

**New Jersey Technical Proposal – SGA/DFA PY-11-01**  
**Workforce Data Quality Initiative**

**STATEMENT OF CURRENT CAPACITY—CURRENT WLDS**

The New Jersey Department of Labor and Workforce Development (LWD) has a **partial** workforce longitudinal data system (WLDS). This proposal outlines a vision for a fully-developed WLDS that is aligned with the overall workforce development goals of the Department, called the “New Jersey Talent Connection Initiative.” These goals, set forth in LWD’s Workforce Innovation Fund proposal, include (1) focusing workforce investments in key industry sectors of the state economy; (2) creating a critical foundation of workforce preparation aligned with employer skill demands; and (3) integration of unemployment insurance and re-employment efforts. The goals are to be achieved through a sustained emphasis on program evaluation and performance-based metrics. A robust WLDS is critical to the successful implementation of more effective re-employment and talent development in the state.

As it enhances its WLDS, LWD will work closely with the John J. Heldrich Center for Workforce Development in the Edward J. Bloustein School for Planning and Public Policy at Rutgers, The State University of New Jersey (Heldrich). For over ten years, LWD has contracted with the Heldrich Center to maintain the current, partially-developed WLDS. As an academic research institution employing seven faculty and staff with PhDs and broad training in research methods, the Heldrich Center has extensive data management and analytical capacity. Heldrich staff members are uniquely qualified to lead the move toward a fully-developed WLDS, not only because of their partnerships with New Jersey education agencies and their longstanding role in the LWD’s current WLDS, but also because of their substantial research and evaluation portfolio. Since 1998, the Heldrich Center at Rutgers has been authorized by contracts and data sharing agreements with LWD, the New Jersey Department of Education (NJDOE), and New Jersey Higher Education (NJHE, previously the Commission on Higher Education) to create a

workforce longitudinal data system and to use the data in it to calculate labor market outcomes for all individuals in WIA Title I-B occupational training programs, all adults served under WIA Title II, individuals participating in post-secondary adult vocational education programs under the Carl D. Perkins Act, graduates of the state's public colleges and universities, and individuals enrolled in private training programs on the Eligible Training Provider List (ETPL) in the state. Since 2005, LWD and NJDOE have partnered to jointly fund (at \$59,000 and \$14,000, respectively) the WLDS at Heldrich.

The primary purpose of LWD's current WLDS is to report on the employment outcomes of all training providers in the state and to allow Heldrich, which has substantial data analytic capacity, to conduct (with LWD's approval) quasi-experimental evaluations of workforce programs. LWD's current WLDS is limited, however, because (a) it does not include any data from some critical workforce data systems, e.g., UI claims, Wagner-Peyser, Trade Adjustment Assistance, and Vocational Rehabilitation; (b) some data files in the WLDS, such as data on WIA trainees, are missing key data fields, including basic demographic information; (c) it contains no data on the P-12 system; and (d) data in the WLDS is used only to a limited extent for performance reporting and detailed data analysis. Under the WDQI grant, New Jersey intends to increase the data in its WLDS through a strengthened partnership with Heldrich in order to allow for more comprehensive reporting and more robust data analysis and program evaluation.

#### Design of Current WLDS

The WLDS has evolved since its inception in 1998 and currently includes New Jersey Unemployment Insurance Wage Records, WIA Title I-B completers of occupational training programs, WIA Title II, Carl D. Perkins Act adult vocational education program completers, NJHE enrollees and graduates, as well as data on the completers of all private training programs

on the Eligible Training Provider List.

The WLDS is designed as a series of flat files, all of which contain social security number (SSN). The WLDS uses SSN—and only SSN—as the unique personal identifier to link all files. Nonetheless, many of the data sets, including WIA Title I-B, include name and date of birth. NJ Unemployment Insurance (UI) wage record data contains employees' full names. There is, therefore, the potential to link these current WLDS files to data files generated from other state data systems, such as the statewide longitudinal P-12 data system, that contain name (and date of birth) but not SSN. Some of the data files in WLDS are incomplete in the sense that they do not contain all of the data fields from these systems. We indicate below the data files for which we have a complete and incomplete set of fields.

- *UI Wage Record Data.* UI wage records include the following information for individuals who are employed: employee SSN, employee name, quarterly earnings paid, number of weeks worked during the quarter, tax identification number of the employer, and industry of employment (in the “employer” file). New Jersey UI wage records contain more than 4 million records per quarter. LWD extracts a flat file from the UI wage record system and transmits it to Heldrich via secure file transfer protocol (sftp) every quarter. The WLDS contains UI wage record data, dating from the first quarter of 2000 to present. Completeness of file: *Complete*.
- *WIA Title I-B Training Completers.* NJLWD maintains WIA Title I, Wagner-Peyser, Trade Adjustment Assistance Act, state-funded training programs (e.g., Workforce Development Partnership Program), Reemployment Eligibility Assessment, Migrant Seasonal Farm Workers, Veterans Services (DVOP and LVER), National Emergency Grants, and WorkFirst New Jersey (General Assistance, Temporary Assistance to Needy Families (TANF) and Supplemental Nutrition Assistance Program (SNAP)) data in the America’s One-Stop Operating System

(AOSOS). Although AOSOS includes service information on clients participating in a wide variety of workforce programs as well as comprehensive demographic and personal history data, LWD sends Heldrich only the following information on WIA Title I-B training recipients: SSN, training start date, training end date, field of training, and the One-Stop Career Center that referred the client to training. A flat file is extracted from the data and transmitted to the Heldrich Center via sftp every quarter. The WLDS contains this data from the third quarter of 2007 through the fourth quarter of 2011. Completeness of file: *Incomplete*.

- *WIA Title II Data*. NJLWD is responsible for calculating outcomes of WIA Title II Adult Education programs. The data contain detailed demographic information, including race, gender, disability status, and education level as well as the literacy-related outcomes. SSN is the unique personal identifier in the file, which also includes date of birth. Each quarter, Heldrich accesses a secure server maintained by an NJLWD contractor and downloads a flat file via sftp. The WLDS contains this data from the third quarter of 2010 through the second quarter of 2011. Completeness of file: *Complete*.
- *Perkins Act Data*. For the past seven years, NJDOE has contracted with the Heldrich Center to calculate employment outcomes for individuals who complete Perkins Act adult vocational education programs at New Jersey high schools and postsecondary educational institutions. NJDOE collects data on students in adult vocational programs through its Vocational Education Data System (VEDS), a student-level data system that tracks adults who concentrate in a vocational-technical field. Data fields in VEDS include SSN and classification of instructional program (CIP), as well as demographic information. Heldrich receives the VEDS data via email in a password-protected zip file once per year. Heldrich has VEDS data for the academic years ending in 2002 through 2011. Completeness of file: *Incomplete*.

- *Higher Education Student Unit Record Data.* NJHE maintains the Student Unit Record (SURE) data system, which warehouses detailed information from all 31 public (and nine independent) two and four-year colleges and universities in the state on students who enroll in and complete degree programs. The SURE data file includes student SSN, basic demographic information, date and program of enrollment, date of completion, degree or certificate awarded, and program of study (by CIP). For students who attended high school in New Jersey since 2009-2010, the SURE data files have included the unique personal identifier from NJDOE's Statewide Longitudinal Data System (the NJ SMART ID). NJHE sends Heldrich this data annually via email in a password protected zip file. Heldrich has in its possession, SURE data from the 2006-2007 academic year through the 2010-2011 academic year. Completeness of file: *Incomplete.*
- *New Jersey Consumer Report Card (NJCRC).* All training providers on the state's ETPL must report student information for their students and trainees. The NJCRC includes student SSN, demographic characteristics, program of enrollment, and dates of enrollment and exit from programs. However, because only some fields—SSN, enrollment date, exit date, completion status, and CIP code—are mandatory, most demographic information is missing. An NJLWD contractor maintains a secure server that Heldrich accesses quarterly and downloads an enrollment flat file and an exit flat file via sftp. The WLDS contains NJCRC enrollment and exit data from the November 2001 to present. Completeness of file: *Complete.*

#### Capacity for Maintaining Data Security

LWD has contracted with the Heldrich Center to maintain the WLDS in some form since 1998. Heldrich has significant experience with maintaining a high level of confidentiality and security of the files in the system. The majority of the data that Heldrich receives is transmitted

via sftp, which encrypts the data during transfer. Data received from NJHE and NJDOE are transmitted in password-protected zip files.

Heldrich maintains extensive electronic and physical security protections to safeguard the confidentiality of the data in the WLDS. Data is downloaded electronically via sftp using a dedicated computer that can only access (i) the secure servers from which the data are downloaded via sftp and (ii) the Workforce Development Research (WFDR) secure server where the data are stored. The dedicated download computer is a “freeze” computer that securely deletes working files every time the computer is shut down or rebooted. Only two computers can access the WFDR secure server. The WFDR secure server sits behind three firewalls and the data is stored in an encrypted volume. All data analysis takes place via remote client on the WFDR secure server, which is loaded with all necessary data management and analysis software (e.g., SAS, Stata, R). Confidential data are never stored on personal computers.

As for physical security, the server is stored in a locked room on the Rutgers University campus within a computer lab that requires key card access, and the lab is inside a building with monitored access and that is locked at night and on weekends.

#### Results / Reporting from Current WLDS

The WLDS helps LWD meet some of its reporting obligations, and the reports produced from WLDS data are designed to enhance job seekers’ ability to make career decisions. WIA requires LWD to maintain and set standards for the state’s Eligible Training Provider List (ETPL). To inform jobseekers’ decisions, LWD contracts with Heldrich to analyze WLDS data to calculate the employment rate, retention rate, and average earnings of training providers on the ETPL (as well as adult vocational and higher education programs). These calculations are performed by merging UI wage records with data on the following: (1) WIA Title I-B occupational training

completers, (2) Title II Adult Education program participants, (3) Perkins adult vocational education program graduates, (4) two and four-year public and independent college graduates and (5) private training provider program completers.

These calculations support the generation of a number of reports. First, the employment results for all training providers on the ETPL are disseminated to the public via the NJCRC web site, [www.njtopps.org](http://www.njtopps.org). Labor market outcomes are shown at the program level, the cluster level and the provider level for the first, fourth, and eighth quarters after program completion. Second, quarterly reports for frontline One-Stop Career Center staff are produced to assist them in helping customers select an appropriate training provider. Separate reports are generated for each of the state's 19 WIBs that detail the employment outcomes that their customers realized after completing the training programs to which their WIB referred them. These localized results assist the WIBs in setting performance benchmarks that training providers must meet in order to receive customers with Individual Training Account (ITA) vouchers, and they allow One-Stop staff to gauge the performance of the training providers in their area so they can better assist their customers in making career and training decisions.

The current WLDS generates performance reports that assist NJDOE in meeting reporting requirements under the Perkins Act. In the course of calculating outcomes for the NJCRC, Heldrich also calculates employment outcomes for NJDOE's adult Perkins students. Heldrich returns these results to NJDOE so they can be reported to USDOE.

With explicit written permission from LWD consistent with the terms of the data sharing Memorandum of Understanding between the two organizations (see the "Attachment to the Technical Proposal"), LWD permits the Heldrich Center to conduct quasi-experimental evaluations of federally and state-funded workforce development programs. Using data in the

WLDS Heldrich, has conducted a number of such evaluations, such as a quasi-experimental evaluation of New Jersey's (state-funded) Workforce Development Partnership (WDP) program. For that evaluation, LWD provided Heldrich with data on WDP participants as well as UI claims data so that researchers could use probabilistic matching techniques to select a comparison group. Heldrich calculated employment and earnings outcomes for WDP participants and the comparison group of similar individuals who did not participate.

LWD also gave permission to Heldrich researchers to use WLDS data to conduct a quasi-experimental evaluation of a state-funded (LWD and the NJ Schools Construction Corporation) pre-apprenticeship program. The comparison group was selected from WIA Title I-B training completers, and outcomes, including earnings growth, were calculated for program participants and the comparison group using UI wage record data. Currently, LWD is permitting Heldrich to use UI wage data and WIA Title I-B training completer data to create a comparison group and calculate outcomes for the TAACCCT-funded programs for the state's TAACCCT grantee, Raritan Valley Community College.

The Heldrich Center's quasi-experimental evaluations are not conducted on a regular basis and depend on outside funding to support the research. LWD anticipates that the expansion of WLDS, along with WDQI funding and other support after the end of the grant period, will allow program evaluations to be conducted for the state on a more regular and systematic basis, as the proposal explains below.

#### Partnerships with Other States

New Jersey has built relationships with other states for the purpose of sharing UI wage record data. New Jersey is currently a member of The Regional Area Data Exchange (TRADE), which is a system for the quarterly exchange of UI wage record data among five additional states

(Pennsylvania, Ohio, Maryland, Virginia, and West Virginia) and the District of Columbia.

Although this data is not currently incorporated into WLDS, LWD will look to incorporate it, to the extent permitted by the TRADE data sharing agreement.

New Jersey is also a member of the Wage Record Interchange System (WRIS) and in November 2011 signed on to WRIS2. New Jersey has access to data on individuals employed by the federal government through the FEDES system; however, it is not currently used. Both of these sources will be investigated further for inclusion in the expanded WLDS.

## **PLAN OUTLINE**

### Objectives for WLDS and Plan for Achieving Objectives

The overarching goal of the expansion and overall improvement of New Jersey's WLDS is to support priorities consistent with the New Jersey's Workforce Innovation Fund proposal (the New Jersey Talent Connection Initiative), which is also the foundation of the state's Unified State Plan. Specific priorities can be linked directly to the larger initiative.

(1) **Performance Metrics.** As part of the governor's performance budgeting and accountability initiative as well as for improved program management, LWD is implementing a balanced scorecard approach to support better metrics for all workforce programs. The metrics go beyond the Common Measures and will also inform the State Employment and Training Commission's (state WIB) accountability initiative. Expanding WLDS will allow for timely reporting of the new metrics and may result in additional measures for the current rubric.

(2) **Evaluation.** Improved and increased evaluation of workforce programs is, combined with metrics, a critical element for learning more about our ability to serve clients, particularly the hard to serve populations. The expanded WLDS will facilitate quasi-experimental evaluations of workforce programs. This will be especially helpful for intense interventions of special

populations such as the long term unemployed and youth.

(3) **Improved integration of UI and workforce programs.** There are 400,000 New Jerseyans unemployed and many who are likely to exhaust their UI benefits. The expanded WLDS would give the state the ability to revisit its profiling method to identify individuals who are likely to exhaust their benefits. LWD could use Heldrich’s ability to develop an improved profiling method using machine learning techniques.

(4) **Sector talent development and stackable credentials.** Merging education and workforce data will enable LWD to learn important things about career pathways and understand the flow of workers in and out of the labor market. LWD is committed to targeting resources to specific sectors of the state’s economy, including through the Talent Networks, which focus on the industries in order to connect employers, job seekers, the state’s One Stop Career Centers and educational institutions. The WLDS will enable analysis to support talent development in specific sectors.

All of these elements relate to one of the core missions for both LWD and NJDOE: to be transparent and accountable to New Jersey taxpayers.

In order to realize these research and reporting outcomes, LWD has a number of objectives for upgrading the WLDS. The first objective is to have all of the workforce data that the Department collects—UI wage, UI benefit, WIA, Wagner-Peyser, TAA, Vocational Rehabilitation, Workforce Development Partnership Program—linked in a single system (by SSN). Although UI wage and some WIA data are currently in WLDS, as indicated earlier, the system is incomplete. Having all of these data systems linkable is the necessary first step toward being able to conduct quasi-experimental evaluations of workforce programs – and to produce meaningful and user-friendly reports for each of LWD’s key audiences—job seekers, One-Stop

staff, employers, program managers, and public policy makers. To facilitate the merging of data files, this step will also consist of ensuring that all data files in WLDS contain as complete personal identifier information—SSN, full name, date of birth—as is available.

Second, LWD will work with NJHE to obtain data prior to the 2006-2007 academic year for the WLDS. NJHE has collected data in the same format dating back to 1990, and the long time series will allow for long-term analyses of the labor market experiences of college enrollees and graduates.

Third, LWD seeks to incorporate NJDOE SLDS data into its system in a manner that is consistent with the Family Educational and Privacy Rights Act (FERPA). To date, LWD has not maintained any NJDOE identified student-level data. Indeed, this is part of the reason why LWD originally formed the partnership with the Heldrich Center, so that Heldrich could house NJDOE data and calculate Perkins outcomes for NJDOE. Once the SLDS data is in WLDS, the critical task will be to successfully link it to the other data systems. Because the SLDS does not contain SSN, Heldrich will work with LWD, NJDOE, and NJHE to develop a system to match SLDS data with workforce data. The first step will be to match all records in the NJHE SURE data that have an NJ SMART identification number (SID) with their records in the SLDS. NJHE data has SSN as an identifier, so these individuals will be able to be linked with workforce data. However, because NJHE has SIDs for the past three years of students and because many students do not enter post-secondary education or attend college out of state, LWD and Heldrich will also work with the state agencies to link SLDS and workforce datasets using full name and, when available, date of birth. To facilitate the matching, Heldrich will use LinkPlus, probabilistic record linking software developed by the US Centers for Disease Control.

Fourth, once the data have been matched, LWD intends to have Heldrich generate reports

for various audiences and conduct the data analyses that the proposal describes in the section after next. A detailed list of project activities appears on the last page of the Technical Proposal.

### Statewide Longitudinal Data System (SLDS)

New Jersey's SLDS is "NJ Standards Measurement and Resource for Teaching (NJ SMART), a comprehensive data warehouse, student-level data reporting, and unique statewide student identification (SID) system" (NJDOE, 2012). The unique personal identifier in the NJ SMART system is the NJ SMART identification number (SID). Since 2007, NJDOE has assigned an NJ SMART ID to more than 1.8 million students, 1.4 million of whom are currently in the system. As of December 2011, NJ SMART had fully implemented ten of the twelve elements from the America COMPETES Act and had partially implemented elements #8 and #9 – creating the student-teacher link and collecting course and transcript data – which are due to be completed in Fall 2012.

The SLDS now collects more than 150 student-level data elements through a web-based portal from over 650 school districts and charter schools for every student from pre-Kindergarten through adult high school. Among those data elements are information about enrollment, graduation, transfer, dropout, special education status, and a host of demographic information. Additionally, via the unique student identifier, NJDOE has connected this data to the Department's statewide annual assessments (including regular assessments, alternate assessments, and English Language Learners assessments), garnering information about student achievement outcomes, students not tested, and relationships between and among student demographic variables and special education classifications and placements.

### Sustainability

The WLDS will be a cornerstone of LWD's efforts to strategically target its resources to only the

best and most effective workforce programs. When the grant ends, therefore, it will devote its own resources to support the continuation of the WLDS.

LWD and NJDOE have sustained the WLDS through a joint partnership to fund Heldrich to calculate performance outcomes for the NJ Consumer Report Card (NJCRC). Because the scope of the WLDS will expand dramatically as a result of the work LWD is proposing under WDQI, the funds that LWD and NJDOE currently allocate to support the NJCRC are insufficient to support the expanded WLDS after the WDQI grant ends.

LWD intends for the WLDS to function as a more robust, comprehensive system for generating the reports that LWD needs to produce in order to better evaluate programs and produce performance metrics, particularly metrics in excess of the Common Measures. These capabilities will generate more comprehensive reporting to facilitate enhanced decision making. LWD intends to sustain the WLDS beyond the three years of the WDQI grant by continuing to use the funds it currently allocates to maintain the NJCRC. In addition LWD will target funds from workforce programs that have reporting and / or evaluation requirements and pursue innovative opportunities such as financial support from foundations and outside researchers who have previously funded projects or data requests (see Attachments for a letter of support from the Nicholson Foundation). The WLDS will be key for both mandatory reporting requirements and performance metrics. These metrics are part of a statewide initiative focused on performance-based budgeting. At LWD, a balanced scorecard approach to metrics has been developed in order to supplement data from the Common Measures. These additional forces underlay how critical it is to sustain the WLDS. Therefore, at a minimum, this combined amount would support a continued partnership with Heldrich to maintain and manage the data in WLDS, as well as rigorous evaluation of LWD programs. Responsibilities for additional reports and analyses

would be undertaken by LWD staff or Heldrich, depending upon resources available and the level of specialized expertise needed

## **PARTNERSHIP STRATEGIES**

Although New Jersey's current WLDS is limited in the breadth of the data that it houses, LWD has established solid partnerships with state agencies that will provide a foundation for expanding WLDS and the scope of the partnerships. As noted previously, the current WLDS is a joint partnership between LWD, NJDOE, and NJHE. Through the WDQI grant, LWD will expand these partnerships in order to increase the breadth of the data in WLDS and the breadth and the depth of the reporting and analysis that the WLDS produces.

The WLDS will have a Data Provider Committee that will include state agencies providing data to the WLDS. Participants will include LWD (including its Divisions of Labor Planning and Analysis, Vocational Rehabilitation Services, and Information Technology), the State Employment and Training Commission (the state WIB), NJDOE, NJHE, the NJ Department of Human Services (which has agreed to provide Supplemental Nutrition Assistance Program (SNAP) data), the NJ Office of Information Technology (NJOIT). The committee will meet quarterly and will establish policies for sharing data, particularly in response to outside requests.

There will also be a WLDS Steering Committee, which will be chaired by the Assistant Commissioner for Labor Planning and Analysis at LWD. It will also meet quarterly and include consumers and users of the WLDS, such as WIBs, community colleges and other postsecondary institutions, and foundations. The institutional research committee of the New Jersey Community College Consortium has already begun meeting with LWD to discuss employment outcomes for their student success model. In addition, the Nicholson Foundation, which has

previously sponsored studies using workforce data, has agreed to serve on the committee. The Steering Committee will recommend data analyses to be conducted using WLDS data, identify additional data items to be collected, discuss the results of data analyses and reports, and develop strategies for disseminating findings.

#### Partnerships within State Workforce Systems

LWD will be the central leader in advancing and overseeing the development of the WLDS. Within the Department, the Division of Labor Planning and Analysis (LPA) will be the lead unit in the implementation. As a first step, LPA will be responsible for writing data sharing agreements permitting Heldrich access to workforce data. LWD, as a whole, will be responsible for sharing data dictionaries of its data systems, working with Heldrich to identify the data fields that should be included in the flat files that it generates for WLDS, creating flat files from its UI wage, UI claims, WIA Title I, Wagner-Peyser, Trade Adjustment Assistance, and Vocational Rehabilitation data, and setting and obtaining input from key stakeholders on the research agenda. LPA will take the lead in working with the different units within the Department, including Unemployment Insurance (UI) and Vocational Rehabilitation Services (DVRS), to accomplish all of these tasks, including generating the flat files to be transmitted to Heldrich. In this process, LPA will work closely with the Department's Division of Information Technology (DIT) to establish automated processes for generating the required flat files on a regular basis, e.g., quarterly. LPA will also coordinate LWD's efforts to ensure data quality with UI and DVRS.

Finally, DIT will take the lead in, and will have ultimate responsibility for, ensuring the security of the data in WLDS. DIT will establish procedures for secure data transfer and will be responsible for monitoring, and conducting audits of, Heldrich to ensure that the highest level of data security is being maintained.

## Partnerships with State Education Agencies

LWD recognizes the value of relationships with secondary and postsecondary education, as workers enter and exit the labor force to and from educational institutions. The Department continues to pursue innovative partnerships, most recently partnering with the state's 19 community colleges to deliver services for the USDOL Reemployment and Eligibility Assessment. In addition to relationships with individual institutions, LWD has strong partnerships with NJDOE and NJHE. LWD currently partners with NJDOE to fund and provide the data for Heldrich to calculate employment outcomes for participants in WIA and Perkins-funded training programs. This partnership will continue after the awarding of the WDQI grant.

NJDOE is the State Education Authority and maintains New Jersey's SLDS, NJ SMART, while NJHE maintains the higher education unit record data systems. Through NJDOE's most recent SLDS proposal to USDOE, LWD has strengthened its working relationships with NJDOE and NJHE. In that proposal, NJDOE committed to working with LWD to match students in NJ SMART with workforce data. As FERPA requires student confidentiality to be strictly maintained, LWD, Heldrich and NJHE are in the process of negotiating Memoranda of Understanding, i.e., data sharing agreements, with NJDOE to allow Heldrich and NJHE to have appropriate access to data in NJDOE's SLDS. The intention is for NJHE, NJDOE, and LWD to create a partitioned but shared data warehouse of P-12 data along with the higher education data, in a P-20 longitudinal data system. Under the terms of the MOU being negotiated, Heldrich will receive the P-20 longitudinal data extracts. The scope of the data that will be available, in terms of grade levels and data fields, is being negotiated.

Before any data transfer occurs NJDOE, NJHE, LWD, and Heldrich will work together to develop data extract and transmission procedures between NJDOE and NJHE and between

NJHE and Heldrich.

NJDOE and NJHE will also play central roles on the WLDS Data Providers Committee.

#### Partnerships with State Research Universities

LWD and the Heldrich Center have collaborated on projects throughout the center's 15 years, including evaluating the Department's workforce programs, developing online career awareness tools for high school students, and developing profiles of key industry sector skill needs. It is based on this history of collaboration as well as Heldrich's resources as a university research center that LWD decided to form a partnership for the expanded WLDS. Heldrich currently has a Memoranda of Understanding for data sharing with LWD and NJHE. Both agreements require that Heldrich treat all data containing personally identifiable information as highly confidential. The MOU specify that the data must be stored only on a server that is highly secure from unauthorized access. Both MOUs prohibit Heldrich from disclosing data to anyone outside of Heldrich and set forth specific purposes for data use. The MOUs permit Heldrich to use the data for research purposes as agreed state agency that owns the data.

Before starting any research project, Heldrich submits an application describing the research procedures to Rutgers University's Institutional Review Board (IRB). These research projects have typically been classified as "Expedited" and all of them have received IRB approval, in some cases with minor revisions. Under WDQI, Heldrich will continue to seek IRB approval for individual research projects. In addition, it will seek IRB approval at the outset for hosting the data in the expanded WLDS.

LWD has identified a number of roles for Heldrich in this continued partnership. Heldrich will be responsible for (a) maintaining the WLDS, by helping to specify requirements for data extracts (flat file generation); (b) working with NJDOE, NJHE, and LWD (and other partner

agencies) to ensure that secure file transfer procedures are implemented; (c) receiving, processing, cleaning, merging, and managing the data in the WLDS; (d) conducting data analyses and generating reports to facilitate enhanced decision making.

#### Partnerships with Additional State Agencies

LWD will also work with the NJ Department of Human Services (NJ DHS) and Heldrich to negotiate a Memorandum of Understanding between Heldrich and NJ DHS in order for Heldrich to receive access to Supplemental Nutrition Assistance Program (SNAP) data.

### **DATABASE DESIGN, DATA QUALITY ASSURANCE, AND PROPOSED USES**

#### Personal Identifier

The current WLDS is a series of flat files linked by SSN. The expanded WLDS will retain this system design after the receipt of the WDQI grant. SSN is currently and will continue to be the personal identifier in the WLDS. This will, of course, present a challenge for linking (matching) records obtained from NJDOE, which do not have an SSN, with records from the workforce system. Heldrich will assign SSNs to records in the NJDOE SLDS data using the following procedure. First, because the NJHE data will contain the NJ SMART ID, Heldrich will first match NJDOE SLDS data and NJHE data using the NJ SMART ID. Students who are matched will then be assigned the SSN that is in the NJHE data.

For records in the SLDS that do not appear in the NJHE data, Heldrich will match SLDS records with workforce records using name and, when available, date of birth. As a first step in the process, Heldrich will conduct an initial assessment of the completeness of the name and date of birth fields in all data files. Record linkage would then be performed using some of the methods, such as probabilistic linkage, described in Stevens (2012) and Mulrow et al. (2011). Heldrich will likely use the probabilistic record linkage software LinkPlus, which was developed

by the CDC, to facilitate the linking of records, and which can accommodate misspellings by allowing phonetic matches. Heldrich will develop a set of rules for linking NJDOE SLDS data and workforce data and then merge the data files. For any record in the NJDOE data that are matched on name and, when available, date of birth with records in the workforce data files, an SSN will be assigned to that record and the associated names and dates of birth in both files will be deleted.

LWD and Heldrich will work to establish whether SSNs in the data files are consistent with SSA's social security number issuance rules. In order to establish the validity, the Heldrich Center will develop checks on match quality by taking advantage of all of the data in WLDS -- for example, by ensuring that an SSN assigned through the matching process to an 18 year-old, 2009 high school graduate did not appear in the wage record data before 2006.

#### Data Systems Not in Current WLDS that Will Provide Data to Expanded WLDS

Under the WDQI grant, WLDS will be expanded in two ways. First, the data files that currently reside in the WLDS but are missing some important fields will be replaced with the complete versions of those files. Of the LWD data systems that contribute data to the WLDS, Heldrich has complete data from four of the six files in the current WLDS: UI wage (since 2000), NJDOE data for individuals who completed adult vocational education programs (since 2001), private training provider data (since 2002), and WIA Title II (since 2010). The files that are missing key fields and will therefore be replaced include the data on WIA training recipients and NJHE student record data.

Second, in addition to the data systems described earlier in the "Current Capacity" section that supply data to the current WLDS, under an expanded WLDS, the following data systems will also contribute data to the WLDS on a regular basis. At the start of the project, LWD and

Heldrich will work with the owners of each of the databases to obtain data dictionaries and identify the data elements in them that should be included in the flat file extracts for the WLDS.

- *UI Claims Data.* New Jersey's UI data system (LOOPS) is maintained by NJOIT, which extracts UI claims data on a weekly basis and provides it to LWD. Under WDQI, LWD and Heldrich will ensure that the extract includes full name and date of birth. NJOIT will also transmit the extract that it currently provides LWD to Heldrich on a weekly basis.
- *Wagner-Peyser, WIA Title I, and Trade Adjustment Assistance Act Data.* On a quarterly basis, LWD currently sends Heldrich limited data on individuals who receive training through WIA Title I, but it does not send Wagner-Peyser or TAA data. LWD stores data on all three of these programs in the America's One Stop Operating System (AOSOS). As indicated earlier, AOSOS uses SSN as the personal identifier and collects basic demographic data as well as data on the services received by the customer and the dates of service. Under the WDQI, LWD will generate flat file extracts from the AOSOS Oracle database for transmission on a quarterly basis to Heldrich for inclusion in the WLDS. LWD will also transmit files from previous quarters dating back, at a minimum, to the first quarter of 2007.
- *Division of Vocational Rehabilitation (VR) Data.* VR data is not collected by AOSOS, but VR includes similar data elements, such as demographics, services received, and dates of service, as well as each individual's documented primary and secondary disabilities. Data from its Oracle database will be extracted into a flat file for transmission to Heldrich every quarter. LWD will also transmit files from previous quarters dating back to the first quarter of 2007.
- *NJDOE SLDS: New Jersey Standards Measurement and Resource for Teaching (NJ SMART).* Since FERPA mandates strict protection of student confidentiality, LWD, Heldrich and NJHE are in the process of negotiating Memoranda of Understanding, i.e., data sharing agreements,

with NJDOE to allow Heldrich and NJHE to have appropriate access to data in NJDOE's SLDS. The intent is for NJHE, NJDOE, and LWD to create a partitioned but shared data warehouse of P-12 data along with the higher education data, in a P-20 longitudinal data system. Under the terms of the MOU being negotiated, Heldrich will receive the P-20 longitudinal data extracts. The scope of the data that will be available, in terms of grade levels and data fields, is being negotiated.

- *Supplemental Nutrition Assistance Program (SNAP) Data.* Data on the SNAP program are a longitudinal record of individuals' receipt of supplemental nutrition assistance. The data include social security numbers, benefit amounts, and basic demographic information.

#### Data Discovery Process

Before adding data to the WLDS, the Heldrich Center will do a complete inventory of the datasets and data fields in the current WLDS. Heldrich will then create a data dictionary for each of the datasets within WLDS, which it will share with LWD and partner agencies to identify data files and fields missing from the system. LWD and each state agency will share the data dictionaries for each of the data files that they will contribute to WLDS.

#### Data Quality Measures

LWD will take the lead in checking data quality. The Analysis and Evaluation section within Labor Planning and Analysis conducts annual Employment and Training Administration (ETA)-mandated data validations of the Unemployment Insurance, Workforce Investment Act, Trade Adjustment Assistance, and labor exchange programs. The purpose of these validations is to ensure the accuracy and reliability of reportable data and ensure data quality standards are being met. The validation processes operate within the general framework and schedules established by ETA. The general structure of a typical validation includes: report verification;

sample selection and document review for the quality portion (the document review process involves field visits to review procedures and required documentation); data processing and review; transmitting validation information to ETA; and, communicating the validation findings to the program areas. These validation efforts are designed to strengthen the workforce system by ensuring that accurate and reliable information on program activities and outcomes are produced and reported. These existing efforts will be a critical component of the WLDS.

The matching of workforce data with education data by name and date of birth will also be subject to extensive data quality checks as described earlier in the Personal Identifier section.

### Security Measures

Heldrich maintains extensive electronic and physical security protections to safeguard the confidentiality of the data in the WLDS -- from data transmission, storage, and management, to data analysis and data sharing.

*Data Transmission Security.* Heldrich downloads sensitive and confidential data via sftp using a single dedicated computer that can only access (i) the secure servers from which it downloads confidential data via sftp and (ii) the Workforce Development Research (WFDR) secure server where Heldrich stores all sensitive and confidential data. This “freeze” computer securely deletes working files every time the computer is shut down or rebooted.

*Data Storage Electronic Security.* The WFDR secure server is secured at a number of different levels. First, the machine is protected by a local area network-based firewall that resides on Rutgers’ private space. Only machines at Rutgers can reach the firewall where this system resides; access is limited to the explicitly specified systems for SFTP and Remote Desktop. In the event someone did get through the firewall, the system’s host-based security settings only allow these two programs to be seen. Only one other production host is behind the firewall where

this system resides. The data itself is also maintained in an encrypted volume. Therefore, system intruders cannot access the data without an encryption password. Only two computers can access the WFDR secure server.

*Data Storage Physical Security.* The server is stored in a locked room on the Rutgers campus, within a computer lab that requires key card access. Only the Director of Information Technology for the Bloustein School (the school to which Heldrich belongs) has a Medco server room key -- which cannot be duplicated. The computer lab, in turn, is inside a building with monitored access and that is locked at night and on weekends.

*Data Analysis Security.* All data analysis takes place on the WFDR secure server via remote client access. All data analysis software (SAS, Stata, and R) sit on the WFDR secure server. Heldrich staff access these programs remotely to process, clean, manage, and analyze data. Individual-level data never leave the WFDR secure server during the course of data analysis; only aggregate results are saved from the server to a personal computer and disseminated.

*Disaster Recovery.* The hardware for the WFDR secure server is a Precision 390, so there are many spare parts in case of hardware failure, and the hard drive was cloned after the build to create a fully functional backup hard drive. An externally connected USB drive is also needed for ongoing backups. If there was a hard drive failure, the spare would be inserted into the machine and data would then be restored from the USB drive to the data drive on this system. Truecrypt will connect to that data drive before copying files over.

*Data Sharing Security.* Because the Heldrich Center only has client software, it is only able to securely transfer (upload and download) files to the state's secure network. As with data downloads, these data uploads occur via sftp. To ensure compliance with FERPA, Heldrich will only share data back to the disclosing state agency. At the direction of LWD, however, Heldrich

will prepare and share with LWD (via sftp) analysis data sets created from data that LWD supplies to Heldrich that LWD can use to generate its own customized reports and conduct separate analyses.

### Planned Reports and Deliverables

As explained previously in the “Plan Outline” section, LWD is interested in expanding WLDS in order to conduct additional analysis and reports in support of the overall vision of the New Jersey Talent Connection Initiative outlined in the state’s proposed Workforce Innovation Fund grant application. The expansion will result in the capacity to conduct quasi-experimental evaluations of workforce programs, the potential to improve the models that it uses to identify individuals who are likely to exhaust their UI benefits, improve the timeliness, relevance, and usefulness of the reports and information that it disseminates for the use of job seekers, and use education data to study the labor market experiences of disadvantaged groups, including high school dropouts and people with disabilities.

LWD will charge the Heldrich Center with the responsibility of using the data in the WLDS to conduct the following studies. The studies are organized in the chronological order in which LWD will expect Heldrich to conduct them. In addition, specific connections are made to how these studies support larger LWD initiatives.

1. Quasi-experimental evaluation of WIA Title I training programs. Heldrich will use data from Wagner-Peyser and use probabilistic matching techniques to select a comparison group of individuals who are as similar as possible to the individuals who received ITA vouchers and use difference-in-difference models to compare the wage growth of the two groups over time.

*(LWD initiative: evaluation)*

2. Using machine learning techniques to improve the predictive ability of LWD’s UI exhaustion

model. Currently, LWD uses a regression type model to forecast the likelihood that an individual who has recently become unemployed will remain unemployed and exhaust his or her UI benefits. Heldrich will use state of the art machine learning algorithms, e.g., neural networks, ensemble models, to develop a model that can better predict which individuals receiving UI are likely to exhaust their benefits. By improving on the current predictive model, Heldrich will allow LWD to target services directly to the job seekers who most need the assistance. *(LWD initiative: better integration of UI and reemployment services)*

3.Reports for improved performance metrics. Heldrich will work with LWD to identify measurements for their expanded performance metrics. Heldrich will then calculate these outcomes as a pilot for comparison with the outcomes currently calculated. Discrepancies in the numbers across the two sets of calculation will be addressed by examining not only the mechanics of the calculations, but also the logic behind how the measures are being calculated. The purpose of examining the logic of the calculations will be to ensure that the calculations that Heldrich is producing accurately reflect how well New Jersey's workforce system is serving job seekers. *(LWD initiative: performance metrics)*

4.Enhanced reports for the public. Currently, through the NJCRC, the Heldrich Center uses the WLDS to calculate the labor market outcomes of all training programs in the state. Heldrich then posts these results for the public to view on the njtopps.org web site. Heldrich's editor/graphic designer will work with LWD to test several formats for developing more user-friendly reports. LWD and Heldrich will lead focus groups with job seekers to gather their input as to the types of information and displays they find useful and appealing.

5.Labor market outcomes of hard to serve populations. After it has merged the education and workforce data, Heldrich will work with LWD and NJDOE to conduct two studies of hard to

serve populations. These populations may include the long-term unemployed, youth and / or high school dropouts or individuals with disabilities. For example Heldrich will work with LWD and NJDOE to develop a system for classifying an individual as having “dropped out” of high school, as opposed to, for example, having moved to a different state. Then Heldrich will match those students classified as drop outs with their records in the UI wage record data in order to examine their employment experiences. In addition, Heldrich may follow students in the NJDOE data who have individualized education programs. SLDS data will be matched with DVRS data and UI wage record data to see how many of these individuals transition from high school into the vocational rehabilitation system and how many transition into employment. *(LWD initiatives: evaluation; stackable credentials, understanding the flow of workers in the labor market.)*

6. Research aligned with the NJ Talent Networks. Fitting in with the talent network/sector orientation of the Department, LWD will seek to look at the turnover rates in different industries for low-wage, youth workers. The research will match high school graduates from NJ SMART with the UI wage data, as well as look at the outcomes of high school graduates who are enrolled in college by matching with the NJHE SURE data. In order to get a sense of the industries that might provide better opportunities for individuals with only a high school diploma, The research will then look to observe which industries high school graduates have longer tenure on the job / longer tenure in the industry. The research will also look to measure earnings gains over time for high school graduates by industry. *(LWD initiative: sector talent development )*

All reports will be available on the LWD website, distributed through the department’s electronic mailing lists, and shared with specific stakeholders through the Data Providers and Steering

Committees. These groups will also be encouraged to distribute to their wider audiences.

## **STAFFING CAPACITY**

The current WLDS, located at the Heldrich Center, is staffed by one full-time Database Programmer, one Data Analyst who devotes ten percent of his time to WLDS data analysis, and one Hardware and Data Security Specialist who set up and manages the hardware and sets and implements the security requirements of the WLDS. The Database Programmer has four years of experience processing, cleaning, merging, and managing data in the WLDS. The data analyst has nine years of experience analyzing longitudinal workforce data and conducting quasi-experimental evaluations of workforce and education programs. The Hardware and Data Security Specialist is the Bloustein School's Director of Information Technology, has 20 years of experience in network and system administration, and holds both the GIAC Security Essentials Certification from GSEC and GIAC Certified Incident Handler Certification from GCIH.

LWD and Rutgers propose to expand the staffing to include:

- *Project Manager (PM)*. A half-time project manager will be employed by LWD to coordinate activities within LWD and serve as a liaison between the state partners and the Heldrich Center. The PM will also serve as staff support to the Data Provider and Steering Committees. The PM will be responsible for the project timeline and budget. To ensure proper management, the incumbent must possess a college degree in education, public administration, or a related field and have at least five years administrative and project management experience in a workforce development or equivalent setting.
- *Database Manager*. Heldrich will hire a full-time Database Manager. The Database Manager will have ultimate responsibility for maintaining all aspects of the WLDS, including the design of the system; the processing and cleaning of all data files; the linking of data records; the

extraction and loading of files into data analysis software (State, R) for Data Analysts; and the documenting of all processes. Because the WLDS will consist of a series of flat files, the Database Manager will be required to have knowledge of and experience using data management and analysis software, including SAS, Stata, and R. The Database Manager will be a member of the WLDS Steering Committee and will provide significant input into building partnerships and facilitating data transfer from LWD, NJDOE, and NJHE to Heldrich (and back to those agencies, as needed). The Database Manager will be the primary point of contact for the state agencies. The Database Manager will be required to have a bachelor's degree in computer science or management information systems and 10 years of relevant experience.

- *Programmer.* Heldrich's current database staff member will report to the Database Manager and will be responsible for the day-to-day cleaning, processing, and linking of all data files, as well as the creation of data files that the Data Analysts can use to conduct their studies. The Programmer will also contribute to the partnerships by communicating with LWD staff with respect to data quality.
- *Data Analysts.* Three current, full-time Heldrich staff members, all with PhDs in the social sciences, will serve as Data Analysts on the project. The Data Analysts will be responsible for working with LWD, NJDOE, NJHE and the Database Manager to ensure that the data fields needed for sophisticated data analyses are included in the flat files transferred to Heldrich. The Data Analysts will also be responsible for working with LWD to develop templates for reports based on data from the WLDS, providing input to the WLDS Steering Committee about data analyses to be performed, and conducting quasi-experimental evaluations of workforce programs and enhancing LWD's UI profiling model using machine learning techniques. The lead Data Analyst will participate on the WLDS Steering Committee.

- *Report Designer.* Heldrich will devote a small portion of the resources from the grant for developing user friendly report formats. A Heldrich Center staff member with significant graphic design experience will work to develop reports that are user-friendly and appealing to job seekers and the public workforce system.
- *Hardware and Data Security Specialist.* The Bloustein School's Director of Information Technology will continue to serve as the Hardware and Data Security Specialist.

All of these staff will be jointly responsible, at the direction of the Hardware and Data Security Specialist, for maintaining the security of the data in the WLDS. Current and future Heldrich staff are subject to federal and state laws, including Title 20 CFR 603.9 (security safeguards for disclosed UI program information), FERPA, NJ Administrative Code Chapter 12: 15 – 2.2 (authorized disclosure of unemployment compensation information), and the NJ Privacy Act, governing the handling and authorized disclosure of sensitive and confidential information. The MOUs that Heldrich and LWD have developed require Heldrich staff to comply with these laws and regulations. The MOUs require all staff accessing confidential data sign agreements indicating that they understand that they are required to protect confidential data. All Heldrich are required to participate in regular computer security training.

Finally, The Database Manager, the Programmer, the three Data Analysts, the Report Designer, and the Hardware and Data Security Specialist will all be employed by, or under contract with, Rutgers University, The State University of New Jersey.

### **BONUS POINTS—OTHER DATA LINKAGES**

In November 2011, LWD became a full member of WRIS2. The current WLDS includes Title II Adult Education data. The current WLDS also includes data on career and technical education program adult completers so Heldrich can calculate employment outcomes for adult

vocational education programs to help NJDOE meet Perkins Act reporting requirements.

Through the WDQI, LWD will add unit record data from NJ DVRS to the WLDS. Finally, LWD will also work with NJDHS and Heldrich to create a Memorandum of Understanding to allow SNAP data to include in the WLDS.

<b>TIMELINE OF ACTIVITIES</b>	Qtr1	Q2	Q3	Q4	Q5-6	Q7-8	Q9-10	Q11-12
<b>Activity</b>								
Heldrich data inventory of current WLDS	■							
Steering Committee meetings	■	■	■	■	■	■	■	■
Data Providers Committee meetings	■	■	■	■	■	■	■	■
LWD, NJDOE, and NJHE provide Heldrich data dictionaries for relevant data systems	■							
LWD negotiate MOU for sharing WIA Title I, Wagner-Peyser, TAA, DVRS, WDPP, and UI benefits data with Heldrich	■	■						
NJHE negotiate MOU with Heldrich for prior year higher education data files	■	■						
NJDOE negotiate MOU with Heldrich for SLDS	■	■						
LWD work with NJOIT to securely transfer workforce data files to Heldrich		■	■	■	■	■	■	■
Heldrich cleaning, processing, and linking of workforce data		■	■	■	■	■	■	■
Heldrich conduct Study 1 WIA evaluation			■	■	■	■		
NJHE transfer prior NJHE data files				■	■	■		
NJDOE transfer SLDS data to Heldrich				■	■	■	■	■
Heldrich conduct study 2 UI profiling					■	■		
Heldrich cleaning, processing, and linking of education data with workforce data					■	■	■	■
Heldrich conduct Study 3 WIA reports					■	■		
Heldrich conduct studies 4 – 6							■	■