

Technical Proposal – South Dakota Longitudinal Research of Workforce Participants

Statement of Current Capacity

Although the South Dakota Labor Market Information Center (LMIC) currently has the ability to conduct longitudinal research, the system in place includes several stand-alone databases, making it cumbersome and in need of an upgrade to become a fully functional state longitudinal database system (SLDS). The Workforce Data Quality Initiative (WDQI) will be utilized by the South Dakota Department of Labor and Regulation (DLR) for the development of a workforce SLDS. The planned development will improve efficiencies and assist training providers in measuring the outcomes of the workforce system. In addition, the application will include the ability to query outcome information and create automated outcome reports.

Capacity to Maintain Secure Data Storage

LMIC is part of the DLR and is in a physically secured office with password protected work stations. The data in the SLDS will be housed within the state computer system and meet all state and federal security requirements. More information about specific plans for data security can be found on page 13 of this proposal.

Planned or Established Partnerships with Education

Though the DLR was not able to secure SLDS funding from the 2009 and 2010 solicitations, the preparation activities for those opportunities laid a solid foundation for establishing vital partnerships with the education system. DLR continues to engage with our state educational agencies in order to improve readiness for similar Data Quality Initiative activities. The DLR has full support from both the South Dakota Board of Regents (BOR) and Department of Education

(DOE), including planned data-sharing agreements (see attached letters of intent). The DOE oversees the state postsecondary career and technical institutions. These partnerships will improve the availability of educational outcomes on a long-term basis while promoting greater coordination between education and workforce programs. Backing from these partners is demonstrated through letters of support included in the attachment to this proposal.

Existing or planned data linkages for data sets

As LMIC is part of the DLR, LMIC will have access to wage records data and other employment and services training data, including UI claims, Workforce Investment Act (WIA), Wagner-Peyser, Trade Adjustment Assistance (TAA), National Emergency Grants (NEG) and other USDOL discretionary grants including the Disability Employment Initiative (DEI), the State Energy Sector Partnership (SESP) grant and Adult Education and Literacy (AEL) data.

The development and implementation of the SLDS will be utilized to track the paths of workforce program completers. Data sharing agreements already in place with the surrounding states of Iowa, Montana, Nebraska and Wyoming allow for the tracking of workers across borders and enable additional research, such as identifying industry-specific commuting patterns of residents. In addition, this system will also serve our project partners, including the BOR and DOE, to measure and track how coursework leads to employment and wage advances over time for program completers.

Program evaluations are key resources in determining whether publicly funded programs are achieving intended outcomes as effectively as possible and at the lowest possible cost. The DLR

will develop and fully implement the SLDS, enabling linkages between existing longitudinal workforce data and program completers to assist partners in measuring outcomes of the workforce system. The DLR has a history of collaboration and data sharing to track individuals through linking data sources. More than a decade of experience has laid a solid foundation for the DLR to continue linking to additional data and implementing a robust and comprehensive SLDS.

Full support from BOR and DOE and their ensuing partnerships will create the capacity to link data from workforce and education databases to track individuals longitudinally from secondary and postsecondary training providers into the workforce. For educational outcomes, DLR will conduct a computer-based data match using state Unemployment Insurance (UI) wage record data and education records. DOE/BOR will maintain oversight and direction of the data match to ensure any disclosure of education records complies with the Family Educational Rights and Privacy Act (FERPA) confidentiality requirements.

Plan Outline

One of the objectives for developing the SLDS is to assist workforce and educational systems in their production of longitudinal information to better meet future demands. Tracking individuals through the education and workforce systems will provide comprehensive information that will assist in the assessment of programs and policies, including those enacted to provide career training, offer quality jobs through the labor exchange, advance careers and help educators improve curriculum. Strategic goals of this project include:

1. Developing a longitudinal database system and populating with historical data sets.

2. Measuring the longitudinal outcomes of the workforce program completers, consistent with all applicable federal and state privacy laws.
3. Upgrading the existing *Eligible Training Provider List (ETPL)* application created in 2002, in which program outcome information can be queried and automated reports produced. This application will help policy makers and practitioners understand workforce and education program performance and assist in making more informed decisions.
4. Enabling the workforce system to be linked to educational systems.

Status of Longitudinal Educational Data System

The Student Teacher Accountability Reporting System (STARS) is a South Dakota Department of Education (DOE) initiative to build an educational SLDS. The STARS SLDS will drive the state's educational initiatives to improved instruction and student performance. STARS is part of a multi-phase effort to create a secure, quality-controlled state data repository that will make data more readily available for research and analysis by internal staff and external stake holders such as state, local and federal policy makers, external agencies, researchers and the general public.

In addition, STARS, at a minimum, seeks to leverage currently existing but disparate data collections to initially populate the repository. The STARS project will bring together historical and current data into one cohesive, consistent and consolidated environment for analysis and reporting. At a high-level these data include, but are not limited to:

- Student demographics, student enrollments, course enrollments and student attendance

- Program participation [Career and Technical Education (CTE), English Language Learner (ELL), Migrant, etc.]
- District and school financial data
- Staff demographics, staff assignments and staff certification
- Assessment data
- School safety and discipline data
- General district and school data

The project began in January 2012 with analysis and design, with a 10-month iterative development cycle. The initial release is planned for the third quarter of 2012. The bulk of the work will be performed during the 10-month period prior to the initial release, with annual rollouts and updates issued to 2017. OtisEd, a nationally recognized IT architect and systems integrator, with headquarters in the state of Georgia, will oversee the design and development of the STARS project. In addition to the architectural design of the STARS environment, OtisEd will provide training and support on STARS to state and school district users.

Along with the development of STARS, DOE has taken proactive steps in terms of data governance and preparation for future linkages with outside agencies. The result of this preparation was the formulation of an ongoing workgroup with the postsecondary governing body (BOR) and the DLR. This group primarily discusses existing data issues, future data considerations and plans for linking systems.

Technical institutes and career and technical education, governed by the DOE, are working on tracking placement at secondary and postsecondary levels. The BOR, which governs the state system of public higher education, started their longitudinal data system in 2002. The system went into production in 2004 and is utilized by all campuses to house student and financial data.

Sustainability

DLR will continue to engage partners and renew and expand data sharing agreements in order to ensure continued maintenance and analysis of the SLDS. DLR will automate data transmission where possible to reduce the burden on partners and lessen project costs long term.

Harnessing the interests of other state partners into support and funding for the SLDS is integral to the underlying partnership strategy of producing robust longitudinal analyses of program completers. The joint workgroup including workforce and educational partners (DLR, DOE and BOR) have discussed plans for sustainability for the SLDS, which could be accomplished through self-funding of all partners involved.

The South Dakota Workforce Development Council (WDC), which oversees implementation of WIA-funded workforce training programs for the state, unanimously supports DLR's WDQI grant application and approved this effort at the Council's April 2012 meeting. The support of the Council, as well as other state entities such as DOE and BOR, plus evidence of the SLDS's dynamic usefulness as it is developed and implemented, could be presented to the state legislature with the hope of securing financial support to sustain the SLDS after grant funding has ended.

Description of Partnership Strategies

Data sharing agreements with other states

As mentioned previously, DLR has a number of wage record data sharing agreements with other states and is actively soliciting additional agreements. A copy of one of these data sharing agreements is included in the attachment to this proposal. Partnerships will be strengthened through analysis that will support workforce and education planning and longitudinal assessment. DLR plans to develop partnerships outside the state workforce system through dissemination, communication and participation. Labor market information (LMI) offices produce monthly publications for a variety of customers — legislators, the executive branch, public and media. Results and analysis will be highlighted in these publications along with links to electronic reports maintained on state websites. DLR representatives will attend conferences when requested and provide presentations.

Partnerships with other state agencies

To create, sustain and strengthen partnerships, the DLR will encourage the development of technical and policy advisory committees comprised of data sharing partners to ensure accuracy within the state workgroup partners (BOR and DOE). These committees will identify the partner goals the SLDS can help attain, such as improved outcome reports, and ensure continued support from partners. Letters of support from DOE and BOR are included in the attachment; both agencies have committed to data-sharing partnerships with DLR.

The DLR will also partner with the South Dakota Bureau of Information and Telecommunications (BIT) to develop and manage the SDLS. BIT will oversee the programming

and hardware/software systems required for the SLDS. BIT will be responsible for the performance, integrity and security of the SLDS; programming the logical and physical design of the SLDS; normalization and storage capacity management; and database backup and recovery systems. Specific responsibilities include programming to meet established needs of users; controlling access permissions/privileges and monitoring security; performance and parameters; mapping out the conceptual design; refining the logical design; and allocating sufficient storage capacity. BIT is also responsible for programming complex queries in response to customer needs; hosting a portal to house these reports; developing data dictionary definitions, data models, metadata repositories and other data management tools. In addition, BIT will develop, manage and test back-up and recovery plans. BIT will also program the application to prepare outcome reports. Information will be delivered through an enhanced Web-based system allowing for delivery of most current data available. This partnership will adhere to ongoing security and confidentiality of the SDLS. An inter-agency agreement will be prepared to ensure BIT will perform information and technology services for DLR as set forth in SDCL 1-33-43(1-6). Since BIT may be accessing protected Personal Identifying Information (PII) in the course of their duties, this agreement will set forth the roles and responsibilities of the agency personnel relative to the protection of PII.

In addition, BIT will redesign and upgrade the Eligible Training Provider List (ETPL) application. The redesign would be normalized in SQL Server 2008, and code will be rewritten from Visual Interdev to Visual Studio 2010 framework 4.0. This will be a Web-based application intended for public interface.

SDWORKS, which is the South Dakota state online database of job seekers and employers, will extract a portion of the data collected in the upgraded EPTL application and utilize this data to synchronize with WIA, TAA, NEG and other federal and state program participants with training classes offered by state-approved providers. Participant statistics will be incorporated once again, with the program performance data presented in aggregate form and available on-demand in real time. Links will be incorporated when available for Facebook, Twitter and LinkedIn pages. In addition to more detailed reports and a more user friendly format, the upgraded EPTL program performance component of the SLDS will incorporate all requirements of the WIA, including:

- List of WIA-approved training providers
- Descriptions of program training provided
- Annual reports regarding program specific performance, including
 - 1) Program completion rates for all individuals participating in the applicable program
 - 2) The percent of all individuals participating who obtain unsubsidized employment
 - 3) Information on program costs
 - 4) Wages at placement in employment of all individuals
 - 5) Training success information for participants who received assistance including:
 - Percent of participants who completed applicable program and who are in subsidized employment
 - Retention rates in unsubsidized employment six months after first day of employment
 - Wages received by participants six months after first day of employment
 - Where applicable – rate of licensure or certification

The estimated cost to upgrade this application is \$124,000.

Description of Database Design, Data Quality Assurance and Proposed Uses

Description of Database Design

The DLR plans to model its platform and security standards after the ones established and in place in Wyoming. Currently, longitudinal databases in Wyoming reside on a Microsoft SQL Server[®] platform. SQL Servers offer several advantages: universal acceptance as a data storage platform, allowance of data from several sources, and ease of access for statistical analysis due to its ability to interface with Microsoft Office applications and statistical packages such as SPSS/SAS.

The server will reside within the state of South Dakota's virtual infrastructure. This design will allow for efficient use of resources, including tangible resources like computer hardware and software, as well as intangible resources such as consolidated management, fault tolerance, disaster recovery and reduced energy consumption.

A Microsoft SQL database will house the data within a Storage Area Network (SAN) system. For the initial development, BIT would utilize the SQL database basic design structure currently in use in the state of Wyoming for security and table space. In order to implement this new system into the current virtual environment, an estimated \$80,000 will be allocated for the host server, data storage and for the software and desktop setups for contract workers.

Personal Identifier

Most programs, with the exception of K-12, use SSN as a unique personal identifier. However, DOE does not currently collect SSNs. Therefore, to track these completers into the workforce,

the DLR will replicate a methodology developed by Wyoming. The DLR will match student records to the Division of Motor Vehicles (DMV) file to obtain SSN. Public transportation is rare in rural states where most youth drive, so the match rate is expected to be high. Using DOB, gender, name and home address, DLR will match student records to the DMV file to obtain SSNs. DLR and DOE will collaboratively develop the capacity to link student identifiers from the education system with SSNs to facilitate wage records and workforce system file matches.

Data Quality Measures

For UI wage matches, partner data is validated by individual program. Additionally, queries will be run to ensure accuracy and include validation queries on UI account numbers, NAICS and institution codes. After matches are performed, individual identifiers will be assigned and maintained on individual records. Identifier keys will be maintained in separate tables. Data quality and standards workgroups will be instituted with DLR, DOE, BOR and BIT staff members as well as stakeholders who are sources of longitudinal data.

Data providers will be asked to validate frequency distributions on all variables provided to the DLR. Providers will be asked to address or explain logical inconsistencies between variables and within variable values, based on documentation obtained from each provider. Providers will also be asked to explain high counts of missing values. The DLR will maintain all file and record documentation (to be made available to researchers) as well as training manuals used by data providers on an ongoing and archival basis for each data set, including mode of collection, state and federal legislation-induced changes, administrative changes (e.g. those implemented by the

National Center for Education Statistics for school districts) and federal reporting changes for state entities.

Scope of the Longitudinal Data

The DLR will implement a SLDS comprehensive system which will have the capacity to link individuals' demographic information, high school and college completer data, UI wage data, and workforce service data. The DLR, along with workgroup partners (BOR and DOE), will contribute to the overall goal of substantially reducing the variation among systems. As referenced earlier in this application, the DLR has access to many of the employment data sets specified in the SGA. Common data sets that are stored in a longitudinal manner and linkable on SSN or UI account number are UI wage records, QCEW and DMV drivers' license records including name, DOB, gender and address. As LMIC is part of the DLR, LMIC will have access to wage record data and other employment and services training data, including UI claims, TANF, SNAP, WIA, Wagner-Peyser, NEG and other USDOL discretionary grants, including the Disability Employment Initiative (DEI), State Energy Sector Partnership (SESP) grant, and Adult Education and Literacy (AEL) data. The DLR also has a contract with FEDES, so information from the U.S. Postal Service, federal civilian workers and the military is also accessible.

Data from DOE and BOR is currently housed within their own data systems. However, data-sharing agreements are planned as all agencies move forward with longitudinal systems. The SLDS will enable additional program partners to access aggregated data to measure program effectiveness and improvement over time. When fully implemented, the SLDS will enhance

analysis regarding demand, education and employment of completers. The linking of postsecondary enrollees and secondary graduates will complete the education and employment relationship picture and offer more comprehensive analysis. The confidentiality of all shared data will be protected in accordance with all state and federal laws as described in the Memorandums of Understanding (MOUs) in the attachments to this proposal. DLR will share longitudinal data, in accordance with established MOUs, to produce joint analysis of the workforce, education and training programs of shared labor markets that cross state lines. Information from wage records will be used to determine where workers are employed. The impetus for the comparisons programs is placement results, labor market analysis, the Carl Perkins Act and the Workforce Investment Act in order to improve the quality of information used to evaluate the success of graduates and training program participants and programs.

Security Measures

As noted in the Statement of Current Capacity, South Dakota's LMIC is in physically secured office with password protected work stations. The SLDS will be housed in the state computer system and meet all state and federal (FISMA-NIST) security and confidentiality requirements. SLDS files will only be available to individuals employed by the state LMI and BIT offices and access granted only to persons with documented need. Confidentiality of records for statistical analysis will be protected by adopting BLS standards for primary and secondary disclosure and any additional standards required by entities providing data to each SLDS.

As mentioned on page 10 within the description of database design, the DLR plans to use the platform and security standards established and in place in Wyoming. The SQL Server allows for

layers of user permissions and access controlled by the Database Administrator (DBA). Access can be granted to individuals at several layers of permission from the Database Owner, which allows universal rights to read only or no permission at all. In Wyoming, the SQL Server uses the SSN as the unique identifier and includes several other fields of personally identifiable information (PII). This model allows access to SSN, but restricts access to data sets where first and last names and addresses are available. The DLR will implement a gateway server design restricting SSN access in the SLDS to a gateway database managed by project DBAs. The gateway strips and archives PII from data sets as they enter the SQL Server. Each SSN is coded to a unique ID which will be archived on the gateway database. As data are brought into the gateway, the SSN will be replaced with a unique ID for previously archived SSNs and new unique IDs are assigned to new SSNs. The gateway also acts as a mechanism for assigning unique IDs via name, demographics and address to acquired datasets that do not have SSNs assigned. As SSN-based data are brought in, they will be archived on the gateway database for use to identify SSNs in current files, new SSNs which each need to be assigned unique IDs, and the assignment of SSNs to non-SSN based records.

Current data acquisition mirrors the technology available to data partners and may not meet the highest standards of security. Therefore, password protected/encrypted files containing partner PII including SSNs, wage records, UI account numbers, NAICS, location codes, residency designation and education enrollment will be transmitted to the SDLS over secure FTP sites.

Planned Report Deliverables

In addition to workforce outcome reports, other planned reports include graduate outcome annual reports on public postsecondary graduates by field of study, degree, demographics, earnings, industry employment, employer size and ownership. This information will also be displayed in the state WIA-eligible training provider list in order to provide user-friendly information to training participants, so they can select the programs that best suit their needs. Annual reports specific to the individual agency/institution will be available online to participating agencies and educational institutions in PDF or other electronic format. Summary data will be made available for agency/institution partners and will be offered in text file or other electronic format.

Returning Veterans

More extensive research will include the development of a process to identify and track returning veterans using the wage record and FEDES system. Information about how many veterans are expected to return to the state and how this will impact employment histories and college enrollment will be provided to partners in the workforce and education systems.

How Partner Data will be Incorporated

Employment outcome reports of training programs will be produced by the DLR and will provide key information about statewide aggregate performance of the workforce system to partners in BOR, DLR and DOE, as well as other state agencies and the general public. Specific aggregate outcomes addressed will include: graduates by major and degree, wages, industry, work location, retention, demographics and wage progression. All partners (DLR, DOE and the

BOR), continue to meet on a quarterly basis as members of an interagency governance council to develop plans for matching educational data to meet the intent of FERPA.

The enhanced knowledge about the impact of state workforce development programs and other program completers, in combination with the minimized reporting burden on state agencies and training providers, will ultimately create efficiency and cost effectiveness through quality information for frontline program management and public policy. These outcomes will be published in report form and electronically on the DLR website. These reports can be used by stakeholders to help steer curriculum development as well as assist workforce and education case managers in identifying programs with completers who typically earn higher-than-average entry-level wages. It will also help the general public and student populations make informed career decisions based on the comprehensive information provided through analysis of the SLDS.

Staffing Capacity

Database Administrator

South Dakota will have a Database Administrator to oversee the hardware/software systems. The Database Administrator is responsible for the performance, integrity and security of the SLDS. Specific responsibilities include: establishing the needs of users and monitoring user access and security; monitoring performance and managing parameters to provide query responses to front-end users; mapping out the conceptual design for a planned database; refining the logical design so that it can be translated into a specific data model; further refining the physical design to meet system storage requirements; maintaining data standards, including adherence to the Data Protection Act; writing database documentation, including data standards, procedures and definitions for the data dictionary (metadata); and controlling access permissions and privileges.

The Database Administrator must have experience as a database analyst and database administrator in order to fulfill these duties and responsibilities:

1. The ability to identify complex problems and review related information to develop and evaluate options and implement solutions
2. Critical thinking to identify the strengths and weaknesses of alternative solutions, conclusions or approaches to database design
3. Analysis of system design
4. Developing, managing and testing back-up and recovery plans
5. Additional requirements include planning, development and troubleshooting.

The database design will incorporate the following principles:

- Data are clearly defined.
- Data remains consistent across the database.
- Provision for data security and recovery control (all data is retrievable in an emergency).

Analyst Duties and Responsibilities

The primary responsibility of the DLR Senior Economic Analyst working on this project will be to serve in an advisory role as the SLDS is built, providing insight and guidance regarding the structure of the database and the reporting features of the system. Insight from the workgroup, including partners from the postsecondary governing body (BOR), DOE and DLR will also provide valuable input as the SLDS is developed.

The analyst will also assist in determining the methodology utilized to generate the statistical analysis to meet the objectives and deliverables described on pages 3 and 15 of this report.

The knowledge and abilities needed for the data analyst position includes knowledge of (1) economic theory and analysis and (2) mathematics and statistics methodologies, including modeling techniques. This position also requires the ability to conduct statistical analysis and present the information clearly and concisely to agency partners and the public. The types of previous experience which make a candidate highly qualified with planning, implementing and conducting analyses with a longitudinal data base include (1) experience administering databases, including a longitudinal database and (2) conducting longitudinal wage records research and analysis.

LMIC Administrator

The LMIC administrator, who will be responsible for administering the grant and overseeing this project, has 18 years of LMI experience. The DLR senior economic analyst will work under the direction of the administrator. The LMIC administrator is responsible for the administration of all BLS contracts, ETA grants and work accomplished under MOUs with other state agencies, other states and the Census Bureau (LED). The current administrator has experience working on the CES, LAUS, MLS and QCEW and ETA programs. The administrator will be a valuable resource to staff in accomplishing the Objectives and Deliverables indicated in this proposal on pages 3 and 15 of this document. With involvement in the management of several current commonplace LMI programs, the administrator has the knowledge, skills and abilities necessary to participate in the planning, implementation and analysis of the proposed SLDS. As the LMI

administrator will administer the grant, she will be working closely with the BIT staff as the SLDS is developed.

Employer for Each Proposed Staff Member

In South Dakota, the LMIC administrator and senior economic analyst on staff are employed by DLR.

As the SLDS will be developed and operated by BIT, BIT staff will be the architects of the longitudinal system. Expected technical staff includes a Database Administrator, Senior Software Developer and a Senior Programmer. In contracting with BIT staff, the LMIC will be charged approximately \$50-\$75 per hour. The rate may change based on the availability of BIT staff and the need to hire contract staff to assist. If this is the case, it is expected the hourly rate could increase to \$100 per hour. Over the three-year period, it is expected approximately 6,500 hours will be dedicated to this project, for a high-end estimate of approximately \$650,000.

Compliance with State and Federal Confidentiality Laws

In addition to 20 CFR 603 and other applicable federal regulations pertaining to the confidential nature of the records maintained by DLR, South Dakota Codified Law (SDCL) 61-3-4 explicitly protects certain employment records collected and maintained by DLR. Any employee or officer of DLR who discloses information deemed confidential under SDCL 61-3-4 is guilty of a Class 2 misdemeanor, punishable by up to 30 days in jail or a five hundred dollar fine, or both. In addition, BIT also requires all new state staff to complete an online training regarding basic security information, terminology and concepts.

Per a federal requirement stemming from the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), all state staff members who access confidential data are required to complete annual confidentiality training. As the principal data gathering agency of the federal government in the field of labor economics, the U.S. Bureau of Labor Statistics (BLS) requires this training once a year in order to maintain its integrity, confidentiality and trust with respondents and data users. All state staff members who work with the statistical data gathered and administered by the BLS are required to complete this training, which includes staff from the DLR and BIT. Therefore, all LMIC staff are sworn agents of the BLS under the CIPSEA and have signed confidentiality agreements.

In addition, all MOUs signed with other states include confidentiality clauses that specify all individually identifiable information will be held in strictest confidence by all partners to ensure individual privacy rights will be protected by strict adherence to the confidentiality provisions of the MOU. Partners agree the records used and created during the course of this comparison program will be maintained and safeguarded in such a manner as to restrict access solely to those individuals who have a legitimate need to use them in order to accomplish the program's purpose. Persons with authorized access to information are made aware of their responsibilities pursuant to the MOU. All information is exchanged in a confidential manner and is to be used solely for statistical purposes.

Other Data Linkages

In addition to the workforce data sources mentioned above, South Dakota has access to other data sets that can enhance its analysis and reporting on the state's workforce and education programs. South Dakota has the necessary agreement in place to participate in the voluntary Wage Record Interchange System (WRIS) 2. The agreement was signed in October 2011.

As mentioned earlier in the grant proposal on pages 2 and 12, additional data sources LMIC currently has access to include participants in the following programs and grants:

- AEL (Adult Education and Literacy)
- TANF (Temporary Assistance for Needy Families)
- SNAP (State Nutritional Assistance Program)
- NEG (National Emergency Grant)
- Other USDOL discretionary grants including
 - DEI (Disability Employment Initiative)
 - SESP (State Energy Sector Partnership Grant)