The *Forum Guide to Building a Culture of Quality Data: A School & District Resource*[3] discusses in detail recommendations to assist education organizations in developing this "Culture of Quality Data." Below is an excerpt from the guide that specifically addresses the quality of data:

"The foundation for any decision support system is the data. Quality data are achievable in a school or district through the collaborative efforts of all staff. Administrators responsible for developing a DSS will want to know the procedures that ensure the quality of the data in each of the computer systems, or underlying databases, that will be part of the DSS. Data flowing into a DSS can come from a variety of sources and from a variety of computer systems. When embarking on the path leading to a DSS, the accuracy of the data in each of these systems is one of the first questions considered.

It is necessary for the organization to have appropriate procedures in place to enter data into a computer, transfer data between computers, and to review those data to ensure accuracy. Additionally, it is essential to review the validity of the reports generated. These procedures are critical for the success of any technological solution. It makes sense to delay the development of an integrated DSS while the organization develops and implements effective processes for managing data.

Accurate generation, storage, analysis, and communication of data, involve a constant effort on the part of schools, districts, and state departments of education. Building an environment in which a culture of quality data flourishes requires identifying all the stakeholders, bringing them together through processes that include defining, revising, and communicating policies that relate to the collection, maintenance, cleansing, and reporting of data. It creates this environment by emphasizing that quality data are the responsibility of everyone, not just the office staff."

Secondly, the availability and usability of the data warehouse will have direct impact on the success of the data warehouse as a decision-making tool. The vision of the data warehouse will allow a broad network of user's access to information that was previously either unavailable or difficult to access. Components necessary to meet the availability and usability requirements of the data warehouse must include attention to developing easily accessible and user-friendly mechanisms to access the data warehouse. As mentioned above, users will access the data warehouse via their Web browsers. As the development of the data warehouse moves forward, it will be crucial for the project team to give considerable attention to the development and deployment of the tools to access the data warehouse.

[3] The National Forum on Education Statistics (2005). *Forum Guide to Building a Culture of Quality Data: A School and District Resource: 2004* (NCES 2005-801). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

The capacity of a data warehouse to store snapshots of information at given points in time (referred to as longitudinal data) allows comparisons and analysis to not only be done across districts and regions, but across time. Currently, this ability is limited to individual systems at best. The realization of the data warehouse will allow longitudinal comparisons and analysis across all the data loaded into the data warehouse.

A Clearinghouse for Data

Implementation of a data warehouse forces decisions to be made about the sources and accuracy of data. Where data may be collected multiple times, by multiple systems, determining how this data are represented within the data warehouse will mean the agency will make determinations about what version of this data represents “the truth”.

Further, consolidating statewide data in one centralized database widens the availability and awareness of the state’s information. Constituents of the data warehouse use this data as a clearinghouse for information, breaking down many of the impediments to the free flow of information between divisions, agencies and the public.

Defragmentation of Data

Data currently collected by the agency is highly fragmented across 200 systems and collection vehicles. Despite the fact that many of these systems are integrated, retrieval and analysis of this information can be an extremely time consuming and staff intensive effort. State and federal reporting tasks are typically intensive efforts that absorb a great deal of staff effort to complete. The agency is not always able to respond in a timely fashion to information requests from districts, the legislature, the media and the public at large.

The data warehouse will serve as a defragmentation vehicle for consolidating data and repackaging it in a seamless, integrated repository. State and federal reporting tasks should become significantly easier and more accurate. Staff time should become more available to spend on tasks that more directly benefit students. Data analysis will also become substantially easier and more comprehensive. Timeliness in meeting informational requests should also improve dramatically.

A Communication Tool – Breaking the Silo Culture

A silo mentality is indicative of an environment where individuals and agencies are narrowly focused on immediate responsibilities and concerns. There is no “big picture” understanding of the organizations larger aims or status. ISBE staff and management both agree that the organization suffers from a silo culture.

By melding data from multiple and disparate divisions, the data warehouse will help to open channels of communication. A successful data warehouse must necessitate a cultural shift away from the “it’s my data” mentality to an understanding of data as a communal tool. Tasks and

resources may be better utilized by capitalizing on individual and group efforts.

For example, use of the data warehouse for trend analysis and the creation of district profiles will better leverage agency resources. A district profile is a comprehensive picture of a district at a given point in time and will contain district data pertaining to:

- Students
- Teachers
- Program participation (District Program Budget and expenditures)
- Financial information (by district and school)
- Facilities (buildings and equipment)
- Technical education
- Special education
- Compliance information
- Migrant information
- Limited English proficiency
- Child Nutrition Services

The development of the Data Warehouse will also drive ISBE to define data collection practices and policies across the agency and between ISBE and the districts and regions. Defining data collection practices and communicating the standards through improved training programs will foster increased communication both inside and outside the agency. Likewise, the increased attention to training and policy development will result in loading quality data into the data warehouse. As will be discussed in the Approach section of this feasibility study, substantial work must be done in this area to move the data warehouse from vision to reality.

Call Support Reduction

A common theme through out ISBE is the need to provide support to requestors of agency data. Information requests typically come from the general public, parents, taxpayers, teachers, legislators, reporters/media, and research organizations and are received by the districts, regions, and agency divisions. By using the data warehouse as the engine to power an online interface to support the display and querying of state education data, the agency should see a significant decrease in these types of support calls. Additionally, requests coming into the agency should require less staff time to respond to either because 1) the requestor can be directed to the Website to retrieve the information themselves, or 2) the consolidated nature of the data warehouse's information decreases the time necessary for staff to provide the data.

State and Federal Reporting

Staff in divisions ranging from Special Education, English Language Learning, Career

Development and Preparation, Nutrition Systems, and Data Analysis and Progress Reporting spends a considerable portion of its time compiling information to meet state and federal reporting requirements. In many cases, compiling this information requires gathering information from multiple (and disparate) systems across multiple divisions. Information may be stored in Microsoft Access databases and Excel spreadsheets, on a mainframe, in relational databases, and as flat files.

Centralization of this data within a data warehouse, in a form optimized for reporting (the dimensional design described earlier) coupled with canned reports designed to generate recurring state and federal reports will significantly reduce staff time committed to these tasks. Resultantly, reports should be delivered in a more timely fashion and staff will be able to spend more time on analysis and activities that more directly relate to directly serving their respective constituents.

4.4. Data Requirements – Content of the Data Warehouse

What types of data will be stored in the data warehouse and how will it be used?

The content of the data warehouse is in large part determined by the targeted user communities and identified business requirements. However, availability and quality of data within the agency, as well as privacy and confidentiality concerns, also dictate the substance of the data warehouse. The following have been identified as content for the data warehouse:

- Students demographic, assessment, and enrollment information*
- Teachers demographic and certification information
- District and school program participation
- District financial information including budget and expenditures
- District facilities (buildings and equipment)
- Specialized student programs including technical education, special education, limited English proficiency, and migrant information
- District compliance and monitoring
- District child nutrition services including free and reduced breakfast and lunches (public, nonpublic and ungraded entities)

***The state legislator is considering the repercussions of including student-level data in state data warehouse systems. Should legislative action restrict inclusion of student-level data, the content of the data warehouse should be modified accordingly. This document is written with the assumption that student-level data will be included in the data warehouse.**

CHAPTER 5. ISBE SYSTEMS

ISBE currently maintains over two hundred electronic systems with approximately 100 of them being Web-based. Driven by the business requirements for the data warehouse, this feasibility study focused on the systems most able to provide data to satisfy these requirements. The following includes descriptions of existing systems targeted for inclusion of data in the data warehouse. Also discussed are ISBE's plans for enhancing, consolidating or retiring these systems. Where possible the following systems have been evaluated for data quality. An assessment of data quality considers both the source of the information, as well as the manner in which it is collected. Paper-based systems, and systems that require re-keying, typically suffer from data issues associated with data entry errors. Systems that rely on self-reporting can be susceptible to poor data quality. Poor data quality can be the result of data entry staff not understanding the types of data being requested, or may be intentionally distorted by the data provider to achieve a certain outcome. Data quality evaluations reflect the opinions of ISBE staff responsible for using or maintaining the systems evaluated.

Annual Financial Report (AFR)

Collected at the district level by the School Business and Support Service Division, the Annual Financial Report collects district-level financial information and produces information for the Fiscal CCD, EDEN, the Illinois State Report Card, and other state-mandated financial reports. Data are collected via spreadsheet (which employs some built-in data edit checks) and extracted to an SQL Server relational database after it is visually checked. Each year's data are imported and stored in a separate database table on SQL Server with historical data dating back to 1996.

There are three types of AFRs received by the district: District, Joint Agreement (coop), and -For-Profits. Only values greater than zero are stored in the database; administration costs are automatically calculated and stored and are not entered by the users. Of the 6000 fields available on the AFR, districts typically populate an average of 700-800 values. Data collected by the AFR is considered to be of high quality.

ISBE does not have a plan for modifying or enhancing this system.

Bilingual Annual Student Report

The Bilingual Annual Student Report collects individual student data on students participating in Limited English Proficiency programs and produces information for inclusion in the Consolidated Performance and Title III Performance report. Bilingual data includes enrollment counts (organized by primary language proficiency), graduation rates, withdrawals and program exit numbers. This report suffers from district late reporting issues. Data are compiled and analyzed in spreadsheet format and is not stored as a relational database.

ISBE does not have a plan for modifying or enhancing this system.

Child Nutrition System (CNS) and Application and Claim Entry System (ACES)

The CNS collects application, claim, and monitoring data for sponsors and sites in the National School Lunch Program, Child and Adult Care Homes Program, Child and Adult Care Day Care Centers Program, and the Summer Food Service Program, and produces a batch file to FRIS for payment of claims.

The student identifier system that was developed in conjunction with the CNS application was originally envisioned as a replacement for the existing mainframe RCDTS system; however it is the sole user of this "new" entity system. Data for the CNS is stored in an SQL Server database and deals primarily with payments. It also contains count information on the number of distributed and free and reduced breakfast and lunches. All Nutrition programs are voluntary with entities applying (online) annually through the ACES (front end) interface. Information is collected at the school/entity level with the quality of the data considered to be high.

The Application and Claim Entry System (ACES) facilitates the online collection of CNS sponsor and site application and claim data, and also allows batch submission and processing online. ACES is a Web-based front-end for food program information collection that is stored within the CNS database.

ISBE does not have a plan for modifying or enhancing either the CNS or ACES systems.

Educator Certification System (ECS)

ECS is the master application comprised of several different front-end, Web-based applications connected to a relational SQL Server database. Many of the aggregated numbers produced from this system may be incomplete or deceiving because they lack context for solid definitions, though the overall data quality of ECS is considered to be high. The following is a description of each of the (sub) systems comprising ECS.

The Teacher Certification Information System (TCIS) is the client interface for managing teacher certification data and processes, and is used by Regional Offices of Education (ROE's), Entitling Institutions, and ISBE to enter, store, and process applications for teaching certificates, endorsements, approvals, and teacher service record data within the TCIS database. Within this interface, teachers are uniquely identified by their social security number (SSN) with a medium term goal of moving instead to a unique teacher identifier similar to that used for students. Data collected by this system is considered to be of high quality.

Educators and district administrators use the Web-based system, the Online Teacher Information System (OTIS), to access TCIS data. The OTIS Web site consists of two portals to certification data, the Administrator Portal and the Educator Portal. The Administrator Portal allows district administrators to easily access educator credentials for purposes of identifying, managing, and evaluating teacher placements, including confirmation of the positions for which teachers are NCLB Highly Qualified. The Educator Portal allows teachers to confirm their credentials, register

and renew their certificates, apply for new certificates, request duplicates, and confirm the positions for which they are NCLB Highly Qualified. OTIS accepts credit cards as payment for application services. Data collected by this system is considered to be of high quality.

The Certificate Renewal Tracking System for Teachers and Administrators (CeRTS) Web-based systems allows educators to enter, edit and track professional development activities for certificate renewal, as well as allowing authorized local, regional, and state level officials to review and approve these activities. All the public school administrators in Illinois are required to use the system. CeRTs has the ability to broadcast messages to all the registered educators.

The Teacher Service Record System (TSR) contains individual salary and demographic data for teachers and administrators employed in Illinois public school districts, and produces information for the School Report Card, Annual Statistical Report, Non-Fiscal CCD, EDEN, and EEO-5 Report. Types of information tracked by TSR include all licensed teachers including those in the districts, Vocational Education, special Education and regional offices (there are 1009 reporting entities (LEAs) and 30 charter schools). Information is stored at the teacher level with information on grades and subjects taught, schools, and tenure. Teacher information is updated annually and is stored by staff, district and school. There are approximately 150,000 teachers in the Illinois schools system with roughly 120,000 records stored in the TSR. System data dates back to 1971 with former mainframe records loaded to the current relational database.

There is a current effort to re-brand the entire set of systems comprising the Educator Certification System with the ECS “label”. Additionally, the current systems, interfaces, and look and feel are being modified and consolidated into a more unified whole.

Electronic Grant Management System (eGMS)

Implemented for the 2005 NCLB application period, the eGMS (also known as eGrants) collects application, budget detail, and budget detail data for the NCLB Consolidated Application, and several other entitlement programs. Reporting from the eGMS includes:

- The sum of all children in a district
- Count of grants submitted
- Contact information
- Amounts spent on specified grants

The eGMS feeds budget detail, budget summary, and payment schedule information into the agency’s The Financial Reimbursement Information System (FRIS) reimbursement system. eGMS produces the initial list of Title I schools each year (along with Annual Yearly Progress data) used to determine which schools must offer SES and Choice and collects performance report information. Special Education/IDEAB and Title III data are not maintained within the eGMS. The dynamic nature of the system’s data model may make it difficult to extract data to the data warehouse.

The eGMS continues to be enhanced and expanded by ISBE.

End of Year Report System

The End of Year Report System collects graduation, dropout, suspension, expulsion, enrollment and other end-of-year data annually and produces information for the Non-Fiscal CCD and EDEN. Data are collected electronically and aggregated at the school level. The report suffers from poor data quality due to both a lack of training at the district data entry level, as well as a lack of district understanding of how data are defined and/or aggregated. Variances in school policies mean that numeric comparisons across schools and districts are not meaningful. Consumers/user of data collected by this system includes:

1. Accountability
2. Federal government
3. Department of Corrections
4. Regional offices
5. Schools
6. Boy Scouts/Girl Scouts
7. Businesses (many of whom sponsor after school programs)
8. Media

The End of Year Report is slated to be replaced by the ISBE SIS system.

Fall Housing Report System

The Fall Housing Report System collects aggregated public school and district enrollment data in various demographic categories, and produces information for the Non-Fiscal Common Core of Data (CCD) and the Electronic Data Exchange Network (EDEN). Data are collected as of an October 1 date and aggregated by grade, race and gender. The Fall Housing Report suffers from a problem of late reporting from districts and questionable data quality (districts are self reporting). The information collected in this report is used by:

1. Nutrition Programs and Support Services
2. Bilingual, Regional Safe Schools
3. English Language Learning
4. School Business and Support Services
5. Fiscal Services (The Textbook Loan Program)

Fall Housing data are collected through the Web (except for Chicago which emails their file as an

email attachment), which is extracted, to the mainframe where it is combined with Chicago data. The mainframe represents the final picture of fall housing data. Flat files are generated off the mainframe for analysis and reporting using Excel. The Fall Housing Reports only includes public schools and is collected at the school level (Previously, it was collected at the district level, but this process has been discontinued). Data on the mainframe goes back to the 1970s.

The Student Information System, or SIS, (currently under development) is being built to replace a number of the collection areas currently obtained through the Fall Housing Report. The longer-term goal of the SIS is to have it completely replace the data collection done through the Fall Housing Report.

Illinois Purchased Care Review Board System (PCRB)

Used by Special Education, the PCRB collects and maintains consolidated financial statement data for over 350 providers of services to special education students, housing data about each provider, and produces education and room & board rates for each provider.

Organizations offering special education services use the PCRB to submit paper reimbursement requests that are keypunched into the divisions SQL Server database. There are data quality issues with the PCRB associated with entry errors. Historical information exists in the PCRB dating back to 1999.

This information is in turn used by various other systems at ISBE (for example SEARS to reimburse the providers). PCRB data are of particular interest to special education specialists with cost and rate information (including per diem rates) for providing services used by other education divisions.

A new system has just been implemented that only contains the current year's data (this system has been integrated with SEARS).

Illinois Report Card System

The Report Card System collects data gathered from other systems and organizes it into a format used for generating annual the state report card. The majority of student assessment data derives from testing vendors contracted by ISBE with information received in an aggregated format by grade, by school, and by district and state.

Data are stored in a SQL Server database that, in addition to generating report cards, is used to provide information for EDEN reporting, the Consolidated Performance Report, and in calculating Annual Measurable Achievement Objectives for Title III programs. Due to SQL Server limits on the total number of fields that can be stored in a table, report card data are split across multiple tables. SAS is used against this relational database for data retrieval and complicated calculations with Crystal Reports used for most reporting needs and the generation of the Report Card PDF files.

Report Card historical information dates back to the original records residing on the mainframe. The SIS system and/or the data warehouse, over time, will replace the source systems and data that are used to populate the Report Card System's database. For the foreseeable future, the Report Card System itself will continue to be used for the generation of the actual Report Card, as well as for delivering report card data to organizations requesting it in electronic form. This approach is deemed important to providing all parties with consistent data that will yield the same results including rounding.

Illinois Student Assessment System

The Illinois Student Assessment System is an amalgamation of statewide, vendor-administered student assessment tests. Test results are collected and turned over to the agency for analysis and distribution. Assessment data are collected at the student level and is used to produce required program accountability information required for No Child Left Behind and the Illinois State Report Card.

ISBE is in the process of implementing a consolidated database for storing and analyzing student test scores. The ISBE SIS is slated to capture assessment data in its next phase of development.

Illinois Student Information System (ISIS)

Maintained as a "stand-alone" system, ISIS is a vendor-provided application containing individual student data on students taking career and technical education courses, and produces required program accountability for Carl Perkins.

Data collection falls into two categories:

- Planning (PAS – Program Approval System), which is used for determining items for which schools have been approved, and
- Entry, Analysis and Reporting – Used by the regions and schools for data entry and analysis

ISIS is used exclusively by technical schools (Public schools do not use.), with regions responsible for some level of data entry. In 21 of the sixty regions, the schools do not complete data entry. The regions are exclusively responsible for the data entry activities. There are over 30,000 technical education students in Illinois. Some regions load only their Tech Ed. students to the ISIS database, while others, for expediency purposes, load their entire student rolls.

Data are collected at the student-level and reported by each district at a school level. Information is received by the agency as "tracking files" loaded to the mainframe-based ISIS system. Information collected includes student demographics, courses, credit and specialized calculations for vocational education training. Assessment data are not part of the ISIS system. The ISIS system is not longitudinal, current data overwrites past year data with the system containing Perkins data covering a rotating two years. Data quality is considered to be good.

There is an effort to move the data collection activity away from the current Access-based interface to a .NET/Web implementation.

ISBE Student Information System (ISBE SIS) – under development

Developed in response to the passage of the "No Child Left Behind Act" (NCLB) legislation (Public Law 107-100), and the anticipated reauthorizations of IDEA and Carl Perkins, the ISBE SIS is designed to meet federal reporting requirements. Currently, the system is in the testing/pilot stage of Phase II, when complete the ISBE SIS will contain:

- Unique student identifiers for each student
- Student demographic data (not including SSN)
- Program participation indicators including indicators for Spec Ed, Tech Ed, Bilingual, Truant Alternatives, and Safe School information.
- Enrollment data
- FTE counts (with logic to ensure students aren't over or under counted

Demographic information collected includes legal name (last, first), birth date and gender. Optional information includes middle initial, race and birthplace. There are constraints on the format/content of the student identifier. It is a 9-digit number that cannot start with 0 nor have the same number repeat in a sequence (to avoid incidents of "666"). Of the 2 million students in Illinois, 1.5 million IDs have been assigned. Phase II of the SIS will contain demographic, enrollment/exit, attendance, performance, and program participation data for each public school student in Illinois.

Once completed, the ISBE SIS will facilitate longitudinal data analyses to respond to federal and state accountability and reporting requirements and will negate the need for several existing individual and aggregate student-reporting systems. Slated for later development are disciplinary and assessment information, and Career and Technical Education/ISIS. The SIS will not contain, due to volume and bandwidth considerations, student course and schedule information.

The frequency of data updates to the SIS is determined by the districts (not dictated by ISBE) with upload files are accepted as XML and CSV (SIFF format is not supported). Batch processing of uploads is conducted nightly. Data changes to SIS may only occur at the district, to the source file, which is then reloaded to the database. SIS only contains public students. Early Childhood (daycares) may be added. There are no plans to capture non-public or religious entities. Enrollment information is collected at the school level. Historical data was not imported into SIS with the first year of system data being the 05-06 school year.

The ISBE SIS is currently in the pilot phase.

Migrant

Collected by NGS, a vendor based out of Texas, Migrant student information is reported at the program level for achievement, participation, and programs offered.

ISBE does not have a plan for modifying or enhancing this system.

Performance Management Information System (PMIS)

The PMIS application matches individual student records contained in ISIS to external databases in order to produce analyses and generate reports of student performance in relation to core Carl Perkins indicators.

The PMIS appears to be a viable candidate for replacement by either the data warehouse or SIS.

Professional Development Provider System (PDP)

PDP provides a mechanism for individuals, professional organizations, districts, and other entities to apply for approval and submit Notices of Professional Development. Districts and institutions of higher education may submit coursework.

ISBE does not have a plan for modifying or enhancing this system.

Registration, Enrollment and Staff Report (Non Public Fall Housing)

This report is the non-public equivalent of the Fall Housing Report and is used to collect and maintain data and reporting for non-public or private schools that register with ISBE. Data for this system is collected annually via paper and then keypunched into an ISBE mainframe database file. Information is collected at the school level and aggregated by category (e.g. gender, race). Paper forms are considered very “dirty” with data clean up extremely cumbersome.

Non-public information is stored in two formats: a master file and a print file. The print file contains the same information as the master file, except that address information is extracted from an external system and embedded into the print file. Current reporting is rudimentary with a great deal of staff effort required to support ad hoc query capabilities. There is a minimum of 10 years of non-public historical data in the mainframe.

There are no current ISBE plans to modify the data collection procedures for Non Public Fall Housing information or move it from the mainframe to a relational database platform.

Special Education Approval and Reimbursement System (SEARS)

The SEARS system collects individual student data on students with Individualized Education Programs, personnel data on all full and part time special education staff, and produces required program information for IDEA and state-mandated special education programs. The data collection for this system is extremely comprehensive and highly confidential. Data elements include teacher and staff names, social security numbers and credentials. Data collected by

FACTS system is used to populate the SEARS relational database. This data collection is primarily used by Special Education Services, but is referenced across other programs, systems and reports. Information in the SEARS system's SQL Server database dates back to 2001. Mainframe SEARS data are available back to the 1970s.

ISBE does not have a plan for modifying or enhancing this system.

Special Education Data System (SEDS) – under development

The SEDS data system will maintain the current in production Due Process System. Information to be maintained in SEDS includes all special education due process and mediation requests and proceedings. SEDS will integrate the information maintained within the current system and develop a new database to assist with the implementation of an effective system for identifying and correcting areas of noncompliance through the state monitoring process; ensure timely resolution of complaint investigations; ensure the mediation process is effective and in a timely manner; ensure due process hearings and administrative review decisions are reached within Part B timelines; and ensure the State has a mechanism in place to compile and analyze data to identify systemic issues and problems. Information in the SEDS system is collected at the student level but reported at the district level. On implementation the system will allow parents to view their own children's data.

The Financial Reimbursement Information System (FRIS)

Serving as the agency's financial reimbursement system, FRIS interacts with the agency's MIDAS financial system to disburse state and federal funds to agency payees including customers, districts, daycares, and schools. Serving as the agency's "checkbook", FRIS contains information on vendors, schools, Region/County/District/Type (RCDT) codes, allocation amounts and breakdowns, and fund sources (There is little to no duplication of data between FRIS and MIDAS with district-level allocations and vendor information residing only in MIDAS.).

FRIS run vouchers multiple times in a day and generates payment requests that trigger MIDAS to cut checks. The data quality in FRIS is considered good because the data are scrubbed before it enters FRIS. FRIS information dates back to 1985. District consolidations are a complex process with multiple impacts to FRIS due to the realignment of districts and reassignment or retiring of identifying information. Cross fiscal year information is difficult to "associate" to a district that no longer exists in its original form – the data warehouse may have issues tracking against consolidated districts.

FRIS is used in reporting grant allotment and expenditure information to US Education Department.

ISBE does not have a plan for modifying or enhancing this system.

The Funding and Child Tracking System (FACTS)

The Harrisburg Project was created as an organization responsible for collecting Special Education information. 100% funded by ISBE, Harrisburg transmission sites are used for the collecting of district Special Education funding and approval tracking information. Data are collected at a student level 4-5 times a year with the frequency increasing. Student Social Security numbers are not collected but the majority of students have received a student identifier. Staff data are collected in the same method as student data and includes records for subjects taught, work assignments, and days worked. Teachers are identified within the system via their SSN. SEARS data are used to generate a federal reporting statistic of unduplicated students by district (the process is run for a 12/1 reporting date).

ISBE does not have a plan for modifying or enhancing this system.

CHAPTER 6. THE IMPACT OF THE DATA WAREHOUSE

"What is the impact of creating a data warehouse and what are the corresponding risks and issues to consider?"

6.1. Overview

The building and implementation of a data warehouse will have considerable impact on the operations of ISBE. As with most system implementations, it is necessary to review not only the benefits and risk of implementation, but also address those areas that will influence or be influenced by the new system. This chapter will discuss the benefits and risks of implementing a data warehouse, as well as discuss how the ISBE data warehouse will impact the privacy, security and confidentiality policies of the agency. Finally, as ISBE further defines its Disaster Recovery Plan, it is important to identify issues related to the data warehouse that will impact the agency's Plan.

6.2. Benefits of Building the Data Warehouse

The integrated and homogenous nature of a data warehouse, coupled with its innate reporting and querying strengths, should bring a broad spectrum of benefits to ISBE and the targeted user group profiles of the data warehouse. The following is a partial list of benefits ISBE should expect to realize with the successful implementation of a data warehouse:

1. An integration of ISBE data from across many years
2. Integration of data from key ISBE systems
3. Automation of the production and distribution of reports via email or the web
4. Provision of an online analytical capability to authorized users at the state, regional, district and local levels
5. A mechanism for allowing authorized users at the state, regional, district and local levels to extract data for further analysis using other analysis tools
6. Easier, quicker and more comprehensive reporting and querying capabilities
7. More internally consistent and higher quality query and reporting
8. Reduced staff time spent on data collection activities with more staff time available for data analysis.
9. Retirement of redundant and/or obsolete data collection systems and a reduced data entry burden on schools, districts and ISBE staff.
10. Retirement of systems built and maintained specifically for the generation and reporting of federal and state documents with a freeing of resources previously used for maintenance of these systems.

11. Improved communications resulting from standardizing data collection elements and a clearer approach to training and data collection education.
12. Improved state and federal reporting both from a timeliness and quality perspective with greatly enhanced capabilities for meeting NCLB's increased reporting requirements
13. Provide a mechanism for ISBE management to monitor district performance and address issues previously identified during federal audits.
14. The capacity to analyze large amounts of historical data for "nuggets of wisdom" that can enable ISBE to provide better services to its constituents.

6.3. Risks of Building the Data Warehouse

Acknowledging and addressing the potential risks involved with the successful inception and completion of a Data Warehouse are equally as important as focusing on its benefits. Below are listed some of the risks that will need to be addressed to ensure the success of an ISBE data warehouse:

1. Lack of strong project direction

   A data warehouse implementation is an incredibly daunting task. It will be resource-intensive, expensive, and a complex undertaking that may last for several years. To assure project success, a strong project sponsor(s) must be attached to the effort. Not only must this individual have a clear and committed vision for the data warehouse, but must have the political and organizational clout to clear organizational and technical hurtles that may impede progress. Both the loss of a project sponsor or the replacement of a strong one by a weaker one can significantly increase the risk of project failure.

2. Lack of, or misdirected, data warehouse missions and objectives

   A data warehouse implementation cannot be successful if it does not sufficiently meet the needs and expectations of its user community. This feasibility study attempts to identify and prioritize the missions and objectives of the data warehouse. Over the development and overall life of the data warehouse these missions and objectives should be regularly reviewed and revised based on evolving user community expectations.

3. Continuing use of the existing RCDT identification system

   RCDT, the current mainframe-based identifier system, has been in use for well over 20 years. In its current state, the system contains duplicate entries, entries for non-existent entities, and mis-assignments of IDs based on incorrect assignment of types. Additionally, the system lacks the ability to assign identifiers to vendors and other entities ISBE would like to track. Building a data warehouse system without replacing this system has been described by several at ISBE as "building a house on a foundation of sand.

4. Performance and/or capacity limitations due to technology constraints

The detailed, historic, snap-shot-in-time nature of a data warehouse means that the volume of records it may contain will far exceed the volume of records found in ISBE's transactional system. Rather than record counts in the thousands or millions, ISBE may find its data warehouse contains records in the billions. ISBE's data warehouse approach must be built around a set of hardware and software tools scalable to those volumes.

5. A continuation of the "Silo" mentality

   Within ISBE there is a strong sense of data ownership within divisions responsible for gathering data. This "silo" mentality is indicative of an environment where individuals and agencies are narrowly focused on immediate responsibilities and concerns. There is no "big picture" understanding of the organizations larger aims or status. Individuals and divisions unwilling to release their hold on data can hamstring or even sabotage a data warehouse effort. A cultural shift towards viewing data as a communal resource must be part of the change that accompanies a data warehouse implementation.

6. Poor helpdesk support, quality documentation and training

   The data warehouse is targeted for use by a variety of users with different levels of comfort and proficiency on computers. Insufficient support for users can critically undercut acceptance and use of a data warehouse and lead to criticisms of the tool that are not related to its development or implementation. ISBE must plan and budget support for all aspects of data warehouse operations including internal setup and maintenance (including ETL processes), internal staff training, help desk support, and product documentation

7. Lack of a data steward function

   A data steward acts as a central policy and oversight body to assure the quality of data that is introduced into ISBE's data collection systems, as well as providing a consistent policy for how collected information is used and distributed. Without this function, the data warehouse is highly susceptible to misuse of data, inconsistent policy decision-making, and an inability to evolve as the needs of ISBE change

8. Dependency on the ongoing success of the SIS implementation

   The SIS system is part of ISBE's longer-term strategy to consolidate, streamline, and improve ISBE's data gathering processes. Inevitably, the data warehouse will have a high dependency on the SIS as a major source of data. While this dependency cannot be ameliorated, the success of both systems can be better assured if the two systems are viewed as interconnected components of an overall data strategy.

9. Inadequate preparation, distribution and training regarding privacy, security and confidentiality policies and procedures

FERPA is a driving force in determining appropriate use, access and distribution of information from the data warehouse. To avoid conflicts with FERPA requirements, as well as state and federal regulations, ISBE must clearly define its privacy and confidentiality policies. The policies must be clear, universally applied, and understood throughout the agency. Even after the implementation of a data warehouse ISBE must continually communicate and train ISBE staff, as well as districts and schools, in the agency's established policies.

10. Historically there have been delays in districts submitting data to ISBE

    Currently, ISBE experiences delays in receiving data from the districts. Many of the districts are experiencing manpower shortages that result in their inability to collect, verify, assimilate and submit data in a timely fashion. Should this problem continue, it could have significant impact on quality of the data in the Data Warehouse.

11. Unrealistic expectations for the development timeline and the usability of the data warehouse.

    To distribute the cost of building the data warehouse, ISBE will most likely use a phased implementation approach. This creates a risk in that users may not be able to immediately access the types of reports they expect from a data warehouse nor be able to complete the types of data queries they desire when the data warehouse is initially available. To avoid this risk, it will be important to communicate the project timeline and the phased implementation approach to internal and external users.

6.4. Privacy, Security and Confidentiality

Privacy, security, and confidentiality issues related to the data stored in the data warehouse will have significant impact on the immediate and long-term success of the ISBE Data Warehouse. Federal and state laws such as the federal Family Education Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 CRF Part 99) and the Illinois School Student Records Act (105 ILCS 20/), as well as the Individuals with Disabilities in Education Act (IDEA, 34 CRP § 300.127 and 300.560-300.576), include provisions that protect student and staff information. While these laws have been in place for some time and ISBE has responsibly developed practices and procedures to address and comply with the laws, it is important to review, modify and formalize existing data access and management practices as ISBE implements the new data warehouse.

For the purpose of this document, we will use the following definitions for privacy, confidentiality and security as used in the NCES' *Forum Guide to Protecting the Privacy of Student Information:*

- Confidentiality refers to an obligation not to disclose or transmit information to unauthorized parties

- Privacy reflects an individual's freedom from intrusion
- Security refers to technical procedures that ensure only authorized and intended parties have access to data.[4]

6.4.1. Systems of Data Protection

There are various approaches to implementing systems of data protection. Each of these should be addressed and incorporated into a data access and management policy.

Data Steward and Custodial Levels

As discussed in the best practices section of this document, the practice of identifying an individual or group to be ultimately responsible for the content of the data warehouse and administer the security measures and manage the data access is very important to the integrity of both of the policy and the data warehouse. In other state education agencies, this responsibility often falls to the CIO of Department of Education or someone with similar experience and expertise. Recognizing that ISBE has limited resources to direct to this purpose, a group of individuals assigned as a Data Quality Assurance Team could equally meet these responsibilities. In addition to the ownership of the data at the state level and the administration of privacy, security and confidentiality issues, this group would also be responsible for the maintenance of data dictionary, how data will be collected and how the districts will provide data to the state. On a local education agency level, districts or other primary sources of the data should be responsible for the quality and integrity of the data submitted.

Maintaining Confidentiality

FERPA regulates that no private or confidential data can be released by a state education agency without parental permission except to the following parties or under the following conditions as stated in 34 CFR Part 99.31 Final Regulations for FERPA:

- School officials with a legitimate educational interest;
- Specified officials for audit or evaluation purposes;
- Appropriate parties in connection with financial aid to a student;
- Organizations conducting certain studies for or on behalf of the school;
- Accrediting organizations;
- To comply with a judicial order or lawfully issued subpoena;
- Appropriate officials in cases of health and safety emergencies; and

[4]The National Forum on Education Statistics (2005). *Forum Guide to Building a Culture of Quality Data: A School and District Resource: 2004* (NCES 2005-801). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

- State and local authorities, within a juvenile system, pursuant to specific state law.

It is important to put measures into place that ensure FERPA law is met. This might include requiring that third parties obtaining data from ISBE must receive written consent from the CIO or designated staff member within ISBE before they can access information. A second measure may include restricting data so that the individual(s) obtaining the written consent could only view personal identifiable information. Finally, there should be a measure in place that stipulates that the data should be destroyed after the purpose under which the disclosure was fulfilled.

Defined Levels of Access

Access to the data warehouse is accomplished via two methods: through the publicly available ISBE website, and through the ISBE's IWAS system. Data available through the public website will be public information only and will not include private or student-level data. District, regional and ISBE access to confidential or student-level data will be through IWAS. To support this functionality, the IWAS system will need to be expanded to include authorization and authentication to the data warehouse.

Access to the data warehouse information should fall into four logical access layers including a public layer, an analytical layer, a transaction layer, and a private layer.

1. A public layer that is accessible through ISBE's Website that consists of pre-defined aggregate reports and are loaded to the ISBE Website as HTML web pages and OLAP Public reports that are refreshed on a daily, weekly, monthly and/or quarterly basis. Data on individual students **will not** be accessed by anyone at this read-only level.
2. An analytical layer that provides multiple OLAP layers (cubes) accessible only to authorized ISBE and District/School staff with OLAP multidimensional data cubes updated from the ISBE Transactional Layer information, OLAP information dimensioned, aggregated, stratified and stored into the OLAP layer, OLAP delivers requested information to spreadsheets or tabular reports for further data manipulation and graphing, OLAP information aggregated and de-personalized protecting confidentiality of consumer information, OLAP data manipulation is drag and drop and drill-downs.
3. A transaction layer containing detailed and de-personalized information (names, addresses, and Social Security numbers have been removed), query-able by Reports Wizard(s). This level of access would be recommended for ISBE staff responsible for state and federal reporting. Access to specific data elements for reporting purposes could be approved on a case-by-case basis.
4. A private layer containing data elements that include private and confidential information. This level of data access should be limited to a select number of authorized staff within ISBE that may need to access this level of data in order to do sophisticated analysis and reporting requiring access to detailed student-level data.

It is recommended that ISBE review data access provisions annually to ensure that changes to the data access provisions are made according to any changes in federal and state laws and/or

administrative rules and guidelines.

Unique Student Identifier/Unique Teacher Identifier

The creation of the SIS system has included the development of a component for assigning unique student IDs. This 9-digit number is generated using a random ID generator that ensures that identifiers are “non meaningful”, cannot start with 0, nor have the same number repeat in a sequence (to avoid incidents of “666”). The unique student ID is used in lieu of individual, personally identifying information (including name, Social Security number, address and birth date), to individually identify each public student in the ISBE school system. Of the 2 million students in Illinois, 1.5 million IDs have been assigned.

The ECS system, which contains most of ISBE’s teacher-related information, tracks teachers using their Social Security number and/or a combination of their first and last names. As a part of ECS’s ongoing enhancement, a unique teacher ID component is to be developed similar to that created the SIS.

Data Sharing Between Agencies

Another provision in maintaining the security, privacy and confidentiality standards is to determine how and when data from the ISBE data warehouse will be shared with other Illinois agencies. By identifying the agencies, setting up reciprocal agreements and identifying the specific data elements will further give credibility to the data warehouse initiative and further ensure the security, privacy and confidentiality standards of ISBE.

Research Standards

An additional aspect of maintaining privacy of the data within the data warehouse, will be to determine the research standards for personal identifiable information. This includes establishing a statistical size limitation for data extracted from the data warehouse. Depending on the data element it may be possible to personally identify student or teacher information when aggregated data are below a certain level. To ensure that personal identifiable information is protected even in an aggregate format, it is recommended to set a minimum cell size 10 for the aggregated data. Some state law regulates this level of detail. Regardless, it is recommended to determine a minimum cell size and publish that information as part of the data access and management policy.

Data Access and Management Policy

A data access and management policy should define the responsibilities of the state and local education agencies regarding data collection, the use of the data, reporting procedures and address the security and confidentiality issues related to the data. Several state education agencies have put similar policies into place as part of their data warehouse and student records

systems. Below are suggested topics to address in a data access and management policy.

- Definition of terms: privacy, security, and confidentiality
- Clearly defined lines of the ownership and identified custodians of the data including who is responsible and what are the responsibilities of the state education agency, the districts and other local education agencies for particular data elements.
- Defined levels of access depending upon responsibility and purpose
- Security measures to ensure records are not lost, vandalized, or illegally accessed.
- Measures put into place to retain the confidentiality of the data and define when and how data can be accessed and disclosed.

Disaster Recovery

An IT disaster recovery plan provides a process for resuming normal business and IT functions in the event of both short-term losses like temporary loss of electronic equipment or more significant disaster like flooding, long-term power outages, etc. The ISBE Technology Support division is in the process of creating a Disaster Recovery Plan for ISBE. The ISBE Disaster Recovery Plan will focus on "the continuity of operations for the data systems, critical data and other software that are typically run on the agency servers".[5] The functionality of the data warehouse while very important will not affect the daily operations of ISBE; therefore, a disaster recovery plan for a data warehouse is slightly different than for a transactional system.

The ISBE disaster recovery plan should include the following items for the disaster recovery for the data warehouse:

- Identification of individuals who are responsible for data warehouse and compile a list of names and phone numbers who can access the data warehouse and carrying out recovery plans should the data warehouse be affected by a disaster.
- Identification of potential problems that will affect the operations and functionality of the data warehouse and development of recovery strategies and plans for how to address each problem. It is recommended to address both short-term and long-term disasters.
- Determination of how ISBE Technical Support will respond when data source systems for the data warehouse are down or otherwise unavailable or when scheduled extractions of data are interrupted.
- Detailed list of the servers used for the data warehouse with corresponding list of names and contact phone number of vendors where replacement parts for hardware can easily be obtained.
- Define when scheduled data warehouse backups should occur, the format of the data backup and where the backup will be stored.
- Review of Recovery plan on a regular basis.

[5] ISBE Disaster Recovery Plan- Thumbnail Summary dated 11/28/06.

User Support

As with the introduction of any system to ISBE's infrastructure, a data warehouse implementation will necessitate user assistance. ISBE's existing help desk support infrastructure will need to be expanded to support a broad spectrum of data warehouse users. Because the user community will vary in both its level of sophistication, as well it access to the data warehouse, help desk staff must be able to offer assistance across a wide range of data consumers from casual users accessing the data warehouse the public interface to analysts needing assistance with reporting and querying tasks.

CHAPTER 7. BEST PRACTICES

7.1. Summary of Best Practices

As part of the information gathering process for the Feasibility Study, MTW Solutions interviewed representatives from Georgia Department of Education, New York State Education Department, Nebraska Department of Education, New Jersey Department of Education and the Wyoming Department of Education who have played key roles in the development/maintenance of data warehouses or similar systems for their respective Department's of Education. The goal of this exercise was to identify best practices within these agencies and gain an understanding of the various approaches used in the development, implementation and management of these systems. Findings from these interviews will provide ISBE with additional tools and approaches they can incorporate into their data warehouse initiative. Following the best practices summary below are individual profiles of each of the states interviewed.

Identified best practices can be grouped into the following categories:

- Accessibility
- Comprehensiveness
- Usability
- Quality
- Management – (CIO, Privacy Policy)
- Availability of Interpretation and Analysis
- Training
- Schools Interoperability Framework (SIF)

Accessibility

Accessibility is a two-part best practice for data warehouses. Ideally, a data warehouse needs to be accessible to its users and be a resource where users can access the information they are seeking. Conversely, a data warehouse must be secure and have limited access to data that is protected by privacy laws. During the state education agency best practices interviews it was evident that a balance of the two can be successfully achieved. All the states interviewed distinguish how different users of the data warehouse access the data and includes the use of username and password security access to regulate the level of access in the data warehouse. For example, Nebraska use a multi-level access approach to their data warehouses that provides public users with the most aggregated levels of data yet is designed to provide agency staff the access they need to produce state and federal reports. Wyoming provides "canned" reports to the general public via a Web application. Both Nebraska and Georgia collect student-level data and

use unique student identifiers to protect the information. The New York State Education Department Data Warehouse is set up differently. Their system is set up as a three level system with student-level data residing at the second level. The second level is where districts and regions can access the information. The third level resides at the state education agency level and is used strictly for state and federal reporting. Student information at the state level is stripped and encrypted with unique student identifiers. All states experience research organizations, including higher education institutions that constantly seek raw student-level data for research projects. The data warehouse needs to be able to supply this information with personal identifiers stripped and quantities masked where cell sizes are insufficient to keep an individual from being identified.

Comprehensiveness

Creating a comprehensive data warehouse will allow ISBE to reap maximum benefits from the data warehouse. Comprehensiveness in relation to a data warehouse refers to whether the scope of the data includes a broad cross-section of data domains and yet includes the required level of detailed data to provide maximum breadth of data analysis and will allow the data warehouse to be effectively used as a data driven support system. Generally speaking, the more granularly data can be stored in the data warehouse, the more value it has. Aggregating data before it is stored in the data warehouse by default places constraints on how data can be used.

Interviews with Georgia and Nebraska illustrated how these states are balancing these requirements. The Georgia DOE Data Warehouse will capture student demographics, Attendance, Teacher data, Special Education data and Enrollment. Data will be collected at the most detailed level possible including the capture of student-level data. Student-level data are protected through the use of unique student ID number system.

Likewise in Nebraska, their data warehouse will ultimately collect enrollment, student program participation, assessment, staff/teacher information, student demographics, eGrants data, Vocational Education data and location and district data. Nebraska is in the early stages of implementing their Student Record System and anticipates that it will take 5-10 years to incorporate all the data domains. Additionally, Nebraska uses unique student identifiers to assign student numbers in the system. While they collect student-level data, the information is protected through the use of unique user identifier system.

Usability

A usable data warehouse will mean different things to the different user groups. The users groups identified in Chapter 4 include a broad range of users. Even within the individual user groups there will be varied levels of technical abilities and reasons for accessing the data warehouse. Georgia, New York, Nebraska and Wyoming identified similar user groups as those identified for the ISBE data warehouse. However, Wyoming said that public use of the canned reports was

limited because of the lack of education regarding the availability of the reports. However, they each addressed the needs of the user groups as much as possible. Typically, this included easy to access "canned" reports for the general public including media, parents and legislators.

A key usability requirement of any state education agency data warehouse, is the ability to produce federal reports. EDEN is still in a draft state from the federal government but should be solidified in the near term. It stands as a series of consistent reporting elements that must be provided to the federal government. Most likely EDEN will be used as the default standard for data reporting in the areas for which it is relevant. Its definitions and content will most likely expand over time. States are starting to use EDEN as a way to set state standards for data collection, aggregation and data definitions.

Quality

The quality of data within the data warehouse is crucial to the success of the data warehouse. Data quality refers to the accuracy, timeliness, meaningfulness and completeness of data. The quality of the data moving from the source systems to the data warehouse has a direct impact on the quality of the reports and queries pulled from the data warehouse. Each of the states interviewed identified data quality as a key factor in the planning and execution of a data warehouse. Each state had put measures into place to address the quality issues of data, some states had more success in this area than others. For example, in New Jersey Department of Education while they had strong planning measures in the early phases of their project to address quality issues, there was no standard ETL tool used as assessment data was loaded in the data warehouse. The information did not use the same student identifiers as other data collection sources. This resulted in potentially "dirty" data in the data warehouse and also results in inability of users to rely on the reports generated by this data. The other states interviewed had ETL tools in place and were pleased with the quality of data that was being extracted from sources systems and loaded into the data warehouse.

Management

Management best practices identified in the interviewed process strongly support the necessity of solid business practices to be in place throughout the implementation of a data warehouse project. Nebraska experienced similar "silo" effect those identified within ISBE. Management addressed this potential problem by ensuring that the original owners of the data have a chance to validate the data before it was loaded into the data warehouse. Additionally, NDE created a formal data access and management policy, thus alleviating the concern of many who previously controlled access to data and then experienced a lack of control to restrict access following the introduction of the data warehouse. Nebraska is using a pilot test period of one year that will include approximately 105,000 student's data. This approach will further allow NDE to address management issues before expanding the data collection efforts. New Jersey experienced several management issues that negatively impacted their data warehouse. Some of the issues

impacting the project included a lack of a competitive search for vendors to build and implement a data warehouse, a change of oversight committee midway through the project, and a change in upper management (including the loss of its Commissioner and a state budget shortfall). In Georgia having a CIO in place to manage the building and implementation of the data warehouse appears to have played a significant role in the successful implementation of the data warehouse a month earlier than planned and under budget.

Communication/Training

Communication and training are imperative for a successful data warehouse. New York State Education Department developed documents for explaining data quality and created a data dictionary to be used for communicating data collection efforts. Additionally, they plan to produce a guide explaining how the data warehouse will be used and distribute that document as part of training program.

Training is a critical component of a successful data warehouse. Training falls into several main categories:

- Quality data entry
- Propagating common definitions for data and data usage
- Training on the use of the data warehouse, reporting and ad hoc interfaces
- Agency wide training on the appropriate use and access to data
- Training about privacy and confidentiality issues related to data collection and access.

Nebraska views training for districts as an on-going necessity and at a minimum a multi-year process for data improvement. Training will be implemented as part of the pilot program in the spring 2006 and is anticipated to extend through the summer to educate districts on proper definitions of data.

School Interoperability Framework

The Schools Interoperability Framework (SIF) is an independent organization that is dedicated to setting standards to define XML definitions and communication methods for moving school data using Extensible Markup language (XML) and other open protocols between unrelated systems. By using SIF defined XML format, student information can be exchanged between different applications even when these applications use different internal field names and store data in different formats.

Software applications can be SIF Certified or SIF Compliant. Software that is SIF certified has been validated by an independent committee and meets the SIF specifications. There are over 100 vendors who are SIF certified. Selecting vendors who have products are SIF certified ensures that the software will work with other SIF certified software. The Schools Interoperability Framework is a recognized best practice to emulate for an education data warehouse because it

minimizes the amount of special programming and reduces the amount of modifications necessary to transfer data between systems.

7.2. Profiles of Related Systems in State Education Agencies

7.2.1. Nebraska Department of Education

Name of System: Student and Staff Records System (NSSRS)

Data Elements Collected:

- Enrollment
- Student Program Participation
- Assessment
- Staff/Teacher including certified and para-professionals
- Student Demographics
- eGrants
- Vocational Education
- Location and district data

Data Dictionary: Created data dictionary to include data elements, definitions, code sets and field characteristics to be collected.

Ad hoc Reporting: Will be available through use of Cognos in the spring 2006.

Pre-defined Reports: Will be available via Web-based access

Access: multi-level access based on username and password; public access to pre-defined reports

Status of System: Beginning of pilot test year that will include approximately 105,000 students data

Cost of System: Estimated cost approximately $2,015,000.00

7.2.2. New York State Education Department

Name of System: New York State Repository System (SRS)

Data Elements Collected:

- Enrollment
- Student Life
- Student Program Information

- Assessment

Data Dictionary: Created data dictionary to include data elements, definitions, code sets and field characteristics to be collected.

Ad hoc Reporting: Will be available through use of Cognos

Pre-defined Reports: Will be available via Web-based access

Access: Multi-level access based on username and password; pre-defined and parameter-driven reports via Web browser; limited access to data in the Repository System to authorized individuals; access to selected data summaries to be viewed, downloaded by the general public;

Status of System: not determined

Cost of System: not determined

7.2.3. New Jersey Department of Education

Name of System: NJ SMART

Data Elements Collected:

- Student
- Assessment

Data Dictionary: Have a Student Handbook that identifies 92 different data elements

Reporting: System has a reporting tool called Ed Analyzer. It provides reports, ad hoc reporting and has minimal data mining capabilities to generate graphs and scatter diagrams.

Access: Will have multi-level Web based access. This functionality is currently not available.

Status of System: Two of the three pilot districts have uploaded information, which includes assessment data from 1999 forward. Students will be assigned a unique identifier beginning November 2005 through October 2006. The Education Data Warehouse will be developed (Phase 1) in calendar year 2005, and will continue to be developed over many years.

Cost of System: not determined

7.2.4. Wyoming Department of Education

Name of System: Wyoming Integrated Statewide Education Data System (WISE)

Data Elements Collected:

- All Student Information from SIS system
- Special Education
- Transportation

- Body of Evidence
- Instructional Management
- Human Resources
- Library
- School Lunch
- State Reporting Engine
- Finance

Data Dictionary: not determined

Ad hoc Reporting: Currently, the IT department uses PL SQL, XML and Crystal Reports for queries. Currently, they are reviewing Oracle and Discover as querying tools with the plan to move towards .NET implementation.

Pre-defined Reports: A large selection of "canned reports" accessible via WISE are available

Access: Public and districts only have access via Web application. Access directly to the data warehouse is protected by username and password. One person within the agency is responsible for approving access.

Status of System: Wyoming DOE will be implementing a unique Student ID system in February 2006

Cost of System: Not determined. The system was built in house over a period of many years.

7.2.5. Georgia Department of Education

Name of System:

Data Elements Collected:

- Student Demographics
- Attendance
- Teacher
- Special Education
- Enrollment
- Finance (to be added in 2006)

Data Dictionary: Not Determined

Reporting: From the start, the Georgia DOE data warehouse was envisioned to be a longitudinal decision support system. Therefore it has extensive reporting capabilities including both ad hoc reporting capabilities and pre-defined reports.

Access: The data warehouse holds student-level data and therefore has access level controls with strong username and password protection processes in place. The Georgia DOE includes a scorecard/dashboard tool for use by the State Superintendent of Schools.

Status of System: This is the third attempt to implement a data warehouse in GDOE. Other attempts have cost GDOE close to $100 million. The project was restarted in July 2003 and implemented in December 2005. There is a department goal to move to a daily or weekly capturing of data rather than the current monthly basis of data collection.

Cost of System: $8.9 million was the approved project budget.

7.3. Data Quality Campaign

The focus on data quality has been an increasingly important goal of multiple non-profit and for-profit organizations as data collection efforts have expanded in recent years. In 2005, several non-profit and for-profit organizations joined together with the goal of coordinating the efforts to improve the quality of education data. The Data Quality Campaign, managed by the National Center for Educational Accountability (NCEA), has a mission ”to provide support for and advocacy on behalf of organizations that create, collect, and use education data in an effort to improve student achievement.” [6]

There is a federal effort put forth that makes recommendations for collecting and distributing high quality data. They have a 10-point list on how to ensure high quality data in longitudinal systems (to date no state has implemented all the recommendations). The ten points are:

1. A unique statewide student identifier.
2. Student-level enrollment, demographic and program participation information.
3. The ability to match individual students’ test records from year to year to measure academic growth.
4. Information on untested students.
5. A teacher identifier system with the ability to match teachers to students.
6. Student-level transcript information, including information on courses completed and grades earned.
7. Student-level college readiness test scores.
8. Student-level graduation and dropout data.
9. The ability to match student records between the pre-K–12 and postsecondary systems.
10. A state data audit system assessing data quality, validity and reliability.

Endorsers of the Data Quality Campaign believe “In building a statewide data system with each of these components, the state should ensure that student records can be easily transferred, student

[6] DataQualityCampaing.org. “About Us” page.

privacy is protected, data definitions and requirements are clear to all concerned, and the data system is organized in ways that facilitate data use and user-friendly reporting."[7]

[7] Electronic Brochure: Creating a Longitudinal Data System - Using Data to Improve Student Achievement. Produced by Achieve, Inc. the American Diploma Project Network.

CHAPTER 8. RECOMMENDED APPROACH

8.1. Introduction

The following approach is designed to be both vendor neutral and compatible with ISBE's existing technical architecture. There are two concepts discussed within this approach that can be implemented as part of a data warehouse implementation, but are not necessary or required: data marts and operational data stores.

A data mart is a specialized version of a data warehouse. The key difference between the two is that the creation of a data mart is predicated on a specific, predefined need for a certain grouping and configuration of select data. Data marts can either be created as part of the ETL process, essentially an intermediary step in the loading of data to the warehouse, or extracted from the data warehouse for use by groups that do not require access to the data warehouse as a whole.

The Operational Data Store (ODS) is similar to a data warehouse in structure, but is updated more frequently and contains only current, low-level data–existing information is overwritten as updates are loaded from source systems and no aggregation of data takes place. The definition and uses for an ODS are inconsistent and frequently changing. The value in an ODS can be its ability to serve near real-time data to select users (and applications) as well to serve as a primary staging area for a data warehouse implementation.

As ISBE evaluates various implementation options it should weigh the inclusion of data marts or an ODS only to the extent that it adds value to the data warehouse as a whole.

The dimensional data model supporting a data warehouse can be either of the *star* or *snowflake* configurations. The difference between these two models is in their physical implementation; Snowflake schemas support ease of dimension maintenance because they are more normalized. Star schemas are easier for direct user access and often support simpler and more efficient queries. From the standpoint of evaluating potential data warehouse solutions, ISBE need not be concerned about which model is employed by a particular approach, as long as the proposed schema supports its business needs.

There are three main components to a data warehouse: ETL, OLAP and Reporting. These components can be implemented separately or as part of an integrated whole; there are vendors and products that have offerings in each, or all, of these market segments. Neither an "all-in-one" package nor an "a la carte" can lay claim to being the "best" solution for meeting ISBE's business needs. Ultimately, the best combination of functionality, support, and cost should be the determining factors in how ISBE chooses to implement its data warehouse.

8.2. The Data Warehouse Infrastructure and Environment

The underlying architecture to support an ISBE data warehouse should meet the following criteria:

1. It should be built on a component-based model that will allow the warehouse to adapt to future changes without a major redesign,
2. It should be implementable using a phased approach that allows new components to be added over time as funding and resources become available,
3. It should support multiple types of analysis tools and approaches including decision support (DSS), data mining, and trend analysis
4. It should provide for integration and reusability of data and components,
5. It should be affordable to build and maintain, and
6. It should be built upon known, stable technologies that leverage ISBE's existing technology infrastructure and expertise.

Conceptually, ISBE's data warehouse encompasses 5 key areas: an Operational Environment, an Information Environment, a Volatile Reporting Environment, a Communications Environment, and an Quality Assurance environment.

The *Operational Environment* consists of information systems resident throughout the ISBE environment. Based on the analysis conducted on ISBE's existing systems the following systems should comprise the data sources for the data warehouse (each system is listed with its associated data domains):

1. Student Information System (ISBE SIS)
    a. Student demographics
    b. Program Participation
    c. Enrollment
    d. Discipline (future)
    e. Program (future)
    f. Health (future)
2. Financial Reimbursement Information System (FRIS)
    a. Reimbursable Program Budgets
    b. Reimbursement Claims
    c. Vouchers/payments
3. Grants Management System (eGMS)
    a. Grant Programs
    b. Grant applications
    c. Grant approvals
    d. Spending by program/time/application
4. Educator Certification System (ECS)
    a. Teacher demographics
    b. Teacher certificates

    c. Professional Development Tracking
    d. Professional assessment data
    e. Professional development providers
    f. Professional activities (future)
5. Annual Financial Report (AFR)
    a. District financial data elements
6. Child Nutrition System (CNS)
    a. Meals provided aggregated by school and day
    b. Child nutrition budgets by school
    c. Site audits
7. Illinois Student Information System (ISIS)
    a. Programs approved by location
    b. Student demographics
8. Special Education Approval and Reimbursement System (SEARS)
    a. Special education student data
    b. Special education staff data
    c. Special education claim and payment data
9. Special Education Data System (SEDS)
    a. Special education due process and mediation requests and proceedings
10. Public Registration, Enrollment and Staff Report (Non Public Fall Housing)
    a. Annual aggregates of student data
    b. Annual aggregates of student events and activities
    c. Provider information
11. Migrant
    a. Migrant student information at the program level for achievement, participation, and programs offered
12. Facilities and Inventory (to be developed)
    a. Buildings and properties maintained at the school level

The *Operational Environment* should include the processes and hardware for conducting data extraction from the source systems, as well as a staging area for extracted data.

The *Information Environment* contains the core components of a data warehouse: a normalized database, reporting facilities, data mining and exploration environments and ETL. Based on ISBE's technical infrastructure, the data warehouse database should be Microsoft SQL Server. Reporting facilities may leverage existing ISBE software, including Microsoft Report Services and Crystal Reports, in addition to OLAP-specific reporting and data exploration applications.

The *Volatile Reporting Environment* is designed to address issues associated with late information reporting and data load dependencies. Using an Operational Data Store (ODS) to store non-aggregate, near real-time data (referred to as volatile data), a set of reporting data marts, and communication with the data warehouse, the *Volatile Reporting Environment* can be used to provide current longitudinal and trend analysis before finalized data has been loaded to the data warehouse. While the use of an ODS and/or data marts is not required of a data warehouse implementation, an alternate configuration should address interim reporting business requirements.

The *Communication Environment* provides the mechanism(s) necessary to communicate data warehouse information to external and web-based users (including, but not limited to, the general public, the media, other agencies, districts and schools). This environment will typically contain preformatted reports and queries, PDF-generated documents and reports, and rudimentary query interfaces. The *Communication Environment* will be tightly integrated with ISBE's existing public information interface (for providing data warehouse to the public) as well as the IWAS portal application in controlling and defining user access and authorization for ISBE staff and district users.

The *Quality Assurance Environment* is concerned with two aspects of data quality: that data contained within the data warehouse is accurate, and that data reported or displayed from the data warehouse is displayed *and interpreted* correctly. The *Quality Assurance Environment* will contain functionality to support data audit. This environment is the "softest" of the defined environments in that it will be governed more by practical policy and staff activity than by hardware or software. Also included in this environment will be the functions performed by the data steward role including common data definitions, training, and data confidentiality and data sharing policies.

8.3. Data Extraction, Transformation and Loading (ETL)

ETL is a complex combination of process and technology that consumes a significant portion of the data warehouse development efforts and requires the skills of business analysts, database designers, and application developers. The ETL is not a one-time event; new data will be added to the data warehouse on a regular basis (typically monthly, weekly, daily, or even hourly - depending on the business requirements related to specific data elements). When evaluating potential ETL vendors, it is essential to determine who will be responsible for writing data extraction routines. Unless otherwise determined, ISBE should expect that a significant portion of extraction efforts will fall to its own staff. The level of guidance and assistance a vendor offers in this area is critical. ETL tools include both transformation and cleansing components. ETL tools typically are geared towards either strong transformation or strong cleansing capabilities, but not necessarily both.

Transformation, or the process responsible for data validation, data accuracy, data type conversion and business rule application, can take place while data are being extracted from the originating Online Transaction Processing (OLTP) system (referred to as inline transformation). Inline transformation implementations are typically less robust and flexible than ones that confine transformations to the transformation component of the ETL tool. Transformations performed by the OLTP impose a performance burden on OLTP database performance and split the transformation logic between two separate ETL components (which adds complexity when ETL logic needs to be modified). As a general rule, ETL systems that employ inline transformation are less preferable than those that do not.

The loading component of an ETL solution is responsible for loading data into the data warehouse database. Data warehouses are usually updated periodically, rather than continuously, and large numbers of records are often loaded to multiple tables in a single data load. The data warehouse is often taken offline during update operations, so that data can be loaded faster and OLAP cubes can be updated.

An ETL approach should:

1. Be able to connect directly to all ISBE data sources, regardless of their format or technology
2. Where possible, use graphical, intuitive interfaces for the setup, scheduling, and maintenance of ETL processes
3. Be automated with well-documented operational procedures that can be modified as the data warehouse evolves
4. Contain a metadata component used for the maintenance of information about the movement and transformation of data and the operation of the data warehouse. The metadata component should document the data mappings used during data transformations and be compatible with ISBE's overall metadata strategy
5. Confine transformation processing to the transformation component (no inline transformation)
6. Have a loading component designed to maximize load efficiency and performance and minimize data warehouse offline time
7. Include training and ongoing support for the compilation and maintenance of the ETL process

Popular ETL tools include Data Junction, Ascential DataStage, Ab Initio, and Informatica.

8.4. Online Analytical Processing (OLAP)

Online analytical processing (OLAP) is the technology that enables client applications to efficiently access data warehouse data. The data cube is the foundation of all OLAP tools. A data cube is a type of multidimensional matrix that lets users explore and analyze a collection of data from many different perspectives, usually considering three factors (dimensions) at a time. Data cubes are not restricted to just three dimensions; most OLAP systems can build data cubes with significantly more dimensions (for example, Microsoft SQL Server 2000 Analysis Services allows up to 64 dimensions).

Not surprisingly, the exact definition of what an OLAP tool is, and what it does, appears to vary by tool and vendor. Within a general context, OLAP tool characteristics include:

- An intuitive multi-dimensional data model that makes it easy to select, navigate, and explore data,

- An analytical query language providing the power to explore complex data relationships,
- Pre-calculation of frequently queried data to ensure fast response times to ad hoc queries
- The ability view and manipulate web-based queries and reports

OLAP storage comes in three forms:

- Multidimensional OLAP (MOLAP) where both the source data and the aggregations are in stores in a multidimensional format. MOLAP is the fastest option for data retrieval, but requires the most disk space (disk space may not be a concern given the continuing drop in prices for data storage and processing)
- Relational OLAP (ROLAP) where all data, including aggregations are stored within the source relational database. This will be a concern for larger data warehousing implementations that have higher usage needs. ROLAP is the slowest for data retrieval. Whether an aggregation exists or not, a ROLAP database must access the data warehouse itself. ROLAP is best suited for smaller data warehousing implementations.
- Hybrid OLAP (HOLAP) is a combination of both the above storage methodologies. HOLAP databases store the aggregations that exist within a multidimensional structure, leaving the cell-level data itself in a relational form. Where the data are pre aggregated, HOLAP offers the performance of MOLAP, where the data must be fetched from the tables. HOLAP is as slow as ROLAP.

The major OLAP vendors are Hyperion, Cognos, Business Objects, and MicroStrategy. Major database vendors have started to incorporate OLAP modules within their database offerings. Microsoft SQL Server 2000 with Analysis Services now includes an OLAP offering (as does Oracle with Express and Darwin and IBM with DB2). OLAP solutions are expensive and, in many cases, extremely proprietary. The following should be considered when evaluating OLAP vendors:

1. Is the infrastructure and technology of the proposed tool compatible with ISBE's technical infrastructure?
2. What is the OLAP vendor's history? How many installs has the vendor made?
3. What cube sizes and transaction speeds are supported? The system should deliver most responses to users within about five seconds, with the simplest analyses taking no more than one second and very few taking more than 20 seconds (*as defined by The OLAP Report, www.olapreport.com,* Nigel Pendse*)*
4. How will the OLAP cube be created – HOLAP, MOLAP, ROLAP?
5. Is a thin client, desktop, or web-based interface used for interacting with the data cube?
6. Are analysis and drill-down accomplished via mouse clicks? Are other modes of analysis available (not all users prefer the same approach)?
7. What data extraction/report result extraction capabilities are available?
8. Is the OLAP tool compatible with ISBE's IWAS-based security approach?

9. Does the tool have the ability to mask data below pre-set aggregated levels and/or hide values below a specified cell size based on security assigned access rights?
10. How much consulting will be required to install the product? If changes are needed after installation, will this require the need of external consultants?
11. How tightly integrated is the OLAP tool tied to the database? Can queries be performed within the OLAP tool?

8.5. Report Creation and Distribution

Data warehouse report creation and distribution holds the greatest range of solution possibilities including the leverage of existing ISBE software (including Microsoft Report Services and Crystal Reports) and skills. Many OLAP tool offerings include a front-end presentation layer that allows users to call up pre-defined reports or create ad hoc reports. Popular report vendors include Business Objects (Crystal Reports), Cognos, and Actuate. The following should be considered when evaluating Reporting Tools (based on information provided by 1keydata.com, www.1keydata.com):

1. The tool should be able to connect to both to the data warehouse relational database and to the OLAP multidimensional data source.
2. The reporting tool must have scheduling and distribution capabilities including the ability to schedule regular, recurring reports. Reports should be distributable via email or Web publishing
3. Access/security should be customizable at the report, folder, column, row, or individual cell level. The tool's security layer should be able to interact with ISBE's IWAS security functions.
4. The tool should be able to format reports based on existing ISBE templates and style sheets.
5. Export capabilities should include, at a minimum, the ability to generate files in Excel, ASCII flat file, and PDF formats. For Excel extracts, the tool should be able to generate both the report's format, as well as its content.

8.6. Phased Implementation

A phased implementation of existing data collections is recommended for the completion of the data warehouse. Sequencing of data collections is based on assessments of each system's data content and quality, format, and "user demand". Based on the evaluation of each of the systems reviewed for the feasibility study, the following systems, and their recommended sequence for data collection implementation is:

1. Student Information System (ISBE SIS)
2. Financial Reimbursement Information System (FRIS)
3. Grants Management System (eGMS)

4. Educator Certification System (ECS)
5. Annual Financial Report (AFR)
6. Child Nutrition System (CNS)
7. Illinois Student Information System (ISIS)
8. Special Education Approval and Reimbursement System (SEARS)
9. Special Education Data System (SEDS)
10. Public Registration, Enrollment and Staff Report (Non Public Fall Housing)
11. Migrant
12. Facilities and Inventory (to be developed)

Each of these systems will have a frequency load rate compatible to the type of data that it will be supplying to the data warehouse. Below is the load rate for each of the twelve systems.

System	Frequency Load Rate
Annual Financial Report (AFR)	Annually
Child Nutrition System (CNS)	Weekly
Educator Certification System (ECS)	Monthly
Facilities and Inventory	To be determined
Financial Reimbursement Information System (FRIS)	Weekly
Grants Management System (eGMS)	Monthly
Illinois Student Information System (ISIS)	Weekly
Migrant	Monthly
Public Registration, Enrollment and Staff Report (Non Public Fall Housing)	Annually
Special Education Approval and Reimbursement System (SEARS)	Weekly
Special Education Data System (SEDS)	Weekly
Student Information System (ISBE SIS)	Daily

Systems not to be included the data warehouse are:

System	Reason for Exclusion
Fall Housing Report	To be replaced by the SIS
End of Year Report	To be replaced by the SIS
Illinois Purchased Care Review Board (PCRB)	Integrated into SEARS
Performance Management Information System (PMIS)	To be replaced by the SIS
Illinois Student Assessment	To be incorporated into SIS

System	
Illinois Report Card System	Information is derived from other source systems already targeted for feeds to the data warehouse
Professional Development Provider System (PDP)	The information is of limited value to a data warehouse
FACTS	Data are available in SEARS
Bilingual Annual Student Report	Data are not stored in a database

Based on business requirements, systems outside of the data warehouse need to be built or modified to provide:

1. Facilities and Inventory information at a school and district level
2. Entity System for providing unique identifiers to all entities tracked within ISBE currently identified by RCDT codes.

It is recommended that the ISBE complete a pilot program with a minimum of two data collection systems populating the data warehouse. This will provide ISBE the opportunity to adjust ETL practices and identify a set of reports prior to building the entire data warehouse.

8.7. Historical Data Loads

The following table outlines the scope of historical data to be loaded to the data warehouse.

System	Data Domain	Date Range
SEARS	Student Data	2001 – present
	Staff Data	2001 – present
	Claim and Payment Data	2001 – present
	IPCRB – Service Providers	1999 – present
	IPCRB – Re-imbursement Rates	1999 – present
	IPCRB – Approved Re-imbursements	1999 – present
End of Year and Fall Housing Report	Annual Aggregates of Student Data	1970 – present
	Annual Aggregates of Student Events and Activities	1970 – present
Public Registration, Enrollment and Staff Report (Non Public Fall Housing)	Annual Aggregates of Student Data	1995 – present
	Annual Aggregates of Student Events and Activities	1995 – present
Child Nutrition System	Meals Provided aggregated by school and day	1999 – present
	Child Nutrition budgets for each school	1999 – present
	Site Audits	1999 – present
EGMS	Grant Programs	2005 – present
	Grant Applications	2005 – present

System	Data Domain	Date Range
	Grant Approvals	2005 – present
	Spending by program, time & application	2005 – present
FRIS	Reimbursable Program Budgets	1995 – present
	Reimbursement Claims	1995 – present
	Vouchers/payments	1995 – present
ECS	Teacher demographics	1971 – present
	Subject taught	1971 – present
	Salaries of individual teachers	1971 – present
	Teacher certificates	1958 – present
	Professional development tracking	1999 – present
Annual Financial Report	District financial data elements	1996 – present

8.8. External Data Loads

Throughout ISBE, there are data elements that are external to ISBE systems which staff reference to compile reports or complete analysis. With the development of the data warehouse, many of these external data elements could be loaded into the data warehouse to "add value" to the ISBE reporting capabilities. For example, tax rate tables from the IRS could be loaded into the Data Warehouse to provide an added dimension to school financing analysis.

8.9. Security, Privacy and Confidentiality

Data warehouse privacy and confidentiality are dictated by state and federal laws and regulations. FERPA is a driving force in determining appropriate use, access and distribution of information from the data warehouse. To avoid conflicts with FERPA requirements ISBE can take one of two approaches to the storing of student-level data in the warehouse.

Student-level data can be brought into the data warehouse in its final form complete with personally identifying information (this may include name, social security number, and address). This level of data should be encrypted and made available only to a small number of authorized users. Alternately, as part of the ETL process, personally identifying student information including student names, social security numbers and street addresses can be stripped from each student record. Student records would continue to be identified by each student's unique student ID.

Both approaches have significant benefits and drawbacks. By stripping personally identifying information from student records ISBE's privacy, confidentiality, and training burdens are significantly eased because the potential for accidental or intentional abuse is reduced. However, the level of research, analysis and overall data "usability" is constrained. Inclusion of detailed student data and identifiers broadens the use of the data warehouse for analysis (e.g. trend analysis based on geographic location can be conducted). Should ISBE determine it wishes to

include this level of personally identifying information, then it will need to invest in the resources needed to enforce a stringent privacy, confidentiality and training program to ensure the appropriate use of the data warehouse. For the purposes of describing data access layers, data models and data flow diagrams, this feasibility study assumes the data warehouse will contain encrypted, personally identifiable information.

Within the warehouse, assigned student identifiers will be associated with individual student records. Teacher records will be associated with both Social Security numbers and unique teacher identifiers (assigned by the teacher identification system as the system comes online). Generally, publicly available reports and queries compiled from student-level data will be displayed in aggregated form with values below the minimal cell-size threshold masked to ensure student confidentiality.

Access to the data warehouse information should fall into four logical access layers:

1. A public layer that is accessible through ISBE's website that consists of pre-defined aggregate reports and are loaded to the ISEB website as HTML web pages and OLAP Public reports that are refreshed on a daily, weekly, monthly and/or quarterly basis.
2. An analytical layer that provides multiple OLAP layers (cubes) accessible only to authorized ISBE and District/School staff with OLAP multidimensional data cubes updated from the ISBE Transactional Layer information, OLAP information dimensioned, aggregated, stratified and stored into the OLAP layer, OLAP delivers requested information to spreadsheets or tabular reports for further data manipulation and graphing, OLAP information aggregated and de-personalized protecting confidentiality of consumer information, OLAP data manipulation is drag and drop and drill-downs
3. A transaction layer containing detailed and de-personalized information (names, addresses, and social security numbers have been removed), query-able by Reports Wizard(s).
4. A private layer containing data elements only available to a limited number of ISBE staff with granted access and authorization to view student-level data.

8.10. Cultural and Organizational Changes

Cultural and institutional shifts must occur to ensure the long-term viability and value of a data warehouse.

A primary institutional shift will be communicating the increased need to focus on quality of data and the impact quality issues have on ISBE and consequently the data warehouse. All levels of staff throughout ISBE, school secretaries completing data entry tasks, as well as managers within the districts and regions need to have an understanding of not only the importance of data quality but understand how the data will be used, and why it needs to be accurate and timely. This increased focus on data quality can be achieved through a well-planned training program that has buy-in at all levels.

Secondly, a data stewardship role needs to be introduced. A data steward acts as a central policy and oversight body to assure the quality of data that is introduced into ISBE's data collection systems, as well as providing a consistent policy for how collected information is used and distributed. This function may be fulfilled via a number of approaches including the designation of a Chief Information Officer (CIO), the formation of a Data Oversight Board, or the introduction of a data management division. Typical tasks conducted by the oversight function should include:

1. Create common definitions for data elements, aggregations and calculations
2. Evaluation of current data collections and identification of areas where data quality is suspect
3. Ongoing training to ensure data entry is conducted properly
4. Evaluation of new data sources for inclusion in the data warehouse
5. Set and enforce policies for the use and distribution of data to ensure all state and federal privacy and confidentiality constraints are met
6. Serve as a knowledge expert on the data warehouse and many of its source systems.

Training will be key to ensuring widespread usage of the data warehouse. Training will vary by user type but should include basic training for staff on accessing the data warehouse and using available query and reporting interfaces. Staff responsible for data entry, particularly at the district and school levels, will need to receive ongoing training on the correct definitions of data, as well as appropriate data entry procedures. Technical staff will need ongoing training on initiating, debugging and enhancing the ETL processes for populating the data warehouse.

8.11. Impact on Existing Data Collection Reporting Systems

Twelve ISBE data collection systems were identified as primary systems that will be added to the Data Warehouse in a phased implementation process. The development of the Student Information System and the consolidation of other data collection processes have allowed ISBE to limit the number of systems and narrow the number of systems that will need to populate the data warehouse. For example, with the completion of the Student Information System, ISBE will be able to eliminate the following data collection systems:

- Fall Housing Report
- Annual Report Card
- Performance Management System

Changes to the Special Education Approval and Reimbursement System (SEARS), that included the integration of the Illinois Purchased Care Review Board (PCRB) system into SEARS will allow for the discontinuation of the PCRB. Also, information now collected by SEARS parallels the data collected in the Funding and Child Tracking System (FACTS). Therefore, it will not be necessary for FACTS and the PCRB to feed into the Data Warehouse. Only data from SEARS will be

necessary to be pulled into the data warehouse.

Some information systems within ISBE, like the Professional Development Provider System (PDP), would provide no added value to the data warehouse; therefore, it is not necessary to include data from PDP in the Data Warehouse.

Finally, it is important to note that some data collection systems that are not stored in a database format like the Bilingual Annual Student Report will not populate the Data Warehouse. Likewise, there are other types of data collection processes like the gathering the Facilities and Inventory information at the school and district level that is currently not collected in an electronic format. This system and others like it will have to be built and later identified whether the data should be included into the Data Warehouse in the future.

8.12. Hardware and Software Recommendations

To support the data warehouse and successfully host the data warehouse within ISBE's infrastructure, it is recommended that ISBE use a separate server for each of the following functions:

- ETL
- Operational Data Store (ODS)/Staging Area
- Data warehouse database server
- OLAP cubes and interfaces
- Reporting/Querying
- Backup and recovery

Some solution offerings may support the consolidation of OLAP and reporting functionality to a singe server. Because much of the software applicable to a data warehouse solution is sold on a CPU basis it is recommended that ISBE purchase servers with 2 to 3 high-power processors rather than a server with a greater number of lower powered processors. Where feasible, ISBE may opt to have a fewer number of physical servers in favor of a single "box" containing multiple virtual servers. With the implementation of a data warehouse, ISBE will need to augment its backup capabilities. Based on ISBE's desire to move away from an IBM blade server configuration, due to high upfront and continuing maintenance costs, it is suggested that the data warehouse solution move to a lower cost blade solution.

Software to support the data warehouse includes both infrastructure (e.g. operating systems, a relational database) software and data warehouse specific software. Data warehouse specific software includes the software to operate the ETL, OLAP, and Reporting tools.

The current market includes a variety of product offerings that can address ISBE's needs for a data warehouse. While it is important not to eliminate products as this point, ISBE would benefit by leveraging existing technical infrastructure and software. By requiring the data warehouse

database be Microsoft SQL Server and using Microsoft Report Services and Crystal Reports, ISBE will be able to leverage existing systems and technical expertise.

Because the data warehouse does not contain transactional data it does not need to be backed as frequently as ISBE's transactional systems. Backups can be timed to coincide with updates to the data warehouse. Recovery procedures should be the same for the data warehouse as for the other systems maintained by ISBE (Refer to the ISBE's Disaster Recovery Plan for more information **Disaster Recovery**). The backup and recovery infrastructure will need to be augmented to meet the additional capacity requirements the data warehouse will incur.

8.13. Precursor Tasks

Before ISBE can begin moving towards implementing a data warehouse solution a number of pre-cursor tasks must be completed. ISBE must, at a minimum:

1. Develop and implement an identifier system to replace the current CDS mainframe implementation
2. Determine the level of personally identifiable student information to be stored in the data warehouse
3. Complete phase II of the SIS
4. Determine and implement a data stewardship function
5. Define and document privacy and confidentiality policies

Each of these areas is discussed in greater detail in the **Next Steps** section of this document.

8.14. Solution Options

ISBE may want to consider several approaches to implementing a data warehouse solution including:

- A commercial, off-the-shelf (COTS) product
- A transfer from another state or agency
- A custom development project
- An a la carte approach, where the data warehouse, ETL, OLAP and Reporting components are implemented separately

A COTS solution has the potential to bring ISBE's data warehouse online relatively quickly. Implementing a COTS solution that has been implemented in multiple other environments increases the likelihood of a successful implementation. However, existing solutions are very costly and will necessitate ongoing licensing agreements and fees over the life of the data warehouse. Also, because of the proprietary nature of most data warehouse COTS solutions, ISBE may need ongoing vendor support when the data warehouse is modified or expanded.

ISBE may wish to evaluate data warehouse implementations in other states and agencies that have developed with federal funds. These solutions should be transferable to ISBE. Such an approach may reduce the cost of implementing a data warehouse as well as decrease the need for ongoing vendor support. The major obstacle to this approach may be in finding an existing solution compatible with ISBE technical architecture.

A custom development approach, possibly using vendors as either knowledge experts or for augmenting technical staff, may be a low-cost approach to implementing a data warehouse solution. The benefits of such an approach would be a “custom-fit” solution specific to ISBE’s needs. The intricacies and technical demands of a data warehouse implementation are substantial with a significantly higher risk of project failure than the other implementation options.

The last approach for ISBE to consider is that of an “a la carte” solution. This approach would allow ISBE to leverage its existing technologies that are compatible with a data warehouse solution. This would include the use of Microsoft SQL Server 2000 with Analysis Services as agency’s data warehouse database and OLAP services. Existing reporting software, including Microsoft Report Services and Crystal Reports, could be used for developing querying and reporting interfaces. With this type of implementation ISBE would still need to develop or contract for services in developing the database layouts and data domains, as well as ETL processes.

Any of the preceding approaches, given the correct combinations of time, skills, and cost, may result in a successful ISBE data warehouse implementation. In evaluating possible data warehouse solutions ISBE should remain open to all possible approaches. However, all other considerations being even, ISBE should look to approaches that leverage its existing infrastructure, staff and skills as described in the “a la carte” approach.

CHAPTER 9. DATA MODEL, DFD & SOLUTION ARCHITECTURE

9.1. Data Flow Diagrams

The following pages illustrate data flows for each of the identified source systems to the data warehouse. The data flows are segmented into 6 major areas. Each area represents a range of data, technology and implementation options – from the least compatible with a data warehouse solution to the most "ideal". Below are listed each of the six areas with their documented ranges, from least data warehouse compatible to most:

	Area	Least DW Compatible	Most DW Compatible
1	Data Granularity	Pre-aggregated data	Elemental data
2	Collection System	Paper-based	Real-time automatic
3	Integration Process	Manual key punch	Closed loop
4	Operational Data Store	Distributed files	Integrated Relational
5	Staging/Verification Database	Not verified	Verified
6	Integrated Data Warehouse	3rd Normal form	Conformed dimensional

Definitions of terms used in the data flow diagrams include:

Data Granularity - The level of Detail represented by the data. For example: fine grain data may be actual transactions and course grain data may be the total number of transactions for a month.

Pre-aggregated - The numbers reported are summarized prior to submission. For example: the sum of all transactions for a month may be collected rather than each of the individual transactions.

File/form Based This type of collection system provides better data integrity than paper based systems, by providing real time edit checks, but still requires manual effort and allows opportunity for error.

Elemental Detail - The lowest level of transaction detail. For example, this your grocery store itemized receipt contains elemental detail.

Collection System - Process by which data are captured from it's source.

Paper-based - Data are collected via paper based forms which require transcription effort and potentially introduce transcription errors.

3rd party - Data collected by an agency not under the direct control of either a district or ISBE

Real time automatic - Data that is collected in an automated way that flows the data through to the collection process as soon as it is captured at the source.

Integration Process - The Integration process matches data from various domains along common dimensions. For example: to integrate student attendance data with student assessments, the common students must first be referenced uniquely. This can also be referred to as managing referential integrity.

Manual key-punch - Reference checking is done by people either based on experience or heuristics.

File consolidation - Files are matched to one another used semi-automated processes such as the sorting functions of Excel, but as a manual process, often errors must be corrected during the process and are often difficult to trace or permanently correct.

Closed loop - The data integration process is fully automated and any errors discovered through the process are immediately fed back to the source for correction. As a result of this process, data quality continually improves.

Operational data store - A preliminary point of data collection that mimics the structure of the source data. These structures provide a level of separation from the source systems that can allow sources to be replaced without adversely impacting the overall system. Although these structures consolidate much of the required data into the same location, they typically to not support flexible comprehensive analysis, as the structures are optimized for transactions, not reporting.

Distributed files - Data are collected in non-integrated file based systems where files provide much less functionality than relational tables.

Centralized files - Data are collected in files that reside in a single place, such as a mainframe server.

Non-integrated relational - The data may reside in powerful relational database management systems, but the function of the system may be limited because the structure was not designed to take advantage of the integrating power of the database management system.

Integrated relational - This type of system places the data in structure that supports integration and in a powerful relational database that can provide all of the necessary function to take advantage of this structure.

Staging/verification database - This is the process of providing feedback to the person responsible for the data that allows that person to confidently verify that the data represent reality.

Non-verified - There is no verification step in the process at all.

Verified by source - Data are verified by the person most responsible for it.

Integrated Data Warehouse - A comprehensive data structure designed for reporting and analysis.

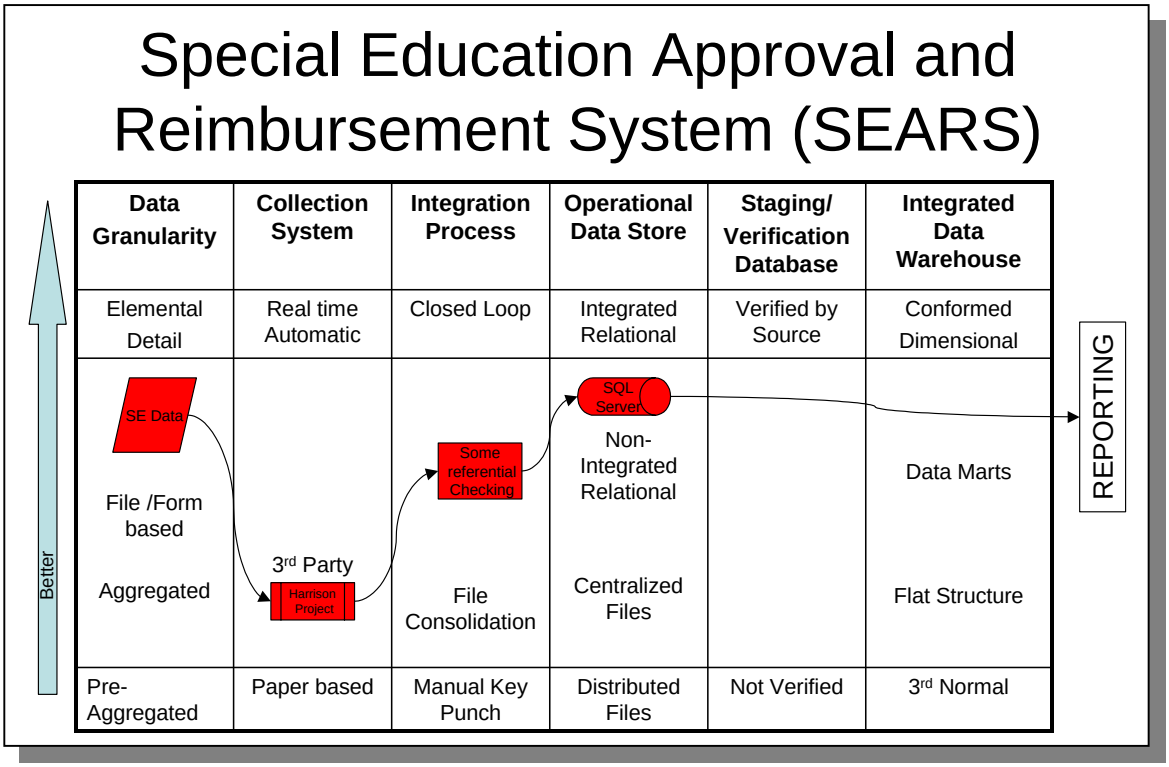
3rd normal - structure more suitable for efficient execution of transactions as it minimizes the duplication of data, but increases the number of tables that would be involved in the satisfaction of a report request.

Flat structure - A wide file structure that allows report requests to be answered by querying a single table, but limits the scope of questions that can be answered.

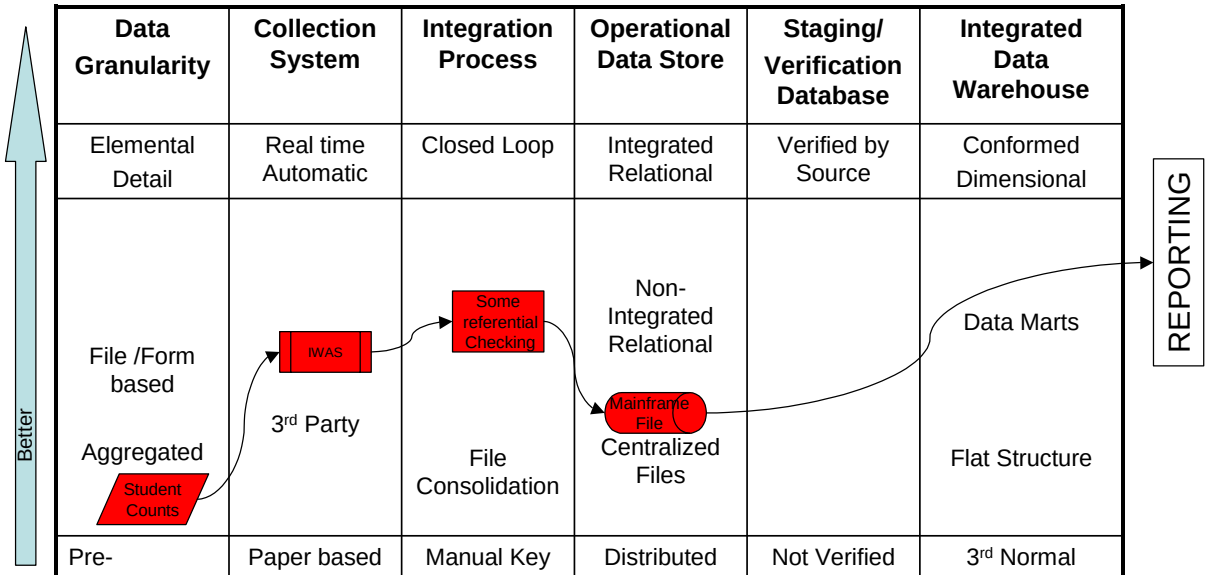
Data marts - limited scope data structures that are designed to answer known questions.

Conformed dimensional - A conformed dimensional structure is designed for efficient reporting and to provide the ability to report on any valid data relationship without knowing in advance that the relationship would be desired as part of an analysis.

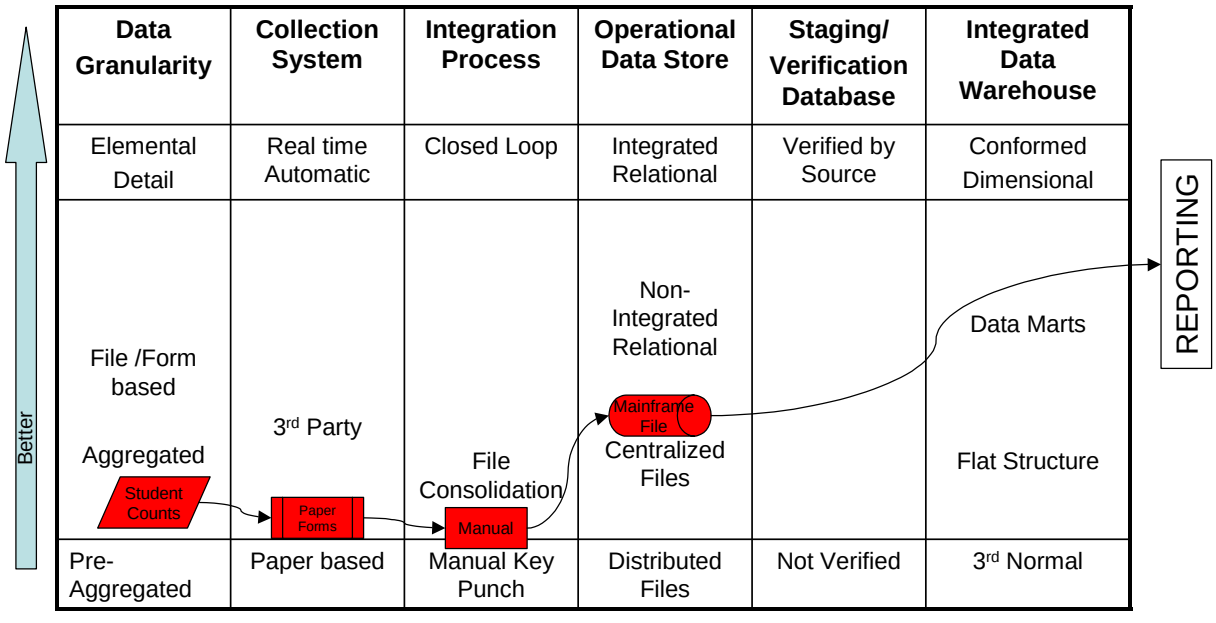
Each data flow identifies where the system currently maps regarding the upper and lower thresholds. This mapping can be used as a gauge for the source system's overall data reliability and quality. Over time ISBE will want to move each of these systems towards the upper end of the spectrum.



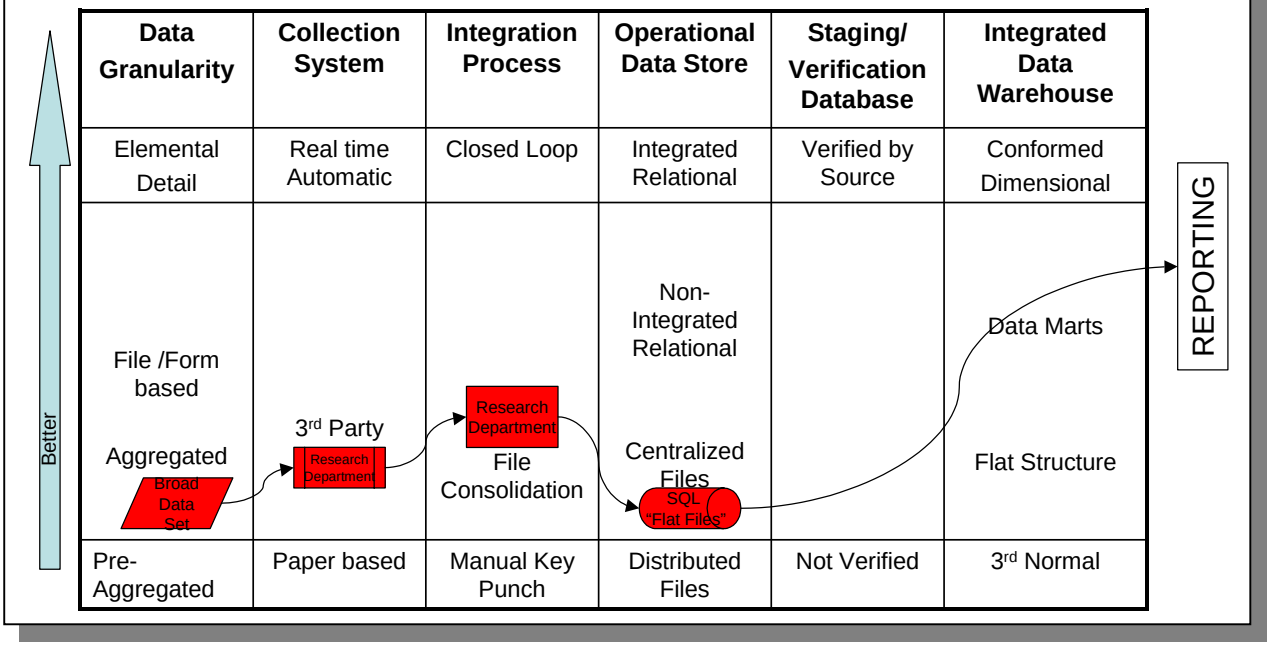
End Of Year/ Fall Housing



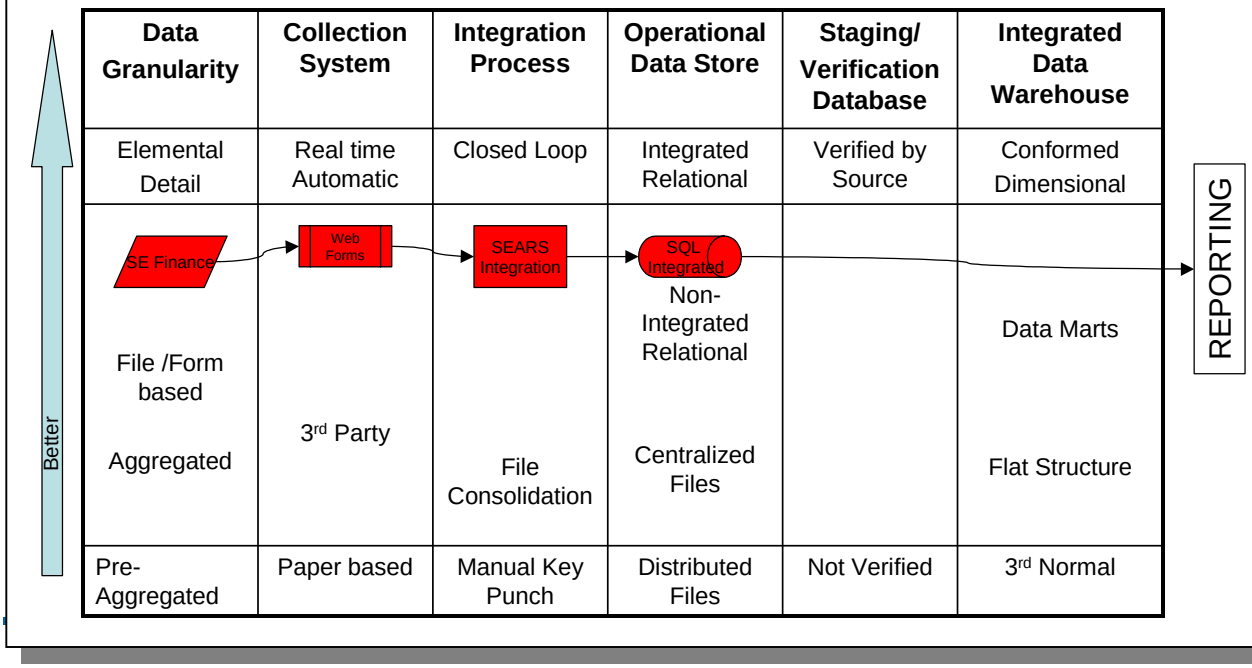
Public Registration, Enrollment and Staff Report (Non Public Fall Housing)



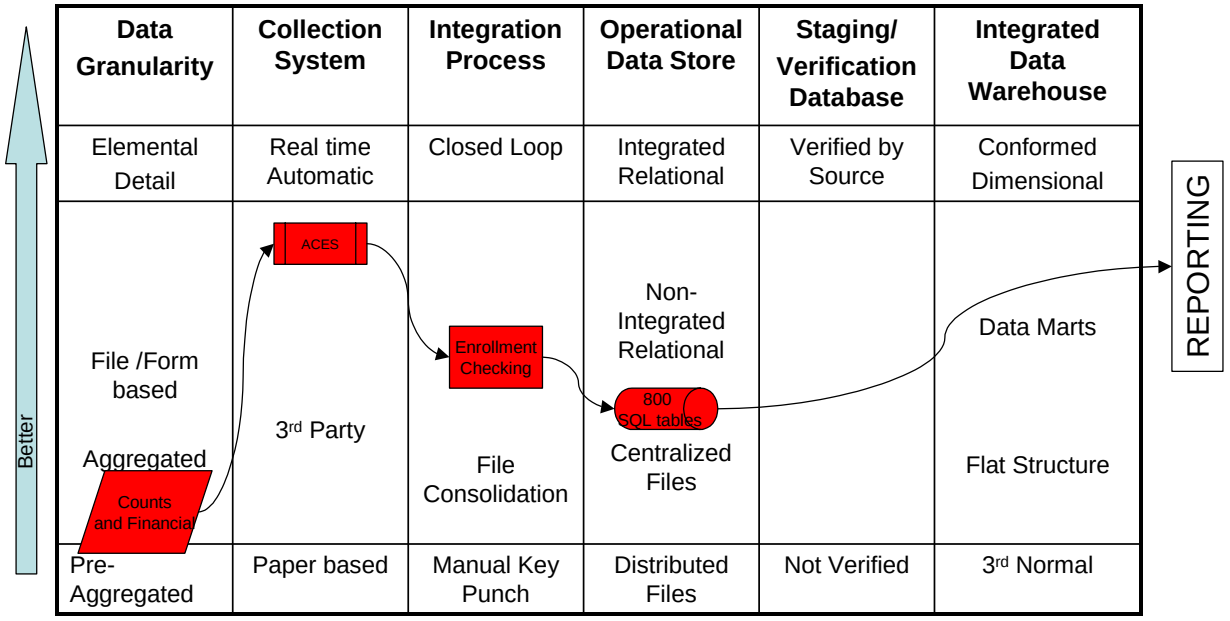
IL State Report Card



Illinois Purchased Care Review Board System (IPCRB)

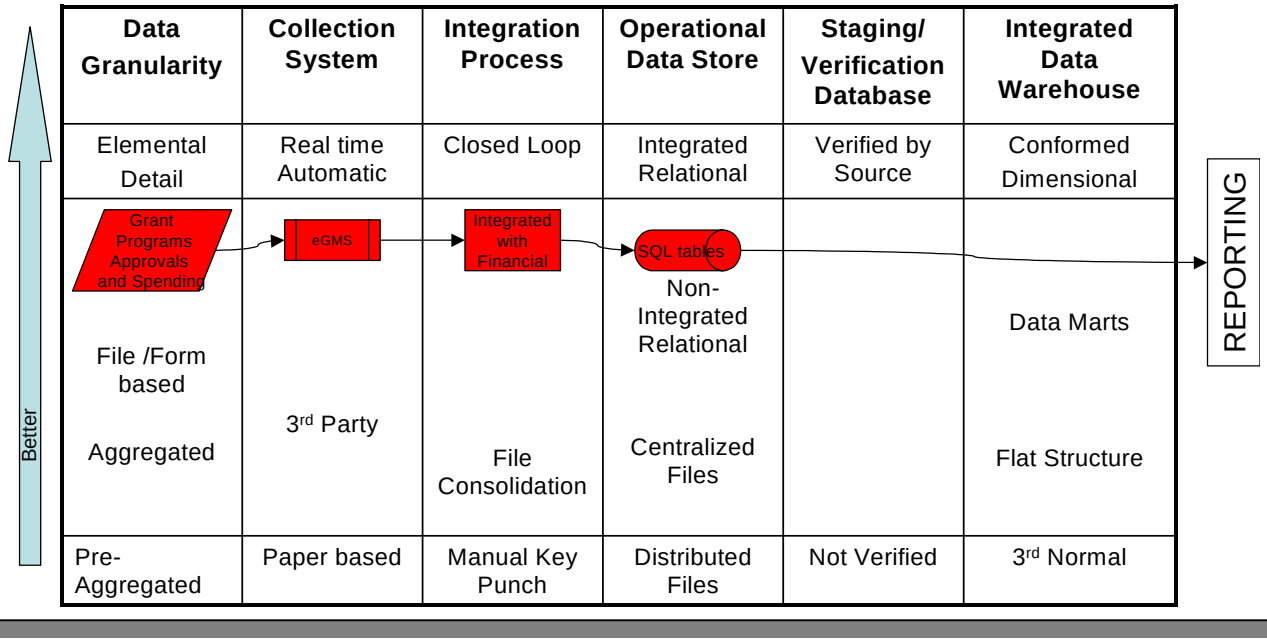


Child Nutrition System



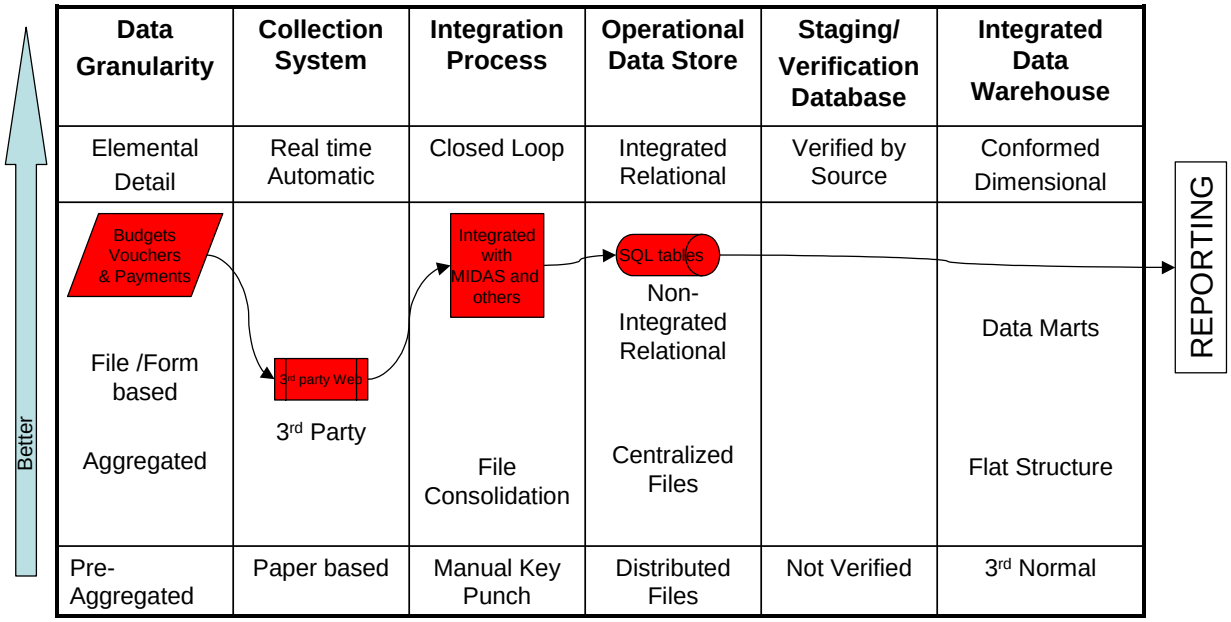
Data Granularity	Collection System	Integration Process	Operational Data Store	Staging/ Verification Database	Integrated Data Warehouse
Elemental Detail	Real time Automatic	Closed Loop	Integrated Relational	Verified by Source	Conformed Dimensional
File /Form based Aggregated	3rd Party	File Consolidation	Non-Integrated Relational Centralized Files		Data Marts Flat Structure
Pre-Aggregated	Paper based	Manual Key Punch	Distributed Files	Not Verified	3rd Normal

Grants management (eGrants)

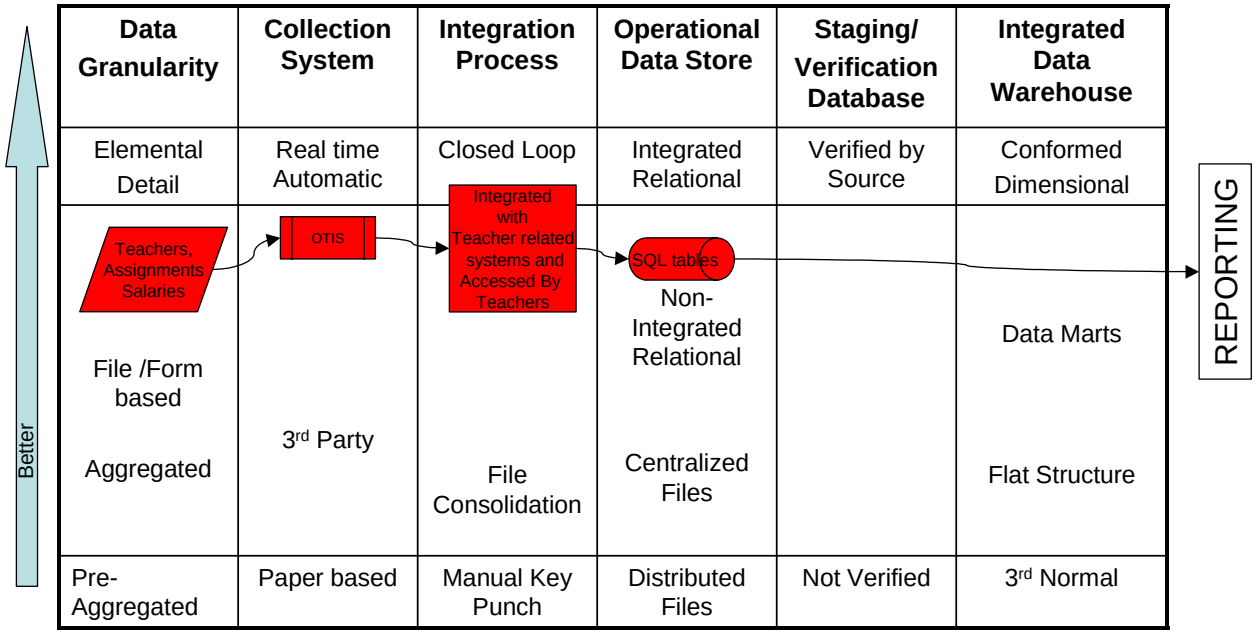


Data Granularity	Collection System	Integration Process	Operational Data Store	Staging/ Verification Database	Integrated Data Warehouse
Elemental Detail	Real time Automatic	Closed Loop	Integrated Relational	Verified by Source	Conformed Dimensional
File /Form based Aggregated	3rd Party	File Consolidation	Non-Integrated Relational Centralized Files		Data Marts Flat Structure
Pre-Aggregated	Paper based	Manual Key Punch	Distributed Files	Not Verified	3rd Normal

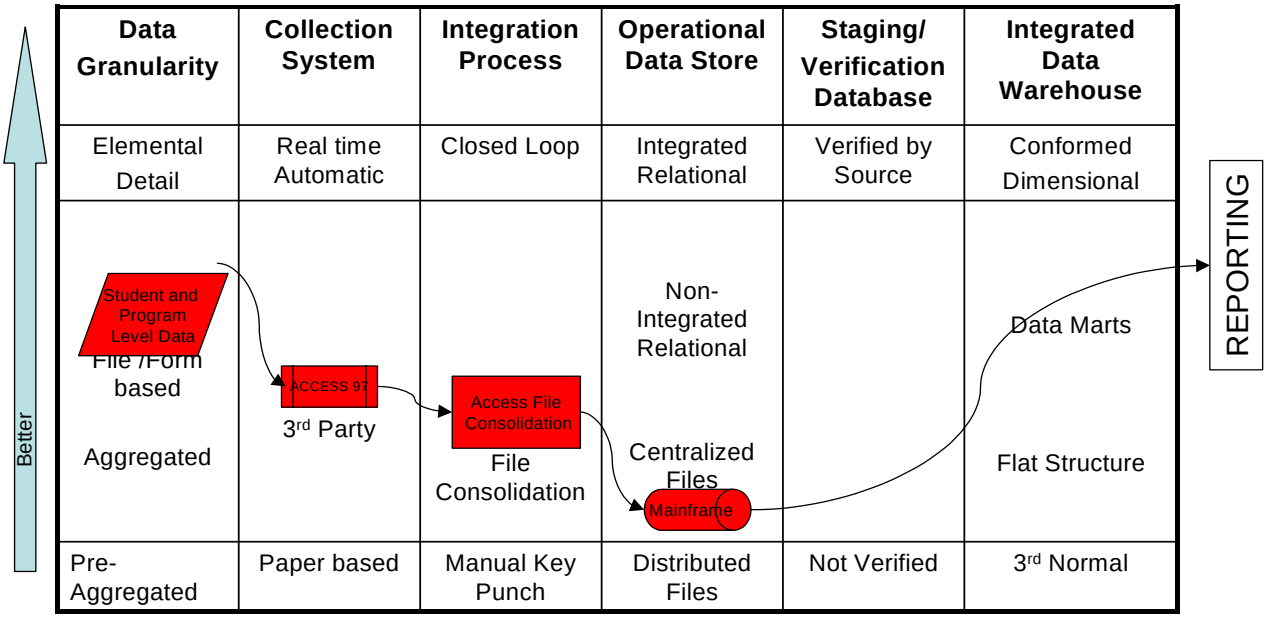
Financial Reimbursement Information System (FRIS)



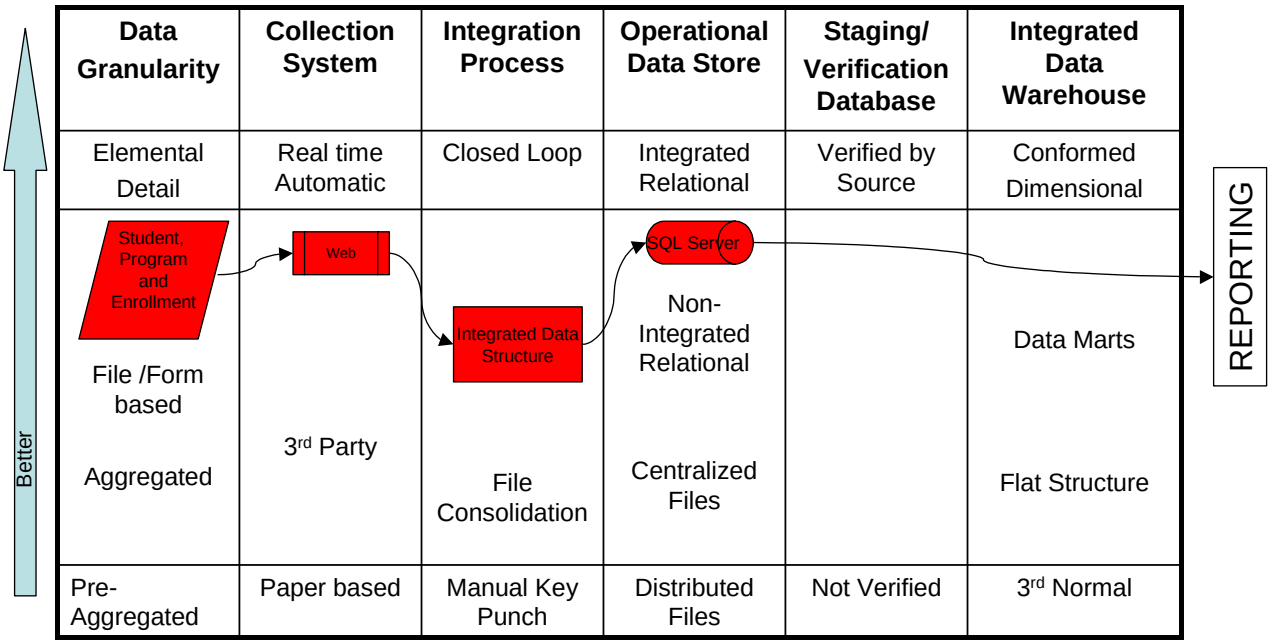
Educator Certification System (ECS)



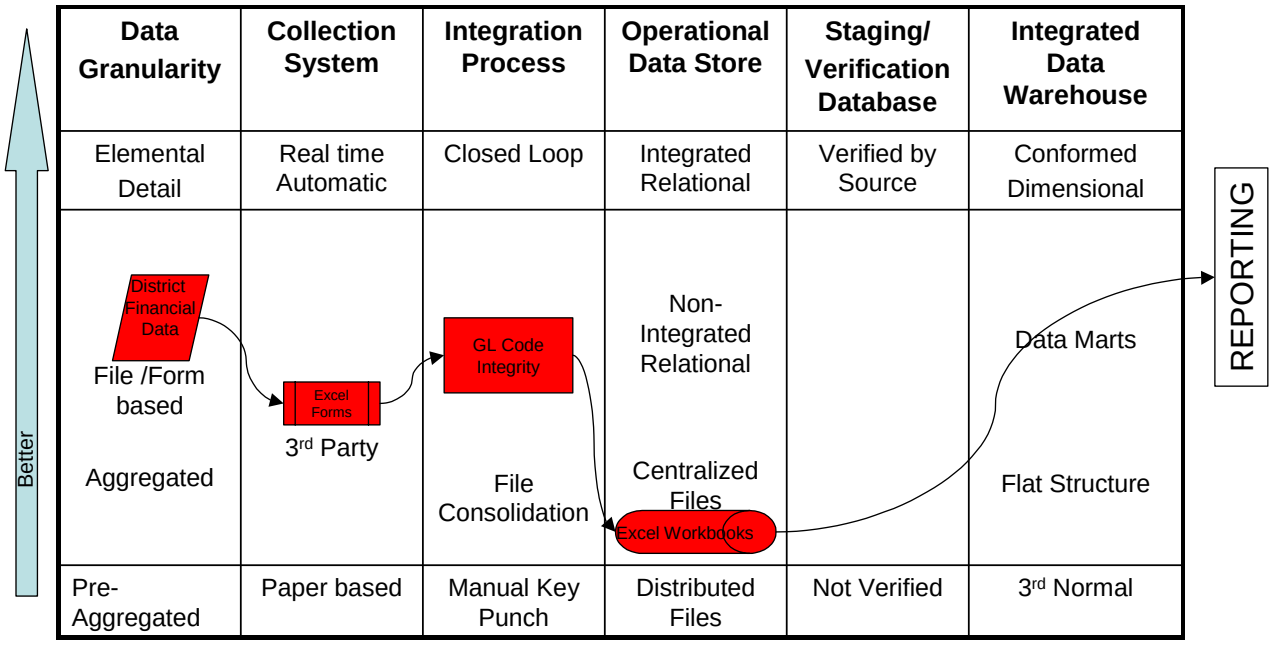
Illinois Student Information System (ISIS)



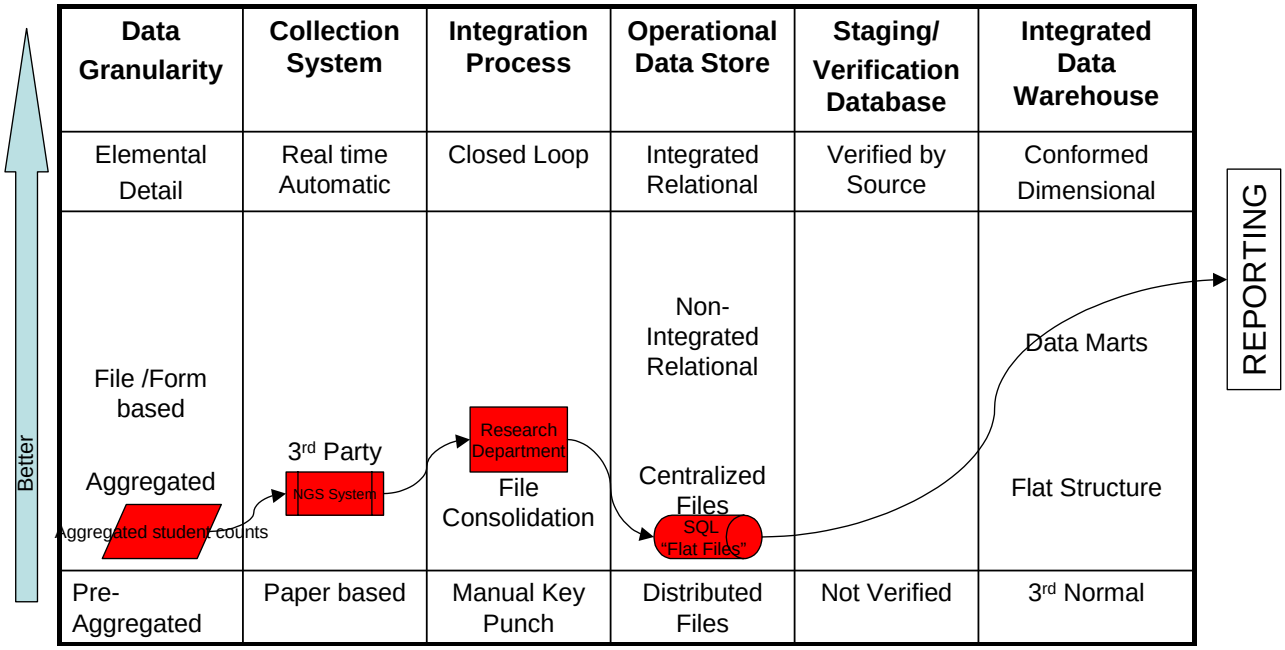
ISBE Student Information System (ISBE-SIS)



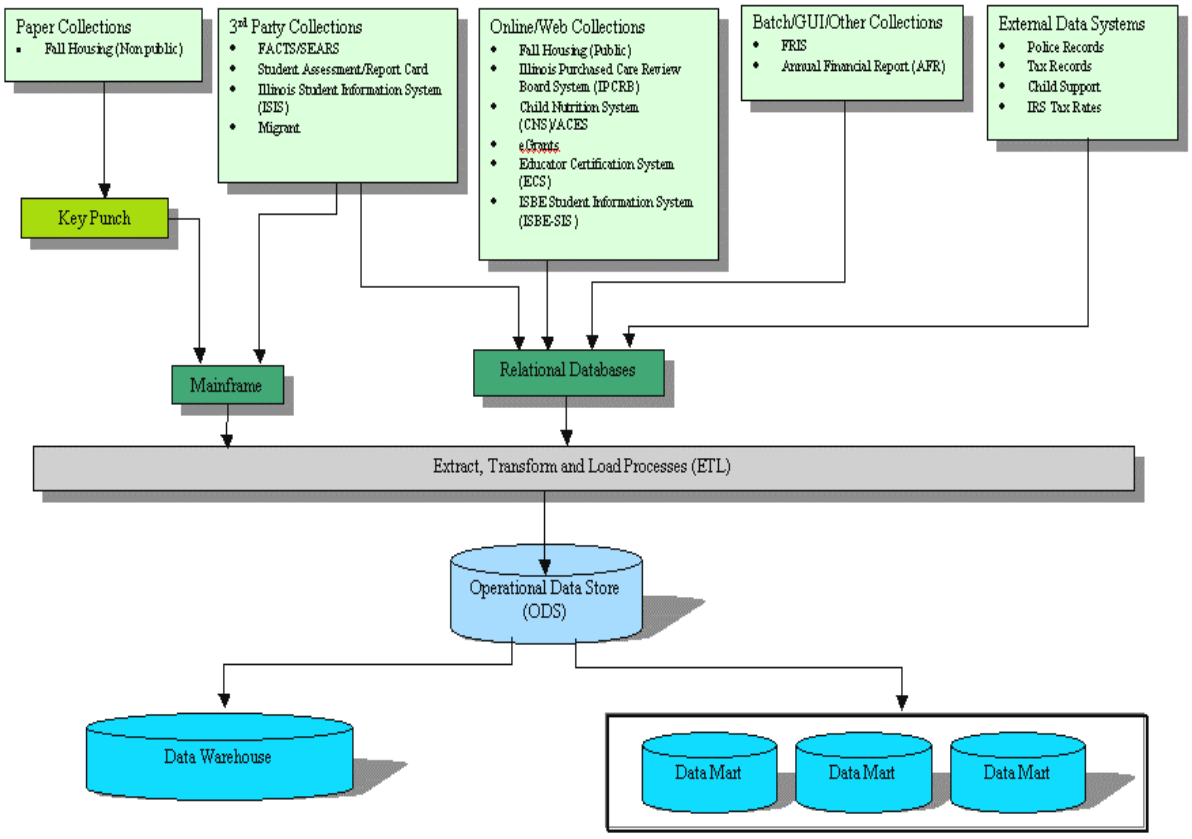
Annual Financial Report (AFR)



Migrant

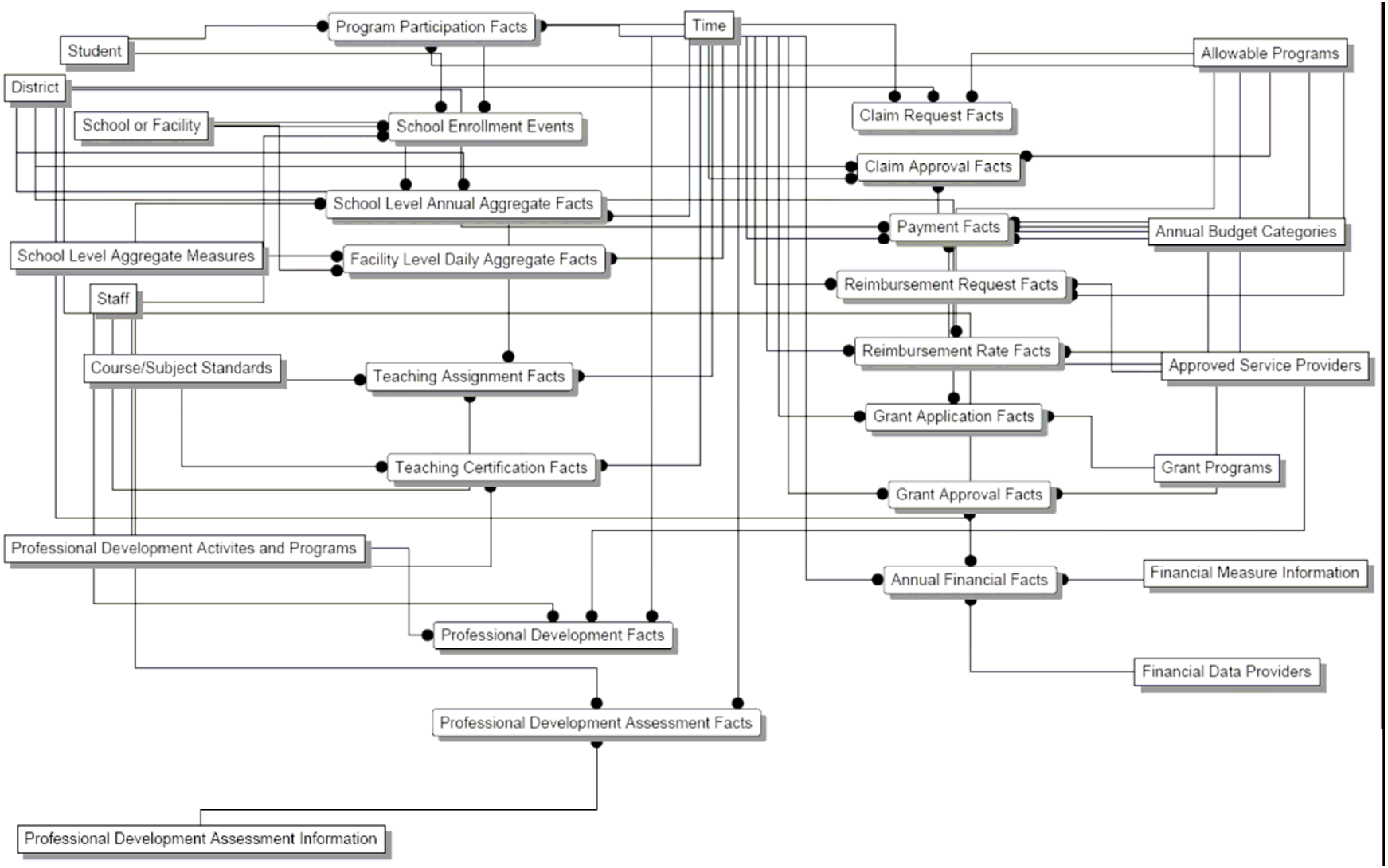


The following diagram illustrates the composite data flows from each of the identified source systems through the ETL process to the operational data store (ODS) to the data warehouse and data store databases.



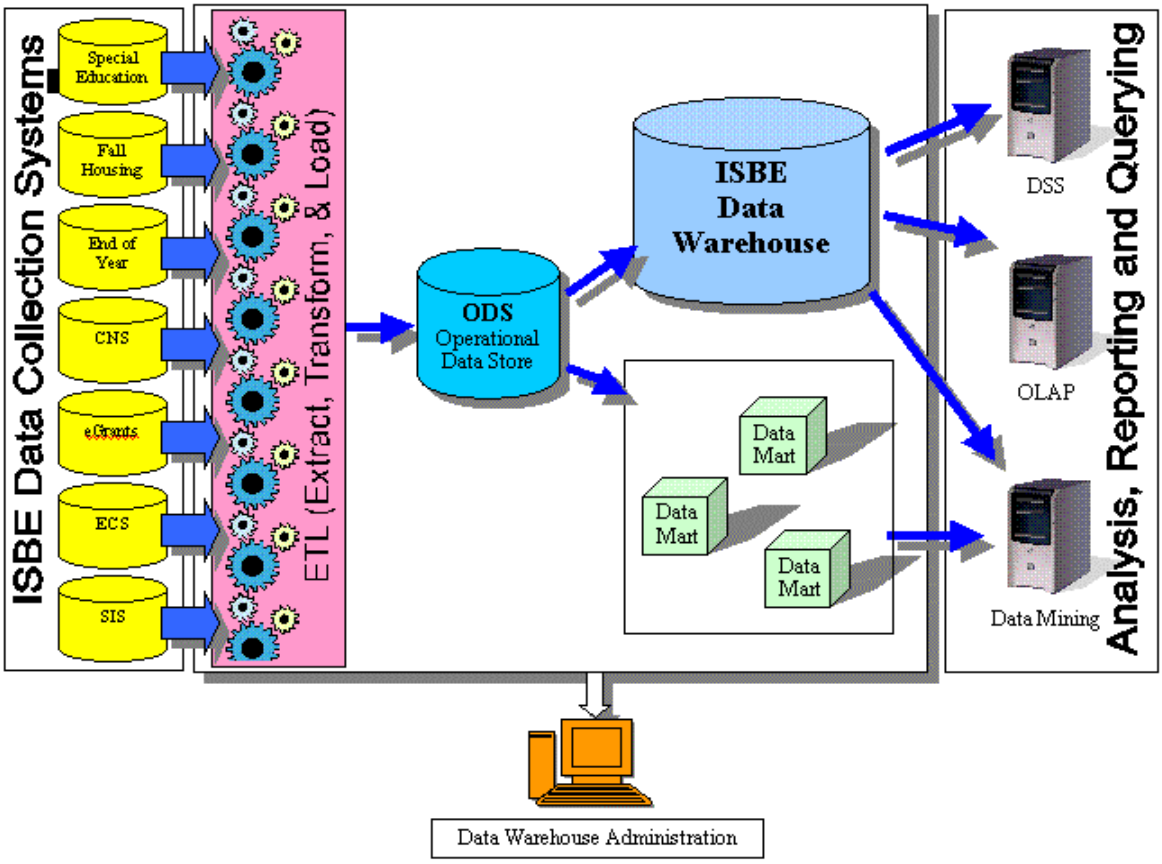
9.2. Data Model

The following data model is designed to cover the twelve data collections specified as data sources for ISBE's data warehouse solution in a dimensional data warehouse. It shows the logical relationships between the data entities at a conceptual level, but does not define the physical structure. This should allow ISBE to clearly communicate the data warehouse's conceptual framework to potential solution providers without locking them into a specific design or framework.


Program Participation Facts
Time
Student
Allowable Programs
District
Claim Request Facts
School or Facility
School Enrollment Events
Claim Approval Facts
School Level Annual Aggregate Facts
Payment Facts
Annual Budget Categories
School Level Aggregate Measures
Facility Level Daily Aggregate Facts
Reimbursement Request Facts
Staff
Reimbursement Rate Facts
Course/Subject Standards
Teaching Assignment Facts
Approved Service Providers
Grant Application Facts
Teaching Certification Facts
Grant Programs
Grant Approval Facts
Professional Development Activites and Programs
Annual Financial Facts
Financial Measure Information
Professional Development Facts
Financial Data Providers
Professional Development Assessment Facts
Professional Development Assessment Information

9.3. Solution Architecture

The following diagram illustrates the conceptual architecture for ISBE's data warehouse solution.



Physical implementations of this architecture will vary based on proposed solutions. At a minimum, ISBE should expect that there would be one or more servers to support each of the following:

- The ETL process
- The Online Data Store (ODS)
- The Data Warehouse
- OLAP cubes and interfaces
- One or more data marts
- The decision support system

Unfortunately, there are no specific guidelines defining the minimum hardware required for

supporting a data warehouse solution. Vendor proposed solutions should address the following server requirements:

1. Required Drive Space
    a. Data storage needs (tables, indexes, temp space)
    b. Data landing space (for inbound files and archives)
    c. Software needs (database, ETL and BI software)
    d. BI software space requirements
2. Memory
    a. Database needs (database working memory and dynamic memory)
    b. ETL tool requirements
    c. BI tool needs (report requirements)
3. Processor count and speed
    a. Each of the above listed servers will have different requirements. Processor requirements for the ETL server will vary based on the proposed solution. Some COTS ETL solutions do not leverage or support multiple processors. CPU counts will impact the number of licenses required for database licensing purposes. The general trend has moved towards fewer but higher, powered processors per server.
4. RAID
    a. Evaluate performance requirements versus RAID approach
    b. Generally, RAID 5 should be avoided due to its significant performance impact.
5. Scalability
    a. Hardware should be expandable
    b. Assume growth will exceed expectations

CHAPTER 10. COST ANALYSIS

Cost associated with the implementation of a data warehouse fall into the following areas:

1. Hardware
2. Software
3. Development
4. Project Management
5. Training
6. Annual maintenance

Cost estimates from this feasibility study were gathered from vendors willing to share pricing information, from best practices discussions and documentation with states implementing, or having implemented, a data warehouse, and publicly available information.

10.1. Hardware

ISBE should assume it will need 6 servers, one for each of the following functions:

- ETL
- Operational Data Store (ODS)/Staging Area
- Data warehouse database server
- OLAP cubes and interfaces
- Reporting/Querying
- Backup and recovery

Some solution offerings may support the consolidation of OLAP and reporting functionality to a singe server. Because much of the software applicable to a data warehouse solution is sold on a CPU basis it is recommended that ISBE purchase servers with 2 to 3 high-power processors rather than a server with a greater number of lower powered processors. Where feasible, ISBE may opt to have a fewer number of physical servers in favor of a single “box” containing multiple virtual servers. With the implementation of a data warehouse, ISBE will need to augment its backup capabilities. Based on ISBE’s desire to move away from an IBM blade server configuration, due to high upfront and continuing maintenance costs, it is suggested that the data warehouse solution move to a lower cost blade solution.

The following cost estimates are based on recent pricing provided to ISBE by Dell and Gateway and include a blade Dell configuration using an EMC storage array in lieu of individual, large drives for each of the servers and a Gateway server and SAN hardware for backup purposes. The capacity of the EMC storage array should allow it to be used by other systems outside of the data

warehouse, or conversely, the data warehouse may be able utilize existing an EMC storage array for the purposes of spreading the cost of the array across multiple systems. The below pricing model attributes the full price of the storage array to the data warehouse.

Item	Quantity	Price	Cost
Blade Enclosure	1	1,200	1,200
Power Edge 1855 with 2 Intel Xeon dual core processors	5	3,200	16,000
Software, Accessories and Server 1-year Maintenance agreement	1	22,500	22,500
Dell EMC CX500 Storage Array	1	160,000	160,000
Gateway 9415R Server	1	5,200	5,200
Gateway 840 Serial ATA SAN	1	7,000	7,000
Total			211,900

10.2. Software

Software costs fall into two main categories: infrastructure (e.g. operating systems, a relational database) and data warehouse specific software. Infrastructure software costs (Microsoft Windows Server and SQL Server) are based on pricing provided to ISBE by Microsoft. Data warehouse-specific pricing is based on a range of cost estimates and sources and will vary widely based on ISBE's preferred solution. A number of data warehouse-specific solutions are priced based on the number of students managed by the solution rather than by fixed price. The cost estimates below include logic for scaling up solutions provided in other states based on ISBE's student population.

Item	Quantity	Estimated Cost	
		Low	High
Microsoft Windows Server	6 @$350 each	2,100	2,100
Microsoft SQL Server	1	400	400
ETL	1	250,000	1,000,000
OLAP	1	250,000	1,000,000
Reporting	1	0	200,000
Total		502,500	2,202,500

10.3. Development

The majority of the costs associated with the implementation of a data warehouse are in development. Commercial off the shelf software (COTS) solutions require extensive setup and customization. The “rule of thumb” for estimating ETL development costs is the initial cost of the tool multiplied by a factor of 7. OLAP and reporting tools trend towards a one-to-one ratio (initial cost of tool x 1). As with the tool pricing discussed above, some development costs are based on total student counts and have been estimated by scaling know state deployments to ISBE’s student population. The below estimates do not include costs associated with the configuration of ISBE’s servers or backup/recovery software (including installation of operating systems, web services, or server administration). For the purposes of cost estimation, the below figures assume ISBE use vendors for 100% of development and customization. Should ISBE determine its data warehouse solution will use internal staff for some (of all) of development these figures should be adjusted accordingly.

Item	Estimated Cost	
	Low	High
ETL	4,900,000	7,000,000
OLAP	250,000	1,000,000
Reporting	50,000	200,000
Total	5,200,000	8,200,000

10.4. Project Management

By their nature, data warehouse implementations require a high degree of project management. Project manager responsibilities typically include overall project oversight including management of development staff, coordination with agency and district personnel, project scheduling and coordination and change control. Project management costs are estimated as between 25% of total development hours.

Item	Estimated Cost	
	Low	High
Project Management	1,300,000	2,050,000

10.5. Training

An implemented data warehouse solution requires ongoing management and support. Quantifiable training costs can be estimated for providing ISBE staff with the skills to setup, monitor and modify ETL processes, create and modify OLAP cubes, and create and modify reports. Most training will be tool-specific and will be provided by the tool vendor, although mature tools with a wide user base may have training offerings available through third-party companies. Training costs are calculated at $1,200 day with the low estimate based on total training of approximately 21 days and the high estimate for approximately 125 days.

Item	Estimated Cost	
	Low	High
Tool training	25,000	150,000

10.6. Annual Recurring Maintenance Costs

Purchased software will require recurring annual maintenance. Costs are estimated at 25% of the initial product cost.

Item	Estimated Cost	
	Low	High
ETL	735,000	1,050,000
OLAP	37,500	150,000
Reporting	7,500	30,000
Total	780,000	1,230,000

10.7. Total Estimated Pricing

The below estimates are for the total cost of development for an ISBE data warehouse implementation. The costs are based on the assumption ISBE will pursue a COTS-based approach with vendor supplied development and customization. Should ISBE pursue a custom development approach for one or more of the data warehouse components it can assume that its upfront product purchasing costs will be lower but that total development hours will increase. The cost of ISBE staff in development and implementation of a data warehouse solution are NOT included in this cost estimate. Annual maintenance is assumed included as part of the initial implementation is not included within this total.

Item	Estimated Cost	
	Low	High
Hardware	211,900	211,900
Software	502,500	2,202,500
Development	5,200,000	8,200,000
Project Management	1,300,000	2,050,000
Training (1 year)	25,000	150,000
Total	7,239,400	12,814,400

10.8. Budgeting Considerations

For budgeting purposes this section assumes a phased data warehouse implementation over three years. Upfront costs for hardware and software are assumed in the first year with development costs, training, and annual maintenance spread across years two and three. Training costs are assumed two halve each year.

	Year					
	1		2		3	
Cost	Low	High	Low	High	Low	High
Hardware	211,900	211,900				
Software	502,500	2,202,500				
Development	1,733,333	2,733,333	1,733,333	2,733,333	1,733,333	2,733,333
Project Management	433,333	683,333	433,333	683,333	433,333	683,333
Training	25,000	150,000	12,500	75,000	6,250	37,500
Maintenance			780,000	1,230,000	780,000	1,230,000
Total	**2,906,066**	**5,981,066**	**2,959,166**	**4,721,666**	**2,952,916**	**4,684,166**

CHAPTER 11. NEXT STEPS

The following tasks have been identified as critical “pre-cursor” steps to successfully lay the groundwork for a data warehouse implementation.

11.1. Identify a Strong Project Sponsor

A data warehouse implementation in an incredibly daunting task. It will be resource-intensive, expensive, and a complex undertaking that may last for several years. To assure project success a strong project sponsor must be attached to the effort. Not only must this individual have a clear and committed vision for the data warehouse, but must have the political and organizational clout to clear organizational and technical hurtles that may impede progress.

11.2. Develop a Unique Identifier System

RCDT, the current mainframe-based identifier system, has been in use for well over 20 years. In its current state the system contains duplicate entries, entries for non-existent entities, and mis-assignments of IDs based on incorrect assignment of types. Additionally, the system lacks the ability to assign identifiers to vendors and other entities ISBE would like to track. Building a data warehouse system without replacing this system has been described by several at ISBE as “building a house on a foundation of sand”. ISBE is aware of the current system’s limitations and has begun investigating options for its replacement. This task will need to be completed, and the system implemented, so that a data warehouse can be designed to incorporate these unique identifiers.

11.3. Determine the level of personally identifiable student information

As has been described within the approach section, ISBE can choose to encrypt and include personally identifying student information within the data warehouse, or strip it out as part of the ETL process. This determination must be made before the data warehouse data schema and ETL processes can be designed.

11.4. Complete Phase II of the SIS

The SIS system will be one of the major feeds to the data warehouse system, with its continuing expansion a key to gathering valid, consistent data for the data warehouse. At a minimum the current phase of implementation needs to be completed before work can begin on the data warehouse. Initiating the data warehouse project before this phase is completed will necessitate the creation of ETL processes that will need to be frequently modified or discarded as the SIS is

expanded.

11.5. Determine and Implement a Data Stewardship Function

A data steward acts a central policy and oversight body to assure the quality of data that is introduced into ISBE's data collection systems, as well as providing a consistent policy for how collected information is used and distributed. This function may be fulfilled via a number of approaches including the designation of a Chief Information Officer (CIO), the formation of a Data Oversight Board, or the introduction of a data management division. Typical tasks conducted by the oversight function should include:

- Create common definitions for data elements, aggregations and calculations
- Evaluation of current data collections and identification of areas where data quality is suspect
- Ongoing training to ensure data entry is conducted properly
- Evaluation of new data sources for inclusion in the data warehouse
- Set and enforce policies for the use and distribution of data to ensure all state and federal privacy and confidentiality constraints are met
- Serve as a knowledge expert on the data warehouse and many of its source systems.

One of ISBE's goals for the data warehouse is heavy utilization by all segments of the education community. ISBE can expect that data contained within the data warehouse will be highly scrutinized. Georgia's Department of Education, in its best practices interview, stated that without implementing a data stewardship function they never would have moved forward with a data warehouse implementation.

11.6. Define and Document Privacy and Confidentiality Policies

FERPA is a driving force in determining appropriate use, access and distribution of information from the data warehouse. To avoid conflicts with FERPA requirements, as well as state and federal regulations, ISBE must clearly define its privacy and confidentiality policies. The policies must be clear, universally applied, and understood throughout the agency. Even after the implementation of a data warehouse ISBE must continually communicate and train ISBE staff, as well as districts and schools, in the agency's established policies.

APPENDIX A – GLOSSARY

Ad hoc Query: A question or query that is not pre-defined or planned.
Ad hoc Report: A report that is not pre-defined or established. A report that is developed using tools as needed, rather than one that was previously created.
Aggregate or Aggregated Data: Summarized data.
Approved schools: These are all private and parochial schools in Nebraska that provide elementary and secondary instruction to children of compulsory age. Non public schools would be classified as approved if they meet all of the requirements of Chapter 14, Regulations and Procedures for the Legal Operation of Approved Non Public Schools.
Buckley Amendment, The: See Family Education Rights and Privacy Act
Business Intelligence: Describes a set of concepts and methods to improve decision making by using fact-based support systems.
Business Rules: The policies, practices, procedures and decision processes of an organization.
Cell Size: The amount of data reported in a cell.
Cell Suppression: Controlling the data reported in a cell in order to protect the identity of individuals represented. Many organizations will provide a rule to suppress data if, for instance, there are less than a particular number of people to report.
Cell: A box into which a piece of data are entered. A spreadsheet is composed of rows and columns of cells.
Confidentiality: Guarantees that personal data will not be released that allows it to be tracked to an individual. ***Confidentiality*** refers to an agency's obligation not to disclose or transmit information about individual students to unauthorized parties.
Criterion-referenced test: A test that allows reviewers to make score interpretations in relation to a functional performance level, as distinguished from those interpretations made in relation to the performance of others.
Cross Tabulation: The simultaneous tabulation of two or more variables.
Dashboard: A display of information that enables individuals and businesses to access, analyze and present critical data graphically, usually in real-time. **See Portal**
Data Aggregator: An information provider or tool that gathers content from several sources and brings it together.
Data Architecture: The framework for organizing the planning and implementation of data resources.
Data Cleaning (AKA Data Cleansing or Data Scrubbing): The process of removing or correcting data previously introduced into a system.
Data Cubes: Consisting of dimensions and measures data cubes allow data to be viewed in multiple dimensions.
Data Dictionary: Definitions of data elements.
Data Driven Decision Making (D3M): The act of making decisions based on data, or information received.
Data Mart: A collection of databases, designed to help managers make strategic decisions about their business. Whereas a data warehouse (q.v.) combines databases across an entire enterprise, data marts are usually smaller and focus on a particular subject or department
Data Mining: The use of statistical and visualization techniques to uncover trends and relationships within massive databases.
Data Model: A collection of descriptions of data structures and their contained fields, together with the operations or functions that manipulate them.
Data Quality (also Quality Data): Accurate, timely, meaningful, and complete data.
Data warehouse: A collection of data designed to support management decision-making.
Decision Support System: The term refers to a computerized system that gathers and presents data from a wide range of sources, typically for business purposes.
Demographics: Data related to the characteristics of human populations.
Disaggregating: An analysis of data differentiated by subgroup or subcategory.
Disclosure means to permit access to, release, transfer, or otherwise communicate personally identifiable information contained in education records to any party, by any means, including oral, written, or electronic means.
Drill down: Allows a user to move between levels of data ranging from the most summarized (up) to the most detailed (down).
Extraction, Transformation, and Loading (ETL): The process of getting data from existing databases, transforming

these data into an interoperable (q.v.) format and loading data into a formal repository.
Family Educational Rights and Privacy Act (FERPA): is a federal law which mandates confidentiality of student records, while assuring parents access to the records. See 20 U.S.C. §1232g (AKA The Buckley Amendment)
Feasibility Study: Determining the needs of the organization, technical support necessary and overall cost of a system before developing it.
Formative Assessment: Allows staff to determine what part of a task a student knows and does not know.
Front end: The part of the software program first seen by the user, usually with icons or text links to the underlying functionality of the application.
Granularity: The level of detail at which information is viewed or described. The more "granular" the smaller bits of information are presented.
Graphic Reports: Presenting data using pictorial representations.
Graphical User Interface (GUI): A front end (q.v.) that uses pictorial representations of information in addition to straight text that enables the use of a mouse, or other pointing device, to move around the page.
Interface: The visible part of a computer system or database allowing the user to select items to view and enter information into the system.
Interoperable: Computer systems or databases that are Interoperable use a set of rules and definitions that enable these software programs to share information, even though the databases might be from different companies. Interoperable rules are platform independent and vendor neutral.
Item Analysis: Identifying the how many students within a population selected each answer to a specific question on an assessment.
Legacy System: An old hardware or software system which may be out-dated in some way, either based on obsolete hardware or using an older user interface (e.g., a character-based interface rather than a GUI). A system that may not be able to interact easily with other computer systems.
Local Education Agency (LEA): An LEA might be a district, county or other intermediate educational organization.
Longitudinal Study: A study that analyzes data for subjects who continue to participate over an extended period.
Metadata (or Meta Data): Data that describe the data contained in the database. Clear, accurate and precise definition of data to reduce confusion or misinterpretation.
Nonparametric Statistical Data Analysis: Data that does not have to conform to specific parameters or rules.
Normalization: The process of reducing a complex data structure into its simplest, most stable structure.
Norm-referenced test interpretation: A score interpretation based on a comparison of a test taker's performance to the performance of other people in a specified reference population.
On-line Analytical Processing (OLAP): A category of software tools that provides analysis of data stored in a database or data warehouse.
Online: The status of being connected to a computer or network or having access to info that is available through the use of a computer or network.
Open data Base Connectivity (ODBC): A set of functions developed by Microsoft that provides access to databases.
Parametric Statistical Data Analysis: Analysis of data that meets the parameters needed for higher-level tests of significance. Some of the key parameters are large sample sizes, normally distributed data, etc. (q.v. Nonparametric Statistical Data Analysis)
Personally identifiable information generally includes, but is not limited to: the student's name; the name of the student's parent/guardian or other family member; the address of the student or student's family; a personal identifier, such as the state student identifier; personal characteristics or other information that would make the student's identity easily traceable. A small set of this information will be essential for assigning identifiers and for identifying students who have transferred from another district within the state or who have returned to the state who already have identifiers.
Portal: A Website or service that offers a broad array of resources, such as e-mail, forums, search engines, and online shopping malls. The first Web portals were online services, such as AOL, which provided access to the Web; now most of the traditional search engines (e.g. Yahoo ®, Google ®, etc.) are Web portals, modified to attract and keep a larger audience.
Pre-defined Report: A report that has been previously developed. A "canned" report.
Privacy: The principle of protecting private information about people, especially in shared or collaborative systems, and of helping to keep people free of distractions. **Privacy** refers to an individual's right to freedom from intrusion due to disclosure of information without his or her consent.
Prototyping: When developing an enterprise-wide system, using a smaller version of that system to test how it will work.

Relational Database Management System (RDBMS): A type of database management system that stores data in the form of related tables.
Scalability: Refers to anything whose size can be changed. How a hardware or software system can adapt to increased demands.
School Indicators: The combination of data such as norm-referenced and standards-based assessments, dropout rates, completion rates, etc., that together provides information about the success of a school in fostering successful student learning.
Schools Interoperability Framework (SIF): An application product standard that allows a computer system to interact with other software functions.
Security: Protecting equipment, performance, and contents when using technology.
SIF: See Schools Interoperability Framework
Slice and Dice: Refers to the ability for end users to view multidimensional data by navigating interactively -- rotating and pivoting how the dimensions are displayed as rows and columns, and drilling down to lower levels of detail.
State Education Agency (SEA): Usually a state department of education.
Stovepipe System: A legacy system that cannot be upgraded without great difficulty and/or is not interoperable (q.v.) with other systems in the organization.
System Architecture: A description of the design and contents of a computer system, including a detailed inventory of current hardware, software and networking capabilities; a description of long-range plans and priorities for future purchases, and a plan for upgrading and/or replacing outdated equipment and software.
System Integration: The combining of two or more computer systems and or software packages enabling these systems to work together efficiently.
Tabular Reports: Reports formatted to look like a table.
Transactional System: An information system designed to store and record day-to-day business information. Systems developed for storing data, but not for analyzing that data.
Transparency: The use is obvious or intuitive.
User Interface: The way a user interacts with data or communicates with the system.

APPENDIX B – FERPA SUMMARY

Reprinted from the **www.ed.gov** Website

Family Educational Rights and Privacy Act (FERPA)

The Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 CFR Part 99) is a Federal law that protects the privacy of student education records. The law applies to all schools that receive funds under an applicable program of the U.S. Department of Education.

FERPA gives parents certain rights with respect to their children's education records. These rights transfer to the student when he or she reaches the age of 18 or attends a school beyond the high school level. Students to whom the rights have transferred are "eligible students."

- Parents or eligible students have the right to inspect and review the student's education records maintained by the school. Schools are not required to provide copies of records unless, for reasons such as great distance, it is impossible for parents or eligible students to review the records. Schools may charge a fee for copies.
- Parents or eligible students have the right to request that a school corrects records, which they believe to be inaccurate or misleading. If the school decides not to amend the record, the parent or eligible student then has the right to a formal hearing. After the hearing, if the school still decides not to amend the record, the parent or eligible student has the right to place a statement with the record setting forth his or her view about the contested information.
- Generally, schools must have written permission from the parent or eligible student in order to release any information from a student's education record. However, FERPA allows schools to disclose those records, without consent, to the following parties or under the following conditions (34 CFR § 99.31):
    - School officials with legitimate educational interest;
    - Other schools to which a student is transferring;
    - Specified officials for audit or evaluation purposes;
    - Appropriate parties in connection with financial aid to a student;
    - Organizations conducting certain studies for or on behalf of the school;
    - Accrediting organizations;
    - To comply with a judicial order or lawfully issued subpoena;
    - Appropriate officials in cases of health and safety emergencies; and
    - State and local authorities, within a juvenile justice system, pursuant to specific State law.

Schools may disclose, without consent, "directory" information such as a student's name, address, telephone number, date and place of birth, honors and awards, and dates of attendance. However, schools must tell parents and eligible students about directory information and allow parents and eligible students a reasonable amount of time to request that the school not disclose directory information about them. Schools must notify parents and eligible students annually of their rights under FERPA. The actual means of notification (special letter, inclusion in a PTA bulletin, student handbook, or newspaper article) is left to the discretion of each school.

INDEX

MTW SOLUTIONS, LLC

3425 Constitution Court, Suite 201
Jefferson City, MO 65109

Tel: 573-893-7997
Toll free: 800.669.9689
Fax: 573-893-6636

www.mtwsolutions.com