

**MASSACHUSETTS DEPARTMENT OF WORKFORCE
DEVELOPMENT**

WORKFORCE DATA QUALITY INITIATIVE

DATA WAREHOUSE/LONGITUDINAL DATA SYSTEM

GRANT NUMBER: MI-21172-10-60-A-25

PROJECT DESIGN

FEBRUARY 1, 2011

DATA WAREHOUSE PROJECT OVERVIEW

The Workforce Data Quality Initiative (WDQI) three-year grant will provide funds for the Department of Workforce Development to accomplish three major goals.

First, expand the planned design and development of a longitudinal data system for unemployment insurance and labor market information to include workforce information from the Division of Career Services and Division of Apprentice Training, and Commonwealth Corporation. The resulting longitudinal data system will comprise a single repository for data records spanning the full range of workforce systems in Massachusetts, including unemployment insurance, apprenticeship training and the One-Stop Career Centers. The data warehouse will link data records and allow us to track individuals and cohorts of participants across unemployment insurance records, employer reported wage records, employment and training services (Workforce Investment Act and Wagner Peyser), Trade and National Emergency Grant services, apprentice data, and federal and state youth programs.

Second, the longitudinal data system will integrate programmatic and performance data with sources of labor market information to analyze how program results tie back to labor trends in the state.

Third, utilizing the cross-agency membership of the Performance Committee of our statewide Workforce Investment Board, the Committee will drive a portion of the grant activities to link the design of the workforce-based data warehouse with the on-going development of the Massachusetts Department of Elementary and Secondary Education's education-based longitudinal data warehouse (P-20) supported by two grants from the U.S. Department of Education for statewide longitudinal data systems (SDLS).

The longitudinal data system will allow the Commonwealth to use cross-program data to evaluate education and workforce programs and answer key policy and funding questions for our State legislators, policymakers, the members of the Massachusetts Workforce Investment Board, and others seeking to provide cost effective services for our citizens. Data and reports will be available to consumers, including educators, program administrators and operators, stakeholders and the public through an easy to use web application.

Project Plan

The DWD longitudinal data system (data warehouse) is a major new project for DWD in the planning and early implementation stages. *Refer to Attachment A for a high-level timeline and milestones.* Oracle's Business Intelligence Suite Enterprise Edition (OBI EE) has been procured as the

application foundation. The platform architecture has been defined. DWD engaged a contractor (Currier McCabe & Associates d/b/a CMA Consulting Services) in January 2011 with expertise in OBIEE products and data warehouse development. The contractor will be responsible for installation of OBIEE and for the development of a detailed multi-year plan for building out the longitudinal data system.

The scope of the DWD longitudinal data system will include the unemployment and workforce programs operating under the administration of the Executive Office of Labor and Workforce Development (EOLWD). DWD within EOLWD comprises three divisions: the Division of Unemployment Assistance (DUA), which is responsible for UI Revenue and Benefits programs and Economic Research; the Division of Career Services, which is responsible for the major federal workforce programs (Wagner-Peyser, WIA and Trade) and services that are delivered primarily through the State's One-Stop Career Centers; and the Division of Apprentice Training (DAT), which is responsible for promoting, developing and servicing registered apprenticeship programs in Massachusetts. Additionally, Commonwealth Corporation (CommCorp), a quasi-public authority within the Executive Office of Labor and Workforce Development, administers State-funded youth jobs programs and sector training programs, and selected workforce demonstration projects.

Grant resources through the WDQI will leverage these existing investments and support additional design and product development to expand and accelerate the scope of the warehouse beyond unemployment insurance data to include workforce participant and longitudinal data analysis. Grant resources will assist the State with the following objectives and outcomes:

1. Complete the reporting requirements for Unemployment Insurance Revenue and Benefits with full implementation of the QUEST revenue and benefits applications that will largely replace the UI legacy systems by the end of Fiscal Year 2011.
2. Create a longitudinal data system comprising a single repository for data records spanning the full range of workforce systems in Massachusetts, including unemployment insurance, apprenticeship training and the One-Stop Career Centers.
3. Link data records and allow tracking of individuals and cohorts of participants across unemployment insurance records, employer reported wage records, employment and training services (Workforce Investment Act and Wagner Peyser), Trade and National Emergency Grant services, apprentice data, and federal and state youth programs.
4. Integrate programmatic and performance data with sources of labor market information to analyze how program results tie back to labor trends in the State.
5. Link the design of the workforce-based data warehouse with the on-going development of the Massachusetts Department of Elementary and Secondary

Education's education-based longitudinal data warehouse (P-20) supported by two grants from the U.S. Department of Education for statewide longitudinal data systems (SDLS).

6. Facilitate examination of policy and programmatic questions that cannot be addressed comprehensively at this time such as, employment outcomes for students enrolled in post-secondary programs and adult education (Title II of WIA), the cost effectiveness of different types of training, and the optimum mix of services for claimants to reduce length of unemployment and realize savings in unemployment benefits.
7. Make data and reports available to consumers, including educators, program administrators and operators, stakeholders and the public through an easy to use web application.

The principal application/system sources for the data that will be brought into the DWD longitudinal data system are supported by the EOLWD/DWD Information Technology Department. These are QUEST (Quality Unemployment System Transformation) a multi-year undertaking to replace the UI legacy system for UI Revenue and UI Benefits with up-to-date web-applications built on an Oracle platform, and MOSES (Massachusetts One-Stop Employment System) which serves the workforce system and Career Centers and tracks all participant and service activities under Wagner-Peyser, WIA, Trade, UI Reemployment Services and Reemployment Assessment programs, workforce and training services for TANF recipients, and individual Career Center programs.

The Economic Research department within DWD/DUA manages and maintains data developed through the BLS cooperative programs as well as UI reporting and UI data quality assurance and validation. The Division of Apprentice Training and Commonwealth Corporation maintain smaller SQL server based applications. These agencies are all under the Executive Office of Labor and Workforce Development (EOLWD) and in the case of DUA, DCS and Commonwealth Corporation have established data sharing processes.

DWD Longitudinal Data System Data Components

The longitudinal data system is being designed to contain the following categories of data at the individual record level.

<u>Data Category</u>	<u>Source system</u>	<u>Owner</u>
UI Employment and Wage Record Data	QUEST	DWD DUA
UI Claimant and UI Benefits Data	QUEST	DWD DUA
Wagner-Peyser	MOSES	DWD DCS
WIA Title I Adult	MOSES	DWD DCS

WIA Title I Dislocated Worker	MOSES	DWD DCS
WIA Title I Youth	MOSES	DWD DCS
TAA – Trade Adjustment Assistance	MOSES	DWD DCS
TRA – Trade Readjustment Allowance	QUEST	DWD DUA
Registered Apprenticeship	SQL database	DWD DAT
YouthWorks (State-funded youth jobs program)	SQL database	CommCorp
Sector Training Programs/Demonstration Projects	SQL database	CommCorp
Quarterly Census of Employment and Wages	Research	DWD DUA
Local Area Unemployment Statistics	Research	DWD DUA
Mass Layoff Statistics	Research	DWD DUA
Business Employment Dynamics	Research	DWD DUA
LED Quarterly Indicators	Research	DWD DUA
Fed. Employment Data Exchange System	FEDES	

Data will be comprehensively linked and integrated into a re-architected data warehouse for up to 1,000 users scalable for public facing and external stakeholders with increased usage, and will feature role specific data marts that provide rapid, relevant, timely and legally compliant data for users. WIASRD level data for workforce programs will be included in the data system. Oracle Business Intelligence software will provide web-based access to dashboards, reports, and ad hoc query functions to be designed and developed with input from workforce partners and other potential users.

The single repository will improve alignment and blended reporting with secure, unified, single definitions and conformed dimensions; will be integrated to provide an enterprise view from high level reporting to detailed transactions; and will reduce cost by consolidating and standardizing environments, processes and data.

Research and Evaluation Goals

The research and evaluation goals for the project are based on the direction set by the Massachusetts Workforce Investment Board's Performance Committee in its October 2008 draft *Report on Governance and Coordination of the Workforce Development System Recommendations* that defined a set of broad objectives for the development of an Integrated Performance Management System to inform policymakers and practitioners regarding the performance of workforce and education programs. The activities of the grant will allow us to accomplish the goals below.

1. Explore the use of wage records to determine employment, employment retention, and earnings outcomes for participants in all publicly funded workforce development programs and educational programs at the secondary and post-secondary level.
2. Explore the capacity to match and share records of “high need” individuals served by multiple programs and funding streams to facilitate improved coordination of service delivery and impact analysis.
3. Explore the capacity to periodically undertake a cost benefit analysis based on net impact analysis, which in turn, would be based on the use of employment (wage) records to determine increase in participant earnings.
4. Support local and state managers in:
 - a. Understanding the impact of publicly funded programs
 - b. Making resource allocation decisions.
 - c. Aligning the investment of workforce development resources
5. Support local managers and staff in facilitating planning and service coordination for customer groups that are served by multiple agencies and funding streams.
6. Provide state and local managers with timely information about progress toward achieving program goals, such as employment, employment retention, placement wages, or earnings increases.
7. Support trend analysis to identify the need for redirection of resources or new investments.
8. Facilitate analysis to support program improvement, including analysis of program impact and effectiveness for targeted customers. Understanding the experience and outcomes for customers who are served by multiple programs.
9. Pursue legislation that enables the use of educational records from secondary and post-secondary systems for the purpose of research and program improvement. The legislation can also explicitly require workforce development programs to share data and track key outcomes. One of the key purposes of this legislation would be to address concerns related to the Family Education and Privacy Rights Act (FERPA).

In addition, the research and evaluation goals for the grant and the Performance Committee align with the goals of the SLDS grant awarded to the Department of Elementary and Secondary Education (DESE) in May 2010. A stated outcome in DESE’s SLDS grant is the establishment of a multi-staged process for linking education, employment and training, health and human services data. This will be accomplished in two steps: (1) centralize current aggregate data, services and resources available, and identify conditions that create demand for services from the Executive Office of Education, Executive Office of Health and Human Services, and the Executive Office of Labor and Workforce Development; and (2) convene a high-level workgroup under the auspices such as the Governor’s Readiness Cabinet to develop options to reconcile privacy and data security requirements with the goal of data sharing

across the range of services needed by at-risk children and families.

The benefits of comprehensively linking and integrating longitudinal data will be the ability to examine the post-program success of federal and state funded workforce programs using wage records, analyze program outcomes with respect to industry/occupation vacancies and other labor market variables, determine the employment outcomes for students enrolled in postsecondary programs and adult education (Title II of WIA) through use of wage record data, examine patterns of unemployment in relation to individual characteristics and economic conditions, and evaluate the cost effectiveness of different types of training.

Technical Architecture

The application needs for the project will be met by Oracle's Business Intelligence product, Enterprise Edition on Windows Operating System (already procured), with multiple application servers that are load balanced and scalable based on the expansion needs. The database needs will be met by Oracle 11g or 10g Enterprise Edition on Windows Operating System using Real Application Clusters (RAC) with multiple nodes having a Data Guard database as a safety net to the data. The QUEST databases are on Oracle 10g environment and there is a proposal to upgrade the QUEST environment to Oracle 11g before the Benefits phase goes live before the end of Fiscal 2011, to make best use of the advanced methods offered by Oracle 11g in establishing a robust and more scalable database platform.

The MOSES databases, the flag ship database for the DWD – DCS operations, consisting of all the Online and Data Mart data for all the needs of DWD – DCS career center users, since 1999, when it went live, has been on TRU 64, Unix servers. The HP Unix servers were built during the 90s using Alpha chips and since the inception of the Moses Project the database has expanded manifold and the servers are functioning at maximum capacity. These servers must be replaced to allow adequate processing speeds for transfer to the data warehouse. For preparing the data for feeding into the Data Warehouse environment, MOSES will require new batch programs to either re-write the data in a de-normalized form or to directly transfer the data from the database to the Warehouse application. For the DWD warehouse, we intend to have all the databases on the same Oracle 11g on Windows Operating Systems.

Data Quality

DWD's validation processes for individual data systems that will populate the data warehouse will be reviewed and business/validation rules that verify the accuracy of data elements will be documented. Workforce program data is held within the MOSES system and the WIASRD level data will be brought into the data warehouse. As we implement the longitudinal data system, it will be

increasingly necessary to evaluate the data consistency and data quality in each of the component source systems.

As the information is relied on for high stake decisions, users must be assured of its accuracy and quality. We will design and implement a data quality and certification process that includes data audits to periodically assess the accuracy of the data. Data audits will identify differences between data sources – UI, MOSES, DAT and CommCorp – and agreements between partners will require reconciliation of variations in completeness for data inputs to the data warehouse. Partner agencies will be expected to make necessary changes in policies and/or procedures to ensure the consistency and accuracy of data entered at the source.

Planned Reports and Deliverables

One first series of specific deliverables and reports will be for UI revenue and benefits. On the workforce side, we will identify and convene workgroups drawn from our workforce partners and the various stakeholder groups to identify specific reports and analyses and to set priorities. The Workforce Boards have been eager to obtain data on employment outcomes of participants from training and on individuals served at the Career Centers based on wage records. While we have been able to provide some wage data files on a very restricted basis, their use required programming expertise on the part of WIB staff. Thus, most areas have had to rely on the federal performance data.

The evaluation and research questions cited above will inform the review of what data are currently in existing source data systems that will be transferred to the longitudinal data warehouse, and the data sets that will need to be built over time. The primary goal is to assist stakeholders, program managers and customers to better understand the impact of workforce and education programs, including the impact of specific services strategies. Over time, the reports and analysis should support making resource allocation decisions and ensuring the effective alignment of the investment of workforce development resources.

The general strategy for reports is to increase and improve the analysis of existing programmatic data in terms of customers groups, activities and services, vendors and providers, and interim outcomes such as completions and certificates. The longitudinal database will facilitate and enhance such reporting in two important ways: (1) the programmatic data will not be bound by the standard (and arbitrary) reporting time periods, cohorts and funding silos needed for federal reporting, and (2) the wealth of available program data will be able to be analyzed along with post-program outcomes on employment, retention and earnings using wage record matching. Initially, access to the data system will be limited for the general users to standard format reports. Longer term, the intent is to develop more ad hoc options for users to mine the data base for the production of customized reports.

As more workforce and education programs are added over time to the longitudinal data base, an increasingly important focus will be to explore the capacity to match and share records of individuals (especially “high need”) served by multiple programs and funding streams. This will support the analysis of the success in the labor market by population groups in various sequences of services and thus to facilitate improved planning and coordination of service delivery.

Ultimately, it is hoped that the longitudinal data base will support trend analyses and, potentially with the assistance of the research community, more rigorous research on workforce programs such as net impact studies and cost benefit analyses.

Personal Identifier

All data sources within the DWD longitudinal data system will use Social Security Number (SSN) as the personal identifier for individuals and the Federal Employer Identification Number (FEIN) or Division of Unemployment Assistance (DUA ID) number for employers. Initially linkages with educational agencies, principally the Department of Elementary and Secondary Education (DESE) will likely occur on an ad hoc basis in order to deal with the difficulties of matching individual DESE records that do not use SSN as the key personal identifier. DESE has invested in software (*CPSI xDUID*) to create and manage unique student identifiers in its longitudinal data system and this product will be investigated for use in matching between DESE and the DWD data system. Agencies within the Executive Office of Health and Human Services that provide services to custodial youth and TANF recipients do make use of the SSN as do institutions of higher education.

Generally, SSN is available in Adult and Community Learning Services (CLS) Adult Basic Education ABE (Title II) records as well as in public higher education records, however, there are numerous confidentiality issues with the linkage or exchange of education records with personal identifiers. These are important issues to be addressed over the course of the grant with the assistance of the MWIB Performance Committee or other advisory groups, such as the Governor’s Readiness Cabinet, and legal counsel. In addition, P-16 records generally do not have SSNs and there is even more work to be done in exploring strategies for linking elementary and secondary records to workforce systems that traditionally make use of SSNs. *Establishing a forum for evaluating privacy and data security issues in sharing data across the range of education, health, human service and employment and training services will be a priority of joint collaboration among these agencies.*

Inter-Agency Agreements

Memoranda of Understanding (MOU) will be developed first with the Massachusetts Department of Elementary and Secondary Education (DESE), and then with the Department of Higher Education (DHE) and Executive Office of Health and Human Services (EOHHS) with respect to specific data sharing and data exchanges. Development of MOUs will begin in fourth quarter FY2011 under the auspices of the WDQI Steering Group and the Executive Office of Labor and Workforce Development.

MA Workforce Board Performance Committee / WDQI Steering Group

<u>Name</u>	<u>Organization</u>
Tamika Correira	Exec. Director, MA Workforce Investment Board (MWIB)
Jennifer James	EOLWD, Deputy Dir. for Workforce Development
Marilyn Boyle	WDGI Project Manager, Div. of Career Services
Rena Kottcamp	Director, Economic Research & Analysis, DWD
John Glennon	EOLWD, Chief Information Officer
Navjeet Singh	Dir., Research and Evaluation, Commonwealth Corporation
Gene White	Deputy Director., Research and Evaluation, Commonwealth Corporation
Jonathan Keller	Associate Commissioner for Research, Planning, and Information Systems, MA Department of Higher Education
Rob Curtin	Department of Elementary and Secondary Education
Vladimir Hyppolitie	SDLS Project Lead, DESE
Glen Daly	Executive Office of Health and Human Services
Donna Cornellier	Adult and Community Learning Services, DESE
Anne Serino	Dir. of Adult and Community Learning Services, DESE
Mary Sarris	North Shore Workforce Investment Board
Ruthie Liberman	Workforce Solutions Group; Crittenton's Women's Union
Elizabeth Kelleher	Mass. Workforce Professional's Association

**ATTACHMENT A - DESIGN AND DEVELOPMENT OF THE LONGITUDINAL DATA SYSTEM
DEPARTMENT OF WORKFORCE DEVELOPMENT**

CORE DEVELOPMENT STAGES	JULY 2010 - JUNE 2011				JULY 2011 - JUNE 2012				JULY 2012 - JUNE 2013				JULY 2013 - JUNE 2014			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
I. Infrastructure and Architecture																
Acquire Equipment/Software/Staff/Contractor	D	D	D													
Development/Test/Production Environments			D	D												
BI Platform - Security/Tools/User Interface			D	D												
Expansion and Enhancements																
II. Data Definition, Data Acquisition, and Data Load																
UI Revenue (QUEST)				D	D	D	D	D								
Employment and Wages (QUEST)				D	D	D	D	D								
UI Claims and Demographics (QUEST)				D	D	D	D	D								
Labor Market Information (for LED)				D	D	D	D	D								
Workforce Programs (MOSES*)				D	D	D	D	D								
Apprentice (DAT)						D	D	D	D							
YouthWorks & Sector Trng Pgrms (CommCorp**)						D	D	D	D							
Fed. Emp. Data Exhchange System (FEDES)						D	D	D	D							
III. Business Intelligence and Business Analytics																
Design/Develop Dashboards - Reports - Queries					D	D	D	D	D	D	D	D	D	D	D	D
Workforce Program Analyses											D	D	D	D	D	D
Web Application for Consumers											D	D	D	D	D	D
IV. Linkages with Educational Data Systems																
Requirements/Data Sharing Agreements/MOUs				D	D	D	D	D	D	D	D					

D - Identifies intense initial development period/requirements/design/programming/implementation; enhancements/refinements on-going after initial implementation
I & II - Data system design, security and access levels, source to target mappings, extract/transform/load, production
II & III - Business/reporting requirements, information portal design, interactive dashboards, ad hoc query, reports
IV - Workforce and educational linkages, analysis, reports

*MOSES covers Wagner-Peyser, WIA, TAA, Career Center Programs
Commonwealth Corporation

DWD Longitudinal Data System - WDQI Grant 02.01.11
Grant Number MI-21172-10-60-A-25