USOE Data Steward Analysis

In the August 21, 2006 USOE Data Group meeting the concept of agency data stewards was addressed, and the decision was made to move ahead with the concept. Within a week of this meeting a draft of a Data Steward Approval document was distributed (Appendix A). Given that it was a draft, two agency staff members (the author of this analysis and Randy Raphael) both of whom originally promoted the idea of data stewards, were surprised to lean that shortly thereafter one section had already appointed their data steward. While the concept of a data steward is very valid and the Data Steward Approval document is an important part of the overall process; other important definitions, descriptions and organizational function implications still need to be addressed. The following analysis attempts to address major components of the data steward initiative at the USOE in more detail. It is arranges in the sections: The Data Steward's Role, Data Steward as Data/Information Analyst, The Presentation Layer, Accountability and Next Steps.

The Data Steward's Role

An entire research project could be designed around the examination of definitions and descriptions of data stewards and their roles. A search of Google with the words: "role" or "definition" or "description" and "data steward", yields over 200,000 hits. See Appendix B for some examples of what can be found from conducting such a search.

Most of the focus of these descriptions of the roles, and organizational functioning of the examples in Appendix B focus on data steward roles that involve data identification, definitions, understanding, collection, business rules, quality and very importantly communications and liaison functions involving both LEAs, state and federal government as well as agency staff members. These communication functions cannot be over emphasized. Ignoring those leads to not only wasted effort and frustration but more importantly it results in poor quality data.

In order to adequately function as the data communicator or liaison who has expertise in a particular area of education data and hopefully with a good overall knowledge of most education data, the steward must understand the way in which data flows and is gathered, assembled and disseminated throughout the system. Appendices C and D contain high level flow charts or schematics of how data moves between the USOE, LEAs, the public and other governmental agencies at the state, national and local levels. The first schematic shows the overall flow while the second depicts the organization of the USOE data warehouse and how it interfaces to other data structures.

Until now the USOE Data Warehouse Group and Computer Services have provided many data steward functions. However, as data volume, complexity and demands grow, this ad hoc group and Computer Services' technical staff are struggling to manage their existing assignments and oversee all other aspects of the agency's second most important resource, its data.

While some members of the Data Warehouse Group could perform many data steward functions most do not have enough time to devote to the job while others lack the background and skills. While Computer Services' staff members are professionals at constructing and maintaining information technology systems and acting as custodians of important databases, user interfaces and business rules, they often lack the business knowledge to make policy decisions about the collection and use of data. Without continual and close consultation with data-aware individuals in the various agency sections programmers, IT managers and database administrators cannot be successful.

Data Steward as Data/Information Analyst

There is an important difference in the way the term data steward is generally being used and what the USOE seeks, as described in the Data Steward Approval from draft. While data stewards generally work at the metadata layer (where a thorough understanding of pertinent data and its business rules is necessary), the USOE also needs those skilled in the technical aspects of data retrieval (from existing data structures) and analysis. This position is generally called a data or information analyst. Such a person would be very involved in working with the data warehouse's so called presentation layer. This type of analyst, working with both the data steward and IT department (Computer Services), is ideally a partner in the construction of these presentation multi-faceted presentation layers.

What is being suggested by the Data Steward Approval form is the merging of these two skill sets. While feasible, this combination may prove to be problematic due to the specialized training needed to do the technical parts of the job. A Utah DHRM job classification of data analyst does exist. The Purpose and Distinguishing Characteristics; Examples of Tasks; Knowledge Skills and Abilities sections of the Utah DHRM job description for an Information Analyst follow.

Purpose and Distinguishing Characteristics

> Incumbents in this job perform at the working level, under general supervision. Incumbents typically perform activities which may include: the utilization of computer software programs to store data, retrieve stored data, analyze retrieved data, and arrange results in a format that effectively supports the organization's production activities; coordinate the use of information analysis related hardware, software, and data with the organization's employees and/or the public; interface with data processing users to automate applications and/or resolve application and operating problems; help customers identify better ways to utilize the system's capabilities; estimate project costs; assure deadlines are met; negotiate common standards for data formats; assess needs for data analysis related hardware, software, and data; and make recommendations to management for improving and/or maintaining effective information analysis support for the organization.

Examples of Tasks

(More specific information about the job can be found in the Purpose and Distinguishing Characteristics. This list contains tasks that are typically associated with the job. It is not all-inclusive and may vary from position to position. Hiring agencies may, depending on the specific nature of the position, modify these tasks and/or identify additional tasks, based on a current position analysis.)

- Adapts automated system(s) to accommodate special and complex agency needs; designs or enhances agency unique automated systems to interface with existing systems as needed.
- Monitors system performance to ensure proper operation and identify possible problem areas.
- Analyzes, summarizes and/or reviews data; reports findings, interprets results and/or makes recommendations.
- Identifies problems or potential problem areas and recommends solutions.
- Writes and/or uses computer programs to generate and analyze data.
- Other tasks as assigned.

Knowledge Skills and Abilities

(This list contains KSAs that are typically associated with the job. It is not all-inclusive and may vary from position to position. Hiring agencies may, depending on the specific nature of the position, modify these KSAs and/or identify additional KSAs, based on a current position analysis.

Knowledge of the Following Theory, Principles, Practices and/or Content:

- computer hardware and system development
- principles, theories, and practices of computer data storage, retrieval, and/or restoration.

Skills/Ability to:

- coordinate computer systems and application design
- use automated software applications
- use State or agency-specific software application program(s)
- communicate information and ideas clearly, and concisely, in writing; read and understand information presented in writing
- speak clearly, concisely and effectively; listen to, and understand, information and ideas as presented verbally
- make a decision or solve a problem by using logic to identify key facts, explore alternatives, and propose quality solutions
- use logic to analyze or identify underlying principles, reasons, or facts associated with information or data to draw conclusions
- read and interpret data from scientific instruments or equipment

- use computer relational database software
- evaluate information against a set of standards

The DHRM descriptions are in some cases broader than the type of data or information analysis role suggested by the Data Steward Approval document. This occurs because the DHRM fits the Information Analyst classification within an organization's information technology division or section and not within individual business sections or units. In these scenarios the role of the information analyst intersects with that of the IT Analyst's which is traditionally more of a system developer's or programmer's one.

Such positions are concerned with not just the infrastructure that contains data that can be analyzed but also the collection/input, operational storage and data maintenance aspects of IT. An example of this would be a operational payroll/personnel systems that includes everything from management of applications to printing payroll checks. What the USOE seems to be striving to achieve is an individual that is more focused on data that can be directly used for making tactical and strategic decisions whether they be at the classroom, school, LEA, state or national level. For this reason it is necessary to consider briefly and more closely what is loosely called the presentation layer. Components of the presentation layer can exist both outside of as well as within the more formal organizational data warehousing structure.

The Presentation Layer

Since it is very important in describing the roles of the data/information analyst and a data steward assuming both stewardship and analysis responsibilities, it is necessary to discuss this data concept and ultimately its constructs and structures in more detail. This discussion will hopefully promote better understanding of the data steward's and data analyst's roles as well as that of enterprise IT.

Definition

It is difficult to find good concise or even rambling technical definitions of the data presentation layer. Often it is considered to be part of or an extension of the data warehouse. Unfortunately, more often than not data warehouse definitions are varied, ambiguous and contradictory. Some articles address the presentation layer, but there are few, if any, good consensus definitions.

In general, the presentation layer manages the flow of information from the lower level or staging part of the data warehouse to the analyst, providing an interface that makes it easier for the skilled analyst to view and work with the data. This layer is where graphical user analysis and reporting tools are most often applied.

What this means is that various data structures represent detailed extracts as well as abstracts (both disaggregations and aggregations) of the complex and

highly relational data stored in the staging or operational part of the warehouse as easier to access structures.

These extracts range from: de-normalized data base tables, simple spreadsheets, pivot tables, flat files/tables on up to more flexible dimensional analysis models such as OLAP (On Line Analytical Processing) cubes (Cognos in USOE's case). Another integral component of the warehouse is collection of computer code or algorithms that embody the enterprise business logic of the warehouse and help build the presentation layer. Examples include the AYP, AMAO and U-PASS "reports". Said another way, the business logic layer contains the actual business rules, expressed as computer algorithms, to provide the services to transform raw data into more useful information much of which is required by statute or rule.

Use

Because it can be so varied, the presentation layer should be defined and administered jointly by IT and data stewards from various agency sections. Actual technical maintenance should be much more limited. At the present time it is a joint effort of Computer Services and the Assessment and Accountability's Results Team and other technically qualified section personnel such as the IT and data analysts in CTE.

Read access to all presentation layer data is granted to these individuals and in the future to data stewards. As noted this presentation layer includes all de-normalized (few integrity constraints) database tables, OLAP cubes (can be Cognos), pivot tables and other data structures (e.g. MS Reporting Services, Sybase IQ). Authorized persons have, and future data stewards can have, secure access on a 24/7 year-round basis except for scheduled times when the presentation layer is undergoing scheduled loading, maintenance or quality assurance procedures.

Accountability

Since accountability is a section of the Data Steward Approval form draft just a few comments about the accountability of data stewards should be made. Although data stewards reside under (hopefully directly under) the section director they must be knowledgeable of, and help communicate both federal, state and agency data and IT statutes, rules and policies.

First and foremost Data Stewards must follow the Family and Education Rights Privacy Act, FERPA, and USOE rules for maintaining and protecting student and educator data. These rules include those provisions that stipulate the protection of laptop files or any other files external to secure USOE servers. There are numerous guides to FERPA and the USOE legal staff members are well versed in it. Appendix E contains an outline of FERPA principals and issues. One very important USOE policy is that no student level

data, especially data that have been de-identified (removal of name, etc.) ever be kept on a notebook or removable storage device.

Next Steps

What the USOE may want to consider initially doing is to focus on the metadata type roles (definitions, rules, communications) of the envisioned data steward position and not as much on information analysis aspects that involve the retrieval, reporting and analysis functions. At best there are probably only 3-4 persons in the agency at the present time that could full both sets of roles (data steward and data/information analyst). Using test individuals in such positions could only be done if they gave full commitment to the job of being a data steward.

We can start with what we already have and put most of the emphasis on: data definitions, adequacy, quality, business rules and communications. Communications or the liaison role specifically means: working closely with districts and IT staffs (both LEA and SEA) as well as with USOE directors or coordinators; attending all data steward meetings both at the USOE and LEAs; and promoting data stewardship at the LEA level.

We currently have skilled data analysts and statisticians on staff at the USOE, but they are a scarce commodity in most sections. It may seem to be economical but it may not be practical to combine the role of data steward and data analyst into one position. While we build capacity for a data analyst positions in each section we may need to establish a small and more centralized pool of such individuals who could also conduct “at large” research on special cross-sectional projects.

It may be productive to have a group of USOE staff with data interests meet and customize a data steward definition for the USOE and work through some interim recommendations.

Appendix A

August 28, 2006

Data Steward Approval - DRAFT

To be completed and approved by Associate Superintendent

A. Appointments

One Data Steward may be approved from each of the following USOE divisions, assuming the division can provide compensation and/or add the assignment onto the work assignment of a current employee:

- Assessment and Accountability
- Curriculum and Instruction
- Educator Licensure and Quality
- Instructional Technology
- Career and Technology Education
- Title One
- Special Education
- Charter Schools

B. Access

The approved Data Steward shall have unrestricted access, without further permission beyond this approval, to manage and store data from the Data Warehouse at the presentation layer (not the original, raw data). This presentation layer of data may be stored, maintained, and administered as the Data Steward deems appropriate. In order to effectively manage data at the presentation layer, 'read' access will be granted to the Data Steward for every table within the data warehouse and associated databases. Data Stewards shall be given access codes and passwords on a regular basis to maintain 24/7, year-round access.

C. Accountability

Data Stewards must follow the USOE guidelines on maintaining and protecting student data, including those provisions that stipulate the protection of laptop files.

D. Permission

I hereby approve _____ as Data Steward
Data Steward Name

for the _____ Division of the USOE.
Division Name

_____ _____

Signature of Associate Superintendent Date

E. Agreement

As the Data Steward for the assigned division, I agree to all access and accountability stipulations and recognize that job action will be taken if I violate any condition of this approval.

_____ _____

Data Steward Signature Date

Appendix B

Exemplary Data Steward Definitions and Descriptions

The concept of a data steward has been around for at least a decade. It's functions of data definitions, business rules, communications, data quality etc. have existed as long as there have been data. In the information age, as is the case with other older more intuitive functions, the increased complexity of or organizations and data require a formal position. The EDEN (Education Data Exchange Network) is one example of a complex data collection and analysis program which is causing many states, most notably, California, to organization data definition and coordination functions with its state level public educations organizations around the idea of data stewards.

1. FROM: http://en.wikipedia.org/wiki/Data_steward

A **data steward** is role assigned to a person that is responsible for maintaining a data element in a metadata registry.

Data stewardship roles are common when organizations are attempting to exchange data precisely and consistently between computer systems and reuse data-related resources.

Data Steward responsibilities

A data steward insured that each assigned data element:

1. Has clear and unambiguous data element definition.
2. Does not conflict with other data elements in the metadata registry (removes duplicates, overlap etc.)
3. Has clear enumerated value definitions if it is of type Code.
4. Is still being used (remove unused data elements)
5. Is being used consistently in various computer systems
6. Has adequate documentation on appropriate usage and notes
7. Documents the origin and sources of authority on each metadata element
8. Has clear and unambiguous data element definition.
9. Does not conflict with other data elements in the metadata registry (removes duplicates, overlap etc.)
10. Has clear enumerated value definitions if it is of type Code.
11. Is still being used (remove unused data elements)
12. Is being used consistently in various computer systems
13. Has adequate documentation on appropriate usage and notes
14. Documents the origin and sources of authority on each metadata element

Benefits of Data Stewardship

1. Consistent use of data management resource

2. Easy mapping of data between computer systems and exchange documents
3. Lower costs associated with migration to Service Oriented Architecture (SOA)

Assignment of each data element to a person sometimes seems like an unimportant process. But many groups have found that people are more likely to trust and use a system where there is a person they can call with question on each data element.

An example of data stewardship is the EPA metadata registry [1]. Note that each data element has a "POC" (point of contact).

2. For a LEA oriented data steward description from the National Center for Education Statistics see: http://nces.ed.gov/forum/pdf/tip_datasteward.pdf. Many of these roles and responsibilities are also applicable on the state level.

Excerpts include:

You serve your administrator by ensuring that the statistical information reviewed by senior staff represents data that have been entered accurately and collected systematically. Furthermore, you enhance the information reporting process through staff development and collaboration with the various offices and programs responsible for producing data and information.

Things to Think About

• Does the information reviewed by your superintendent and senior staff represent facts based on accurate data from programs and offices?

• Does everyone in your school district understand how data are used to benefit the instructional program and provide funds for services?

• Are data collected systematically in the school district?

• Are the staff responsible for entering data trained to do an effective job?

• Is there a process in place that allows “end-users” to request or modify reports?

• **Are you and the information technology (IT) director operating collaboratively**?

Things to Do

• Coordinate the data collection process.

• Provide professional development for staff members leading toward a Culture of Quality Data in the school. The sessions might include:

1. demonstrations that incorporate hands-on training, enabling data entry personnel to become used to the actual data entry screens;

2. examples that actually reflect situations that will be encountered;

3. handbooks or guidebooks, with copies of data entry screens, systematic instructions, and the rationale for entering the data;

4. descriptions of the procedures for obtaining assistance (e.g., help desk phone number, online and/or e-mail query process); and

5. copies of the reports created from the data, enabling trainees who enter information to have a sense of how their work affects the operation of the school.

• Resolve discrepancies in information before reports are forwarded to senior staff.

• Develop a process that allows staff to request new reports or modifications of existing reports.

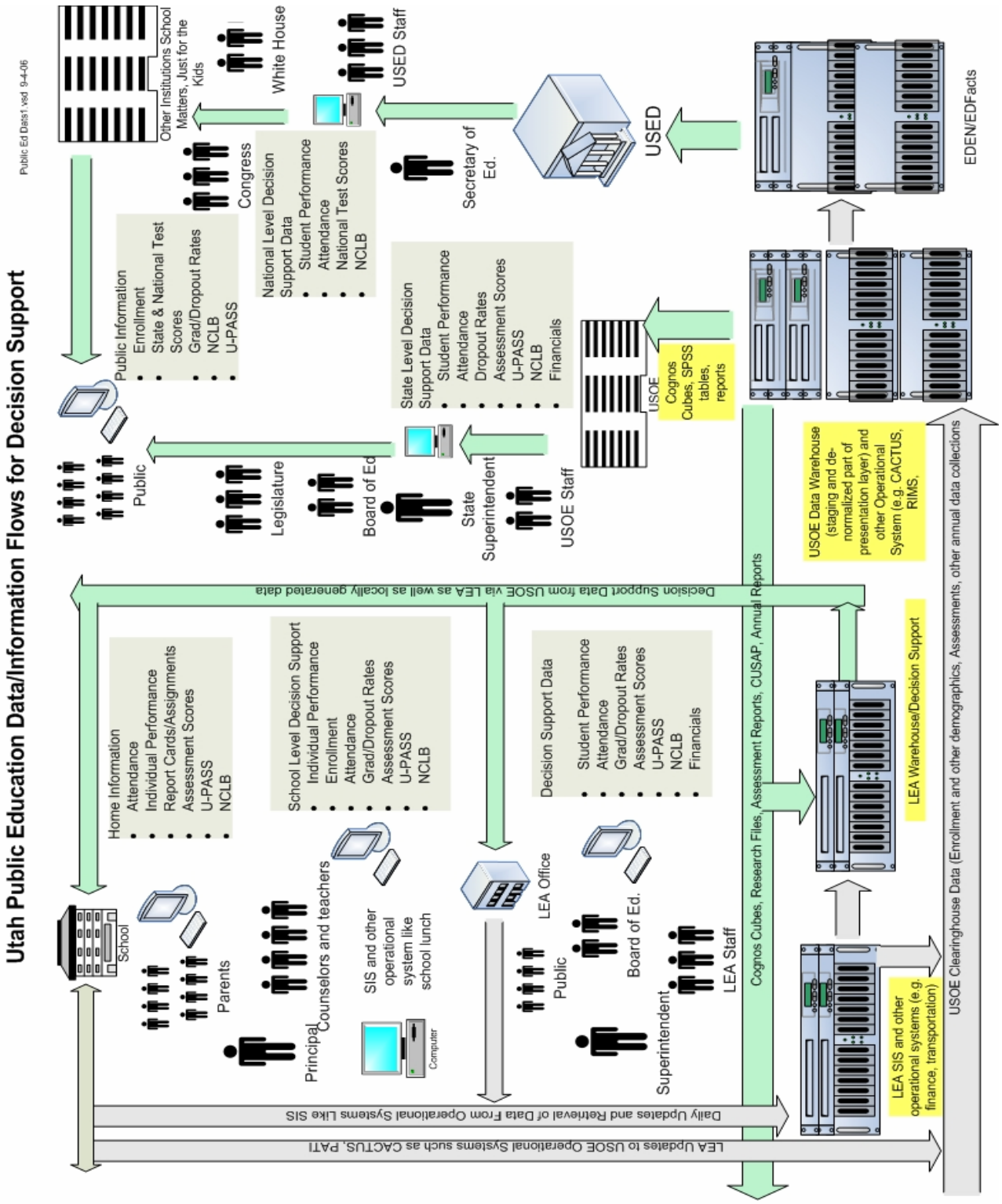
• Collaborate with the district technology director or coordinator to enhance the ability of computer programs to determine effective editing procedures for reports and other information.

Outcomes (What's in it for me?)

By helping staff members to understand the importance of data entry and data collection, and to see the process that leads to data-driven decisions, you are directly involved in courses of action that lead toward improved student achievement and increased services provided to the district and schools.

3. For and additional and very comprehensive discussion of the data steward role see: http://www.tdan.com/i033fe01.htm

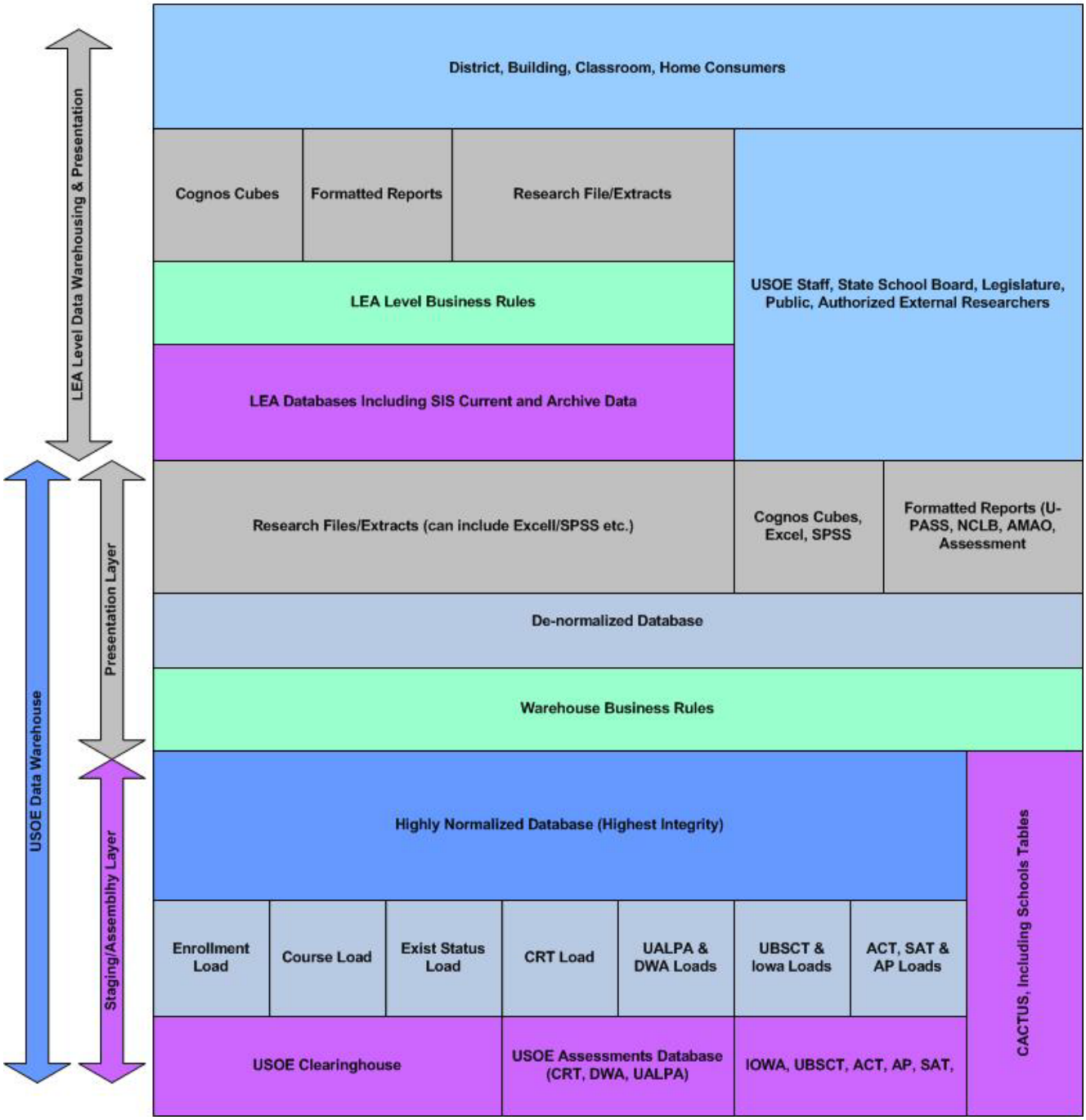
Appendix C


Utah Public Education Data/Information Flows for Decision Support
Public Ed Data1.vsd 9-4-06
Other Institutions School Matters, Just for the Kids
White House
USED Staff
Congress
National Level Decision Support Data
Student Performance
Attendance
National Test Scores
NCLB
Secretary of Ed.
USED
EDEN/EDFacts
Public Information
Enrollment
State & National Test Scores
Grad/Dropout Rates
NCLB
U-PASS
State Level Decision Support Data
Student Performance
Attendance
Dropout Rates
Assessment Scores
U-PASS
NCLB
Financials
USOE
Cognos Cubes, SPSS tables, reports
Public
Legislature
Board of Ed
State Superintendent
USOE Staff
USOE Data Warehouse (staging and de-normalized part of presentation layer) and other Operational System (e.g. CACTUS, RIMS,
Decision Support Data from USOE via LEA as well as locally generated data
Home Information
Attendance
Individual Performance
Report Cards/Assignments
Assessment Scores
U-PASS
NCLB
School Level Decision Support
Individual Performance
Enrollment
Attendance
Grad/Dropout Rates
Assessment Scores
U-PASS
NCLB
Decision Support Data
Student Performance
Attendance
Grad/Dropout Rates
Assessment Scores
U-PASS
NCLB
Financials
Cognos Cubes, Research Files, Assessment Reports, CUSAP, Annual Reports
LEA Warehouse/Decision Support
School
Parents
Counselors and teachers
Principal
SIS and other operational system like school lunch
Computer
LEA Office
Public
Board of Ed.
Superintendent
LEA Staff
LEA SIS and other operational systems (e.g. finance, transportation)
USOE Clearinghouse Data (Enrollment and other demographics, Assessments, other annual data collections
Daily Updates and Retrieval of Data From Operational Systems Like SIS
LEA Updates to USOE Operational Systems such as CACTUS, PATI

Appendix D

09-04-2006 Public Ed. Warehouse1.vsd

Utah Public Education Database, Warehousing and Reporting Abstract



Appendix E

FERPA & STATE LONGITUDINAL DATA SYSTEMS
Outline of Preliminary Analysis

I. Family Educational Rights & Privacy Act (FERPA) – Background

- Prohibits disclosure, without written parental consent, of education records of students by schools and LEAs that receive funds from U.S. Education Department (USED), except for statutory list of authorized disclosures**.**
- Rights accorded to parents under FERPA transfer to student when student turns 18 or enters postsecondary education.
- "Education records" broadly defined to mean any records/documents with personally identifiable data relating to the student maintained by the school or LEA or a person acting for it.
- Two overall circumstances for disclosing data derived from a student's education record without parental consent –
    - If the data neither personally identifiable nor easily traceable to student; or
    - If disclosure fits within a FERPA exception.
- FERPA includes provision barring recipients of authorized disclosures from making redisclosures.
- Under 2002 Supreme Court ruling in *Gonzaga University v. Doe*, no right of parents or others to sue a school, LEA, or state for alleged violations of FERPA.
- USED, through its Family Policy Compliance Office (FPCO), administers FERPA.
- The potential sanction for FERPA violation is cut-off of USED funds, but law requires USED to seek voluntary compliance before seeking that remedy.
    - Enforcement action by USED may be challenged in administrative hearing and subsequently in a federal court of appeals.

II. Clearly Permissible Disclosures (Without Parental Consent

- **Anonymous & Aggregate Data** -- School, LEA, or longitudinal data system (LDS) may use a non-personal identifier to create an anonymous data file for student and disclose the file, without regard to specific FERPA-authorized disclosures, if the key linking file to student is protected from disclosure and data not easily traceable to student.
- **Evaluation/Audit --** School or LEA may give personally identifiable information from a student's education record (PII) to state education authorities (including LDS data warehouse) for purpose of evaluating and auditing federal and state programs, including making accountability determinations under federal and state law.
- **PII on Test Scores, Enrollment, and Graduation** – The No Child Left Behind Act expressly authorizes linking test scores on state assessments, length of enrollment, and graduation records in an LDS and the provision to schools of diagnostic reports based on individual student assessments.

- **Organizations Conducting Studies To Improve Instruction –** School or LEA may give PII to organization conducing study for it or on its behalf to validate or administer predictive tests, administer student aid programs, or improve instruction. LDS clearly could receive PII from school or LEA under these provisions.
- **Sharing student records among schools** attended by student or in which student seeks to enroll, including between elementary and secondary schools and postsecondary institutions and between schools in different states.

III. Open Issues

- **May PII be provided to LDS for purpose of tracking individual students and diagnosing and addressing their individual needs?**
    - FERPA expressly authorizes providing PII to state education authorities only for evaluation and audit of programs.
- **May PII obtained by LDS be redisclosed to schools or other organizations, given general FERPA prohibition on redisclosures?**
- **May PII be disclosed to organizations that initiate their own studies related to improvement of instruction, or must the study be initiated by the school, LEA, or LDS to meet the law's provision that the study be for them or on their behalf?**

IV. Recommended Solutions

- **LDS should be understood to be acting for schools and LEAs in maintaining and analyzing their student education records.** This interpretation --
    - Is consistent with FERPA's definition of education records as including records maintained by a person acting for school or LEA.
    - Reflects the reality that a central purpose of LDS is to debalkanize maintenance of records and use them to assist schools and LEAs in educating their students.
    - Does not disturb USED precedents relating to one state agency not being an authorized representative of another state agency, which related to a different disclosure provision of FERPA that stressed not sharing information beyond certain officials. It is also consistent with FPCO interpretation that disclosures to officials of school and LEA with appropriate interests in the records may include disclosures to contractors.
    - Solves two issues – the legal basis for the LDS obtaining data from schools and LEAs for purposes other than program evaluation and audit of programs; and the authority of the LDS to redisclose the information it obtains from the LEA or school.
        - FERPA provisions barring redisclosure would not apply because the provision of education records to the LDS would not constitute a disclosure outside the school or LEA.
        - For the same reason, provision of the records to the LDS would not be subject to recording and parental notification.

- Permits FERPA to be harmonized with other federal statutes (in particular, the Education Sciences Reform Act) in which Congress supports development of state LDSs, not simply for evaluation and audit purposes, but to improve student academic achievement.

- **Issue state laws, regulations, and/or guidance defining scope and process for LDS implementation of FERPA-authorized disclosures for studies.** This approach –
  - Permits states to define when a study is "authorized" by the LDS on behalf of a school or LEA, and, consistent with common research practice, could include "field-initiated" studies that serve the purposes of the school or LEA to improve instruction.
  - Would establish a process for the approval of such studies by the LDS, including written assurances by the research organization to safeguard and not redisclose PII, perhaps modeled after the Institute for Educational Sciences (IES) licensing procedure.
  - Would implement, and not disturb, FPCO precedent that to come within the FERPA study exception, a study must be "authorized" by a school or LEA.
    - Another possible approach, which *would* reverse FPCO precedent, involves an analysis that USED itself, through the IES process for licensing other organizations to obtain restricted use data for research, has already in practice reversed FPCO's position. IES has a statutory obligation to disseminate its data, but that obligation is subject to confidentiality standards that by law incorporate FERPA. It appears that IES licensing involves a redisclosure of PII under FERPA to organizations initiating their own research projects.

V. Follow-Up Steps for States

- Review and clarify state law/regulations/guidelines on the role of the LDS in maintaining and analyzing student records on behalf of schools and LEAs.
- Take steps to ensure that each school or LEA, in its annual FERPA notice to parents, informs parents of the role of the LDS in maintaining records for the school or LEA; which categories or types of records are covered; the criteria for determining which employees involved in administering the LDS have legitimate interests in having access to the PII; and the procedures for asserting rights under FERPA with regard to these education records.
- Enter appropriate agreements with schools/LEAs or regulate regarding the distribution of responsibilities between them and the LDS in maintaining the records and implementing procedures required under FERPA, including access of parents to the records and their rights to challenge the contents of the records.
- Issue regulations/guidelines on the scope and process for authorizing studies by independent organizations, including privacy protections consistent with FERPA.
- Review state privacy laws.