

OFFICE OF THE STATE BOARD OF EDUCATION

# State of Idaho

---

## Revised Statewide Longitudinal Data System (SLDS) Needs Analysis

Submitted  
November 14, 2012

This document updates the current state of Statewide Longitudinal Data System (SLDS) efforts in Idaho, describes options, and recommendations for maturing to a P-20 to Workforce SLDS.

# Table of Contents

- Executive Summary .....2**
  
- Overview.....3**
  - History ..... 3
  - Current Status ..... 3
    - K-12 ..... 3
    - Postsecondary..... 4
  
- Need for P-20 to Workforce SLDS .....5**
  - Federal Requirements ..... 5
  - Future Initiatives and Grants ..... 5
  - Strategic Plans..... 5
  
- Issues .....6**
  - Gaps..... 6
  - Barriers..... 8
  - Risks ..... 10
  
- Recommendation .....11**
  
- Execution Plan .....13**
  - Phase 1 – Postsecondary Repository and link to K-12 SLDS for P-20 SLDS ..... 15
  - Phase 2 – Maturing the SLDS environment ..... 16
  - Phase 3 – Finalize Design and implement a postsecondary Data Warehouse ..... 18
  - Phase 4 - Transform to P-20W SLDS & Business Intelligence solution ..... 19
  
- References.....21**

## Executive Summary

The updated Needs Analysis is intended to provide the State Board of Education (Board) with an overview of the current status of longitudinal educational data collection, the gaps, barriers, risks, and to provide recommendation regarding the most appropriate path forward for collecting student level data over time.

The Board approved Phase 1 and Phase 2 of this project plan in August 2011. The information contained herein provides a revised scope for Phase 2 and provides revised recommendations to the Board regarding Phase 3 and Phase 4. Staff recommends the Board accept the recommendations and direct staff to move forward with the revised Phase 2 and Phase 3. Revisions to Phase 3 require finalization of the design and implementation of materialized aggregate views. This is a revision from the previously proposed Phase 3 design that would have created a full data warehouse; the current conclusion is that is an unsupported option from the resources and costs required and that the needs for data can be satisfied by building a second aggregated data layer in the postsecondary SLDS. Phase 4 would be the final stage, transforming to a P-20W SLDS with Business Intelligence solutions. The four phased approach provides flexibility and allows Idaho to continue to meet federal deadlines and reporting requirements in a manner that will best utilize resources and aid proper planning and design. The four phase approach limits the burden on the institutions while still meeting the requirements of the various grant information needs and reporting requirements.

## Overview

### Current Status

- **K-12**

The K-12 SLDS, ISEE, began student-level data collection October 1, 2010. Rollout of the initial Schoolnet application has been completed. Enhancements to Schoolnet are being carried out using a grant from the Joe. and Kathryn Albertson Foundation (JKAF).

- **High School Feedback Reports**

The first version of the High School Feedback reports cover a subset of data from 2004-05 and 2010-11 data from ISEE. The High School Feedback reports contains enrollment in postsecondary education, retention, and graduation data. The request has been submitted for the ISEE 2011-12 data and once that data is received and processed through the National Student Clearinghouse a new



set of reports will be issued (anticipated to be December 2012). Efforts are also underway to form a task force to identify additional data elements from the postsecondary SLDS that can be included in future versions to enhance the reports.

■ **Postsecondary**

A single, consolidated postsecondary database has been constructed. The eight public postsecondary institutions have transmitted 2010-11 academic year data. A request has been made for the 2011-12 academic year core data to be provided by December 31, 2012. The data dictionary has been revised. The revised data dictionary will allow for additional data elements beyond the core data to be collected. Once data are imported, and the data validation reports produced and returned to the institutions. It is anticipated that core data covering 2010-12 will be available the first quarter of 2013. The National Student Clearinghouse is being utilized for enrollment and graduation data on students who attend non-public and out of state institutions.

The goal is to eventually expand collecting more detailed private and for-profit data into the SLDS from the institutions interested in participating.

## **Need for P-20 to Workforce (P-20W) SLDS**

### **Federal Requirements/Efforts**

- By accepting American Recovery Reinvestment Act State Fiscal Stabilization Funds, Idaho agreed to four assurances; one of which consisted of implementing the 12 elements of the America COMPETES Act by December 31, 2011, which requires a P-16 SLDS. Idaho currently meets the 12 elements of the Act, but cannot produce the requested reports due to lack of historical data. In July 2012, Idaho received a FY2012 SLDS grant that funds three initiatives:
  1. Enhancements to the Education Unique ID (EDUID) matching system (scheduled for completion by June 2013)
  2. Creation of a Research Request process (scheduled for 2014-15)
  3. Creation of the labor longitudinal data store (completed by June 2015)

### **Future Initiatives and Grants**

For Idaho to pursue future grant opportunities, Idaho must have the ability to track student level data from K-12 through postsecondary education and into the workforce. As part of Idaho's participation in the Complete College America (CCA) initiative, we are required to track the progress on outcomes over time and through systems. This

process is being done manually by the institutions and is very time consuming. Getting the postsecondary SLDS online will significantly reduce the time and effort to produce the data and allow Idaho to eliminate the duplication in the aggregate data currently collected.

### **Strategic Plans**

- The Board, in its Strategic Plan, established the goal to have a P-20W SLDS developed and implemented by 2015. The State Department of Education is also dependent on an SLDS that includes postsecondary data to meet their goal of students being prepared to continue their education without the need for remediation. In addition, the Board has set the goal that 60% of 25-34 year olds have a postsecondary degree or credential by 2020. Idaho needs the capacity to track students over time, a system in place to conduct the analysis of where students are dropping out of the educational pipeline, and the capability to measure the effects of changes in education delivery against this goal.

# Issues

## Gaps

- Data dictionary:
  - The original Data Dictionary reviewed by the institutions was reduced to five core tables due to time constraints. A second full review of the data dictionary was started earlier this year and is nearing completion.
    - Action: The data dictionary is to be completed and published online.
- Data submission request has been made for core tables for 2011-12 academic year
  - Action: Process 2010-11 data collected
  - Action: Collect 2011-12 data
- Data Validation:
  - The Education Unique ID (EDUID) is the link between K-12 and postsecondary. The EDUID system developed and managed by the State Department of Education is utilized to obtain and maintain the identifiers. Because the system utilizes demographic information to create and match individuals, there are opportunities for mismatch. Improvements were made to the EDUID system earlier this year to improve the match rate. These changes include the addition of former names fields, high school attended, and a preview feature to show which records were matched, records where new EDUID's will be assigned, etc. This mismatch will reduce the reported rate for students moving from grade to grade, and on to postsecondary. Some of the causes are:
    - Name changes that are not reflected in the system.
    - Name given to enroll in postsecondary is not same name provided in K-12. K-12 requires a legal name, postsecondary does not.
    - Changes in punctuation can potentially cause mismatch
    - "Seed" files (ACT, SAT, ISAT, Teacher files) caused a number of duplicate entries that are still being rectified.
      - Action Items:

- Investigate methods for identifying the mismatch rate.
    - Enhance the EDUID matching process to improve the match rate. This is a deliverable under the FY2012 SLDS grant.
    - Promote the use of EDUID on high school transcripts to verify identity when student moves to postsecondary.
    - Pursue electronic transcript files to obtain EDUIDs electronically.
- Business rules
  - A basic set of rules to validate the postsecondary data were created along with the Extract, Transform, and Load (ETL).
    - Action Item: Develop enhanced business rules to reflect additional data elements that will be collected in the future.
- Programming
  - The ETL scripts were created for the base tables.
    - Action Item: Develop ETL for additional tables.
- Common Education Data Standards (CEDs)
  - The second version of CEDs has been released and there is an option to match a state SLDS data dictionary. This allows multi-state matching of data elements.
    - Action Item: Load Idaho Postsecondary data dictionary to CEDs website.
- Workforce Outcomes
  - Expanding the P-20 SLDS to a P-20W SLDS (the addition of Labor data) requires establishing necessary agreements and providing data to the Idaho Department of Labor.
    - Action Item: finalize MOU (currently routed for signatures)
    - Action Item: Define format and utilize secure file system for transmission of data.
    - Action Item: Since the Idaho Department of Labor has obtained the driver's license files, need to set up field definitions to also support sending records where SSNs are not available.
    - Action Item: Idaho Department of Labor develop Labor Longitudinal Data store (funded by the FY12 SLDS grant).

- Action Item: Define data needs that require labor data.

- Questions:

- Although a list of potential questions has been developed that the P-20W SLDS could help answer, a clear definition of the needs of the potential users has not been completed. The list is being expanded to include additional labor/education and labor specific questions.

- Security:



- Data security is a major concern. OSBE will leverage SDE's K-12 SLDS security solutions to duplicate these successful strategies.

- Quality:

- A critical requirement of any database is controlling data quality (i.e. data accuracy, standards, integrity, and completeness) from both an I.T. and business perspective.



A Data Management Council (DMC) was established by the Board and guide the development of the policies and procedures necessary to properly manage the data in the P-20W SLDS and serve as the primary review point for all data management activities.

- It is incumbent upon the school districts and institutions to provide clean data. With the wide variety of systems the school districts and institutions utilize, it is not practical to assume perfect data.

- **Agreements**



- Agreements between Idaho Department of Labor and the Board are being processed for signatures. Submitting legislation to mandate the data exchange should be considered in the future.

- **FERPA violation and disclosure of Personally Identifiable Information**

- The recent changes to FERPA provide increased capability to store and share education data, they do, however, add additional penalties as well. The U.S. Department of Education has created a Chief Privacy Officer and a Privacy Technical Assistance Center (PTAC) to assist states with the changes.

- **Stakeholder Engagement**

- The institutions have been engaged in the development of the SLDS Data Dictionary. The Idaho Department of Labor is

supporting the creation of the Labor longitudinal data store. A communications plan needs to be established with data users to ensure an informed and engaged process.

- **Student tracking**



- The Office of the State Board of Education (OSBE) SLDS budget is funding a contract with the National Student Clearinghouse to provide data to all Idaho public secondary schools and provide access for OSBE to submit secondary data files. This system is being utilized for the High School Feedback reports and the ARRA SFSF reports (when data are available).

## **Barriers**

- **Confidential Information and Requests**

- Due to the necessity to collect sensitive data such as personally identifiable information, Social Security Numbers (SSN's), and labor data to build a P-20W SLDS, the design of the postsecondary repository and data collection methods is complicated and time consuming. The amount of identifiable information has been limited in the postsecondary SLDS, this enhances protection, but also raises the work required to match records and respond to data requests.
- The common theme of other states that have a mature SLDS is to highly restrict access to student identifiable data, provide only the required level of information, and set return/destruction dates on the data usage. An SLDS provides a wealth of information that will attract requests for information; therefore, it is critical that the proper processes and procedures are in place before requests are received. The proposed research request website incorporates an automatically generated "Restricted Use Data Agreement" that reflects the data elements in the specific request, specifies the users and timeframe. This system will be built around providing de-identified data further limiting exposure.

- **Distance/Location**

- The locations of the eight public postsecondary institutions make it difficult and expensive to conduct face to face meetings. As much as possible remote meeting technologies are utilized to ensure participation.

- **Time**
  - State contracting restrictions and an inability to hire new staff have delayed the original timeline for implementation. The current timeline is to collect the 2010-11 and 2011-12 core data by December 2013; spend Q1 2013 working on data quality and business rules with a goal of having usable data by the end of Q1. In parallel, a request will be made to populate the other data tables, this will be time consuming as it will require the institutions to develop additional SQL scripts and changes to the ETL process.
  
- **Budget**
  - Consultants and remote access are being utilized to develop the SLDS and reports. This limits the scope of work that can be executed concurrently. This is partially due to space limitations and having no direct access to the domain that the postsecondary SLDS is operating under. The current budget is adequate to perform the remaining work in Phase 2.
  
- **Competing Priorities**
  - There are other major projects currently underway at both State Department of Education (SDE) and several institutions that preclude leveraging some internal resources. These include, but are not limited to, the continuing development of the K-12 SLDS and integration of Schoolnet, other Board initiatives such as Complete College Idaho and Performance Based Funding. It is anticipated that participation of these entities is necessary to ensure the success of the P-20W SLDS. As much lead-time and flexibility will be provided to minimize the impact to other projects. This has continued to be an issue, and in June 2012, financial assistance was provided to most of the institutions to add an additional resource to support the SLDS efforts. This is having a positive impact on the data extraction at these institutions.
  
- **Data Availability**
  - The end goal is the capability to track students from pre-school (in Idaho, from Kindergarten) to the workforce. There are several hurdles to overcome:
    - Obtaining enrollment and graduation data from private and for-profit institutions will be a lengthy process. There may be

interest on their part to track outcomes for their students, and OSBE could provide that link in exchange for enrollment and graduation information from those entities.

- Labor data is an important component to this effort. Typically Unemployment Insurance wage data is utilized. Currently, the only field to match labor data on is the SSN. The K-12 SLDS does not require SSN and postsecondary typically only collects it if the student applies for financial aid; therefore, there is a gap in identifying students who go directly to the workforce from K-12 or those who leave postsecondary education and enter the workforce. The Idaho Department of Labor has reached agreement with the Idaho Transportation Department and has received the driver's license files. This will allow matching K-12 data and postsecondary where we do not have social security numbers. This is a tremendous achievement and is critical to determining workforce outcomes.
- Connecting to a multitude of other state agencies will have to be negotiated individually, but other states have been successful in this endeavor. The participation in the WICHE multistate data exchange project has provided the opportunity to interact with the other states and to discuss the processes they have used to put the agreements in place. The WICHE multistate data exchange project is investigating a governance structure that could be created to continue and expand the multistate data exchange.
- Graduates who join the military or take a federal job are another group that need to be identified and the agreements created to access this information. This is another area where the efforts of other states can be used as a model.
- Idaho participates in the Wage Record Interchange System for education (WRIS 2). This system holds wage data for 22 states currently and includes most of the states contiguous to Idaho. The Department of Labor has agreements with the other neighboring states. There is a restriction that requires the Department of Labor to aggregate the data before release. This somewhat reduces the capabilities of using this data and requires better definition of the cohort.
- There is a fundamental issue with the Unemployment Insurance data collected by IDOL. It does not contain hours worked or an occupation for each worker. Legislation would be required to alter the structure of the U.I. data.



## Risks

- **FY 2012 State Budget**



- The SLDS continues to be funded off the top of the college and universities budget.

- **Personally Identifiable Information Release**

- The P-20W SLDS will contain student level data to allow linking or extraction from multiple data sources. To mitigate the risk of exposing personally identifiable information, this data has been segregated in separate tables that can be secured and the access limited to only the required and approved personnel.
- To safeguard personally identifiable information, any public information requests will require data extracts of the results by internal resources, aggregation, approval from the Data Management Council and the owning institutions.
- Discussions will be planned with other state agencies who routinely deal with sensitive information to ensure that the proper safeguards are in place, including system vulnerability patching, tape storage, administration account control, and access logging.



- MOUs have been developed to manage data extracts for matching to labor data or other data exchanges.

## Revised Recommendation

Staff continues to assert that the construction of the P-20W SLDS should be completed over a period of time, through a four-phased approach. The P-12 SLDS and separate postsecondary repository (to form the P-20 SLDS) have been created. As time and resources allow, we need to incorporate additional data sources, and improve the functionality and use of the SLDS by maturing to a P-20W SLDS. Continuing implementation by adding a materialized aggregate level of data and eventually a Decision Support System to increase the usability and remove the dependency on technical resources to retrieve information.

Adding additional functionality in a phased approach provides early wins, allows Idaho to meet the Federal ARRA reporting requirements, assist the Board in making progress toward its Strategic Plan objectives, and increases stakeholder satisfaction.

The Board should continue as the entity leading the development of the P-20W SLDS toward a common vision across all of education. It is critical that all of the education and labor agencies work together toward a common SLDS goal. The Board's role as the policy-making body for all of public education provides an opportunity to eliminate these barriers and streamline the process. However, challenges will remain in aligning the various institutions and agencies towards the common goal of tracking students from the time they enter preschool through entry into the workforce.



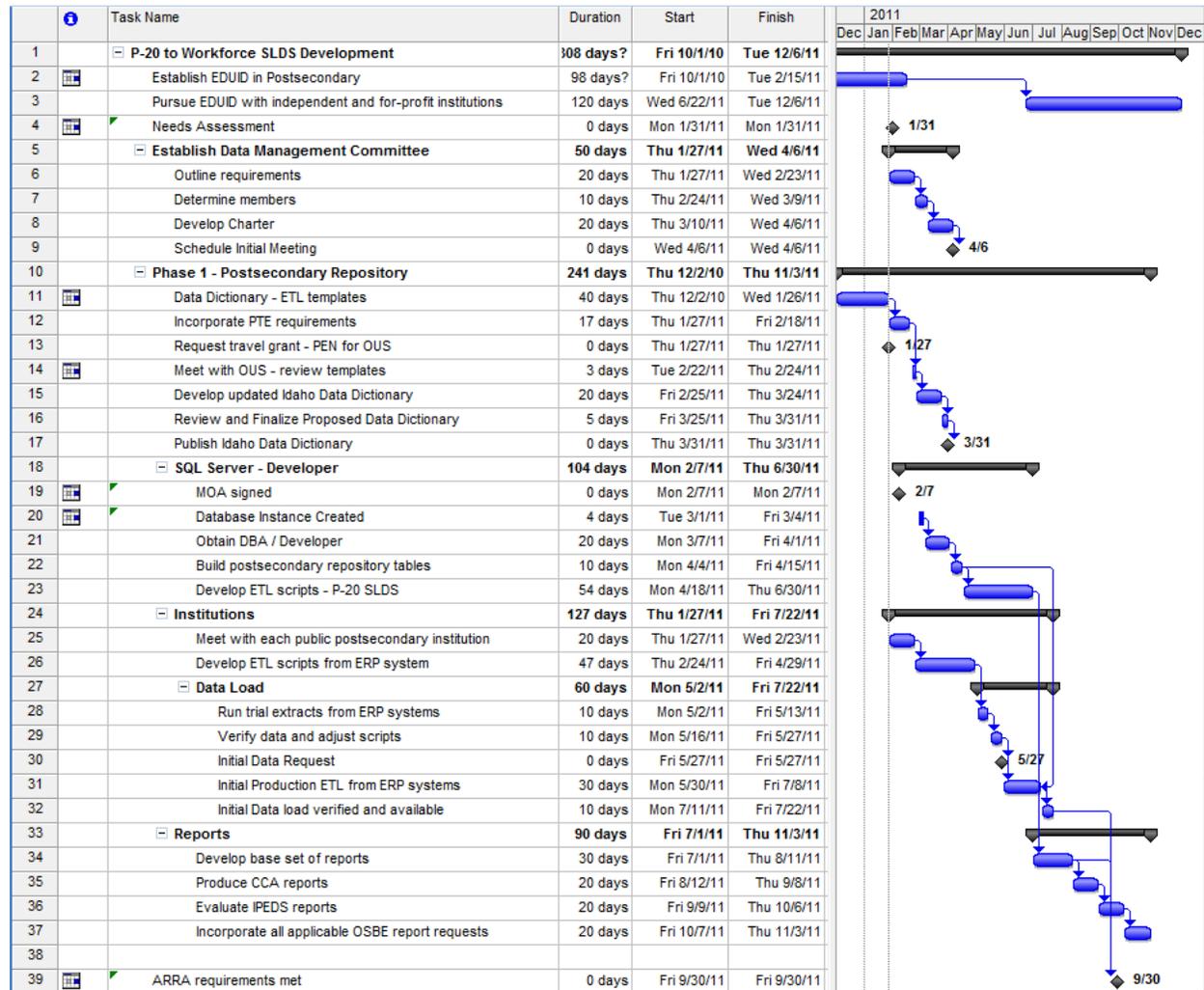
For the SLDS to complete Phase 2 in a timely manner, a commitment is required from all parties involved to make this a priority and to apply the necessary resources to complete tasks when scheduled. The participants required are the State Board of Education, the Office of the State Board of Education, the State Department of Education, the Division of Professional Technical Education, the Department of Labor, possibly the Department of Transportation, the Department of Corrections, all public postsecondary institutions, and if possible, the private and for-profit institutions. Ideally, ample lead time and as much flexibility as possible will be provided when engaging the institutions and departments. However, the reality is that there will be times when the P-20W SLDS will likely need to be given priority over other internal projects and initiatives.



# Execution Plan

## Phase 1 – Postsecondary Repository and link to K-12 SLDS for P-20 SLDS (complete other than reports)

Below is the execution plan and timeline for development of Phase 1.



- The EDUID implementation into the postsecondary institutions project is complete. The cost for this effort was covered by the institutions.
- The postsecondary SLDS database has been constructed on the SDE SQL server cluster. This solution has greatly reduced the cost and timeline for creation.

## Phase 2 – Maturing the SLDS environment

**(Cost \$1M, timeframe complete by June 30, 2013)**

Phase 1 delivered the P-20 SLDS core functionality. Phase 2 matures the environment to provide information to stakeholders, delivery of additional reports, transition of most OSBE data needs to the P-20 SLDS, improvements to the ETL process, and development of additional data sources. The current status of Phase 2 is as follows:

- Training and documentation plan developed (320 hours – internal staff) (open)
- Develop automated import leveraging SDE’s solutions and implement Memorandum of Understanding / Memorandum of Agreements as necessary to include additional data sources and users (400 hours ) (completed)
- Determine and develop standard SLDS reports (1 FTE) (in process)
- Logical model developed (320 hours - consultant or Institution expertise) (open)
- Database Analyst (1 FTE) (using consultant part time)
- Preliminary Design of the Postsecondary Data Warehouse (320 hours – consultant or institution expertise) (revised – design materialized aggregate views – consultant) (open)
- Incorporate workforce data and evaluate other outcome data (480 hours) (in process using FY2012 grant for IDOL portion of work)
- Determine hardware requirements
  - Expand SQL Server environment to support the data warehouse if necessary, or deploy a new solution (open)
- Deliverables:
  - Web ETL file submission (based on SDE’s source) (completed)
  - Reports: (open unless otherwise noted)
    - Integrate federal reporting
    - Transition reports (K-12 to postsecondary) (high school feedback reports developed, enhancements will be needed)
    - Analyze existing OSBE data requests and move to SLDS (in process)
    - Develop ongoing Federal Reports including (in process)
      - Completion of 1<sup>st</sup> year credits within 2 years
      - Tracking Students who enroll in postsecondary within 16 months of graduation
      - Students who complete 24 credits within first 2 years
      - Update of other ARRA reports

- Develop reports to answer critical questions from SBOE, institutions, SDE, and the legislature. (Performance based funding reports underway)
- Design – Investigate incorporation of ISEE data into postsecondary SLDS (open)
  - Determine data elements
  - Develop scope of work and cost estimate
  - **Develop MOU**
  - Execute project
- Design – Materialized Aggregate Views (open)
  - Investigate solutions in place in other states
  - Elemental design decisions made – structure and dimensions
  - Determine hardware, software, and support model

### **Phase 3 – Finalize Design and implement materialized aggregate views**

**(anticipated cost approximately \$500K, timeframe complete by June 30, 2014)**

Materialized SQL Aggregate Views. In the case of education, the materialized views transforms the repository into information that will support the Research Request process and are readily understood by the Institutional Researchers and analysts so they can independently analyze information (within the bounds of the security structure built into the system).

- Determination if P-12 data will be incorporated at this point
- Develop RFP for data aggregation implementation
- Engage institutional experts or consultant to finalize design of the Database structures
- Form committee to determine elements and aggregation level
- Develop materialized views.
- Hire consultant / leverage institution expertise
- Purchase or leverage software to support the database and reporting
- Develop a Business Intelligence roadmap
- Implement solution

## **FY2014 Resources and Cost (major items)**

### **(pricing based on current state procurement rates for consulting, internal = direct labor + burden + indirect costs)**

- Recommendation (implement P-20 SLDS materialized views) \$500K
  - (assumes allowance for internal labor)
- Database Architect Consultant - 240 hours @ \$100 = \$24,000
- Consulting – data crosswalk analysis, determination of data elements, develop views and reports \$50,000
- OSBE labor –
  - participate in design and verify information - 1,000 hours @ \$50 = \$50,000
- Support costs:
  - Reports / queries – 1 FTE for 1 year \$104,000
  - Data Quality manager – 1 FTE for 1 year \$104,000
  - Project Manager – 1 FTE for 1 year \$104,000
  - Database Analyst / SQL Specialist – 1 FTE for 1 year @74.80 \$149,600
  - Server support - .25 FTE for 1 year \$25,000
- Option – incorporate P-12 SLDS data
  - Add Developer/SQL for development - \$125,000

## **Phase 4 - Transform to P-20W SLDS & Business Intelligence solution**

**(anticipated cost approximately \$1.2M, timeframe complete by June 30, 2015)**

Business Intelligence (BI) tools allow self-service data query including drill down capability, ad-hoc analysis, and the ability to provide public access to aggregated data that is meaningful and productive. This expands the scope of the P-20W SLDS to include predictive techniques that will guide educators in optimizing the students achievement.

- Expand storage if required
- Gather requirements and determine solutions
- Review solutions deployed by institutions and SDE
- Develop legislation if required
- Develop and implement additional MOUs necessary to include additional data sources and users
- Develop training and support model
- Research and procure business analytics software
- Deliver training on BI tools and additional predictive analytics

- Expand storage if required
- Develop analytics reports and security model

### **FY 2015 Resources and Cost (major items)**

#### **(pricing based on current state procurement rates for consulting, internal = direct labor + burden + indirect costs)**

- Recommendation – add Business Intelligence tools to the data warehouse \$1.2 million
- Leverage the Decision Support System from another state to base load the capabilities similar to what SDE did for K-12 SLDS. SDE's successful implementation of the K-12 DRS was based on using Nebraska's consultant to assist in installing the base solution. SDE had over an 80% match rate on fields, which made having the system operational in a very short time period reasonable.
- Evaluate other states decision reporting systems and determine a solution
- Decision Support System Consultant 500 hours @ \$100 = \$50,000
- Programmers – modify DRS to match fields 480 hours @ \$75 = \$36,000
- OSBE internal labor – 1,000 hours @ \$50 = \$50,000
- Business Intelligence software and licensing \$100,000 to \$500,000.
- Ongoing support costs:
  - Decision Support Expertise – 1 FTE for 1 year \$104,000
  - Data Quality manager – 1 FTE for 1 year \$104,000
  - Project Manager – 1 FTE for 1 year \$104,000
  - DBA – 1 FTE for 1 year @74.80 \$149,600
  - Server support - .25 FTE for 1 year \$25,000

## **References**

- 1. US Department of Education site visit draft report– August 2010**
- 2. [http://searchsecurity.techtarget.com/sDefinition/0,,sid14\\_gci1366036,00.html](http://searchsecurity.techtarget.com/sDefinition/0,,sid14_gci1366036,00.html)**