



MLDS CENTER

Maryland Longitudinal Data System

Address 550 West Baltimore Street
Baltimore, MD 21201
Phone 410-706-2085
Email mlds.center@maryland.gov
Website www.MLDSCenter.org

**MARYLAND LONGITUDINAL DATA SYSTEM (MLDS)
550 West Baltimore Street
Baltimore, MD 21201**

GOVERNING BOARD MEETING

**March 11, 2016
MINUTES**

The meeting of the Maryland Longitudinal Data System (MLDS) Governing Board was held on March 11, 2016, in the Maryland State Department of Education (MSDE) Board Room of the Nancy S. Grasmick Building. Ms. Kelly Schulz, called the meeting to order at 10:00 a.m. and noted that a quorum was present.

The following Governing Board members were in attendance:

Ms. Kelly Schulz, Secretary, Maryland Department of Labor Licensing and Regulation and Vice Chair
Dr. James Fielder, Secretary of Higher Education
Dr. Ben Passmore, Assistant Vice-Chancellor for Policy Research and Analysis, University System of Maryland (Designee for Chancellor Caret)
Dr. Henry Johnson, Chief Academic Officer, Maryland State Department of Education (Designee for Dr. Jack Smith, Interim State Superintendent of Schools)
Ms. Tina Bjarekull President, Maryland Independent Colleges and Universities Association
Dr. Bernard Sadusky, Executive Director of the Maryland Association of Community Colleges
Ms. Jennifer Mullinix, Math Teacher, Wilde Lake Middle School, Columbia, Maryland
Mr. Steven Rizzi, Vice President, PAR Government
Mr. Brian Roberts, Change Management Specialist, Montgomery County Government

The following staff members were in attendance:

Mr. Ross Goldstein, Executive Director, MLDS Center
Ms. Tejal Cherry, Director of System Management Branch, MLDS Center
Ms. Laia Tideman, Data Management Coordinator, MLDS Center
Mr. Peter Hobbs, Director of Reporting Services, MLDS Center
Dr. Angela Henneberger, Director of Research, MLDS Center
Dr. Michael Woolley, Principal Investigator, MLDS Center
Dr. Laura Stapleton, Associate Director of Research, MLDS Center
Dr. Terry Shaw, Associate Director of Research, MLDS Center
Ms. Dawn O’Croinin, Assistant Attorney General for the Governing Board and MLDS Center
Mr. Robert Murphy, Data Analyst, MLDS Center
Ms. James Dixon-Bobbitt, Executive Associate, MLDS Center
Mr. Chuck Shelton, Senior System Architect, MLDS Center

Ms. Schulz began the meeting with introductions and a request that a report on the recent Legislative Audit be provided. Next, Ms. Schulz asked for a motion to approve the December Governing Board Meeting Minutes. Mr. Rizzi motioned to approve the minutes, which was seconded by Dr. Fielder. The motion was unanimously approved.

Center Report

Mr. Goldstein began by providing an overview of the Legislative Audit. The findings in the audit report are the same findings that were discussed previously after each of the draft (fiscal and IT security) reports

were released. There were a total of three findings (reduced from five). The first finding was that the Center procured services through an existing interagency agreement with the University of Maryland, Baltimore, rather than through a competitive procurement process. The Center agreed with the finding and promised to discontinue the practice. Mr. Goldstein noted that there is one final contract with UMB that ends in June. The second and third findings were information technology security issues. The audit found that: (a) PII data was not properly secured; and (b) the Center's servers were not adequately secured - specifically operating system updates were not applied, anti-malware software had not been installed and software updates not applied. Again, the Center agreed with these findings and noted that the recommended actions have been fully implemented. These issues were due in large part to the hosting situation at the Department of Public Safety and Correctional Services (DPSCS). The hosting arrangement resulted in MLDS staff having limited visibility and access to system controls. That has been rectified by moving the data center to MSDE (which was why the Center was in the process of relocating the data center during the audit).

In response to a question from Mr. Rizzi, Mr. Goldstein stated he was not aware if any audit findings would be directed at DPSCS as the data host, but noted that the ultimate responsibility for security lies with the Center. In response to a question from Dr. Passmore, Mr. Goldstein agreed that negative audit findings will erode confidence in the system. The Center plans to have an independent security audit to review the implementation of the *Data Security and Safeguarding Plan* as well as identify issues that staff may not have thought of or anticipated. Ms. Cherry noted that the Center now has sufficient staff and system access to fully manage the security needs. Finally, Ms. Schulz asked whether staff would know if a system breach occurred and would that be reported to the Board. Mr. Goldstein responded that the incident management plan addresses those issues, including required audit logs, review of audit logs and breach notification to partner agencies. Mr. Rizzi also responded that he has been working with staff to review security documentation. Finally, Ms. O'Croinin clarified that there has not been a security breach.

Mr. Goldstein next informed the Board that the Center has started to receive data requests. Morgan State University requested information for Fall 2014 enrolled students, to include the number employed (based on wages reported in 3rd and 4th quarter of 2014), the industry in which they were employed, and the academic program in which they were enrolled. Delegate Gutierrez requested information on high school students with initial college enrollments broken down by race and by county. The Center has initial enrollments by race on the website - but not broken out by county.

System Development

Ms. Cherry referred to the MLDS Center Data Milestones document and discussed the progress and upcoming activities in the following areas:

- Security and user access - staff has established user accounts, single sign on, closed all guest accounts and unused ports (to protect against unauthorized users coming in through firewall), and reviewed firewall policies. Staff is also working with DoIT to conduct an internal system audit and the process of moving firewall to the DoIT services.
- Development environment - dashboards are being created and new tools are being implemented to manage the data dictionary and inventory.
- Test and development - staff has completed installing servers, creating a fully virtual environment, maintaining firewalls, and upgrading Oracle Software. Staff has also created virtual machines with necessary analytical tools for research staff. All servers have been shut down at DPSCS and plans are underway to move the remaining servers to the MSDE hosting facility. Ms. Cherry did explain that the offsite backup has been delayed, but all data is currently backed up. In response to a question from Mr. Rizzi, Ms. Cherry explained that the offsite backup will not be a "hot site." The offsite backup will serve as a complete data and system backup taken in monthly intervals. Ms. Cherry also responded that the Center will be using Amazon Web Services (AWS) - specifically, the most secure government version. Finally, Ms. O'Croinin

responded to Mr. Rizzi's question about potential legal implications of the Center not having custody of the data by noting that per the contract with AWS, the Center constantly maintains ownership of its data.

- Data load - staff has made ongoing progress loading data, including MHEC enrollment and degree information data for 2008-2015, MSDE Student, Course, Grade, Teacher data, wage data for the third quarter of 2015, NEDP data, and data provided by the Motor Vehicle Administration (MVA). MVA data comes through a secure transmission on a daily basis.

Ms. Schulz noted that the Center's agreement with DLLR requires the Center to return any DLLR workforce data that cannot be matched to a student record and asked whether the inclusion of the MVA data would result in nearly all DLLR data being matched to an MVA record and therefore not returned to DLLR. Ms. O'Croinin clarified that the MVA data was being used to fill in missing information from student records to improve matching. However, the MVA data is not its own sector - the workforce data still needs to be matched to a student record.

Mr. Rizzi asked if there is a research value to understanding the entire picture of Maryland workers. Dr. Woolley responded that it would provide research value because maintaining all workforce data would allow researchers to understand the whole state working population. It would also allow researchers to understand work history that may have occurred before an individual becomes a student. Dr. Passmore noted that the legislature wanted to ensure that the system only included data necessary to improve education and therefore created restrictions to limit the workforce data to records that directly relate to students. Ms. O'Croinin noted that state law restricts the linkage of student and workforce data to 5 years from the date of the latest attendance in any educational institution in the State. After that five years the data must be unlinked, but can remain in the system (up to 20 years). Ms. Schulz stated that it was important for DLLR to understand what the addition of the MVA data does to the percentage of data being used and maintained by MLDS versus data being returned to DLLR.

Data Management

Lia Tiderman noted that MVA started sending its data in January. The data is sent daily. There was an initial file that included all relevant records and the daily filings are just changes. The impact of this data can be understood by revisiting the data matching tables that staff presented in December, 2015. The data match rate is determined by the rate of 12th graders that can be matched to a postsecondary and/or workforce record. The December match rate was 88% and now, with the inclusion of the MVA data it is 93%, which represents a significant improvement. The MVA specifically addresses the issue of being able to match a K-12 record to a workforce record. K-12 data has limited Social Security Numbers (SSN) so matching to workforce data is limited. Carroll County provides a good example of the improvements. In the December data matching table, only 77% of Carroll County 12th graders (610 students) were matched to a postsecondary and/or workforce record. This was a result of the fact that Carroll County does not collect SSNs from students. With the addition of the MVA data, there is now a 94% match rate (over 3,000 students). Finally, Ms. Tiderman noted that the match rate information is included in a memorandum submitted with the meeting materials.

In response to a question from Mr. Rizzi, there was a discussion about potential "blind spots" in the system that occur when Maryland high school students go out-of-state for college. The concern is that the student spends five or more years out-of-state getting his or her degree and then returns to the state to live and work. Due to the five year limitation, the Center would not be able to match that student with a workforce record. This is the case even though the Center receives out-of-state college enrollment data from the National Student Clearinghouse.

Reporting

Peter Hobbs provided an overview of the Center's operating procedures for the development of reports and dashboards. The workflow diagram formalizes a process that shows how the Center interacts with

other agencies, adheres to the data reporting standards, and the activities leading up to the finalization of a product. The process starts with selecting a Research Agenda Question (RAQ), lays out the roles of the team, and establishes a timeline. Mr. Hobbs noted that the three week timeline does not mean that the team is only working on one dashboard at a time. Different dashboards can be at various stages in the workflow process at any time.

Mr. Hobbs informed the Board that a dashboard in response to RAQ 12 is currently posted on the website, and dashboards for RAQ 12, 18, 20, and 6 are all in various stages in the workflow process.

Mr. Rizzi asked whether the development process includes a review by a test panel or consumer group to ensure the content is understandable and appropriate for the general public. Ms. O’Croinin noted that, while there were reviews by outside groups, such as the Research and Policy Advisory Board, there is no test panel. Mr. Goldstein added that currently the dashboards are very complicated and they need to be more accessible to the general public. The original design of the MLDS website was to focus on different communities of interest: educators, policymakers, researchers, and parents and students. The Center needs to find ways to engage those audiences. Ms. Schulz recommended that the P20 Council’s MLDS workgroup could serve as a reviewer. Mr. Rizzi agreed that the workgroup could play a role, but stated that the group was comprised of individuals steeped in education policy. He is looking for non-experts; for example, his mom.

Mr. Goldstein noted that not all dashboards need to serve all people. Dashboards for policymakers and researchers can remain complex and provide in-depth information. But dashboards for students, parents, and teachers should be developed and reviewed with input from appropriate test groups.

In response to a question from Ms. Schulz about how the dashboards are being prioritized, Mr. Goldstein stated that currently topics are being chosen based on available data or the need to explore data that has not yet been utilized (most recently workforce outcomes). However, there will come a time when new mechanisms to decide dashboard priorities will need to be established. Working with the P20 Council, the legislature and the Governor’s office will go a long way to establishing those priorities.

Research Branch

Dr. Woolley provided an overview of four analyses and connected reports that the Research Branch team is currently working on and plans to complete within the next two to three months. First, there is a report that assesses training and retention of early childhood workforce in Maryland. Second, there is a report that tracks students who attain a postsecondary degree and where they are found in STEM workforce. Third, there is an analysis of remedial education that looks at what characteristics predict the need for remedial education and whether students assessed to need remedial education persist in postsecondary education and graduate. Finally, there is an analysis of the workforce outcomes of high school students who do not go on to postsecondary education.

Dr. Woolley also stated that the Research Team is working on administrative issues including reviewing the data suppression rules and the interagency agreement between the Center and the universities for the Synthetic Data Project (under the federal 2015 SLDS Grant).

Tina Bjarekull raised the issue of how the Center is moving forward with the synthetic data project. State law is specific about how data is to be used and managed and that the policy agenda is to be set by the Governing Board. Creating the synthetic data represents a major change for which there are no rules regarding the type of research or how long the synthetic data will be maintained. The issue was brought before the Governing Board and members raised concerns, but staff moved forward with the project despite never obtaining Board approval. Therefore, Ms. Bjarekull stated that the MOU should come back to the Board for review?

Dr. Passmore requested that before discussing the administrative issues the Board should first understand what synthetic data is. Ms. Schulz responded that this is a bigger issue than just the Synthetic Data Project. There should be more input on development of MOUs and plans for the Center. Ms. O’Croinin responded that her review of the legislative intent for the bill the created the MLDS contains nothing that would preclude a project like the development of synthetic data. She also noted that after consulting with representatives of MHEC and MSDE, it was determined that it is not customary to bring a grant proposal or MOU to the board or commission for approval (although the Board could request that in the future staff takes such actions).

Ms. Bjarekull noted that there was a lot of opposition to the MLDS and therefore assurances of that the use of the system would be limited to public policy analysis only and not used for commercial applications. Further, Ms. Bjarekull asserted that the Synthetic Data Project represented a distraction from the primary work of the Center at a time when the system was in development and needed of the staff’s full focus.

Mr. Goldstein responded that if, during the review of the bylaws, the Governing Board wants to direct staff to seek Board approval before moving forward with grant applications or MOUs he will certainly work within those parameters. Mr. Goldstein noted his commitment to serving the needs of the Board. Mr. Goldstein also stated that the Synthetic Data Project was not and will not be a distraction. None of the IT staff worked on the grant and will not in any way be involved in the project once it begins. The project is fully managed by the research team as a separate task with separate staff resources. It also won’t detract from the time the researchers spend on the Center’s work.

Mr. Goldstein also stated that his thoughts about the grant were that (1) State law specifically directs the Center to find additional resources, including federal grants to support the work of the Center; (2) the project is consistent with the work and goals of the Center and will provide resources to help support our understanding and development of the system and expedite research output; and (3) the grant and the resulting project will enhance the prestige of the Center and make the Center more susceptible to attracting future research dollars.

Ms. O’Croinin noted that the legislative history contains no evidence that there was ever an intention to exempt the Center from compliance with the Public Information Act (PIA). Ms. Schulz stated that the Center is not responsible for creating records in response to PIA requests. Ms. O’Croinin agreed that agencies are not required to create records, but extracting information that is in the possession of the agency is not creating a new record. Ms. Schulz stated that there is not a consistent interpretation among Assistant Attorney Generals on this issue and that consultation at a higher level is necessary.

Dr. Passmore asked whether the synthetic data would impact PIA requests? Ms. O’Croinin stated that it would because that data could be more easily shared without having to carefully analyze the response to the request for compliance with complex data suppression rules that are necessary to ensure personally identifiable information is not disclosed. Dr. Passmore noted that regardless of whether there is synthetic data, the Center will have to respond to requests. This led to a discussion of the Center’s responsibility to release data subject to a PIA request. Ms. Schulz requested that counsel from the MLDS and partner agencies confer on the PIA requirements in general and any data requests that may be received until there is clarity on the extent of the Center’s obligations under PIA. Ms. Bjarekull suggested the Board consider proposing amendments to PIA to help ensure that the MLDS is only used for education and workforce policy matters.

In order to have time for the presentation on the synthetic data project, Ms. Schulz moved to that agenda item. Ms. Schulz requested that staff put the final approval of the regulations and bylaws review on next month’s agenda. Ms. Schulz noted that the members received a copy of the current bylaws that contained changes proposed by the Mr. Goldstein and she had distributed changes recommended by her office.

Ms. Schulz encouraged the members to review the bylaws and proposed changes for a discussion at the next meeting. Mr. Goldstein also noted that he would include the bylaws and proposed changes on the agenda for the Research and Policy Advisory Board to get input from their staff representatives.

Presentation of Synthetic Data Project

Dr. Woolley began by noting that synthetic data is used by other government programs - most notably the U.S. Census Bureau for their Security of Income and Program Participation data. However, no state longitudinal education data system has investigated using a synthetic data strategy. Dr. Woolley stated that the project is just one part of an overall grant award to MSDE for a total of \$7 million. The synthetic data project will receive \$2.6 million of that funding over four years. The funding ensures that the project will not detract from other work of the Center, since it will fully cover the costs of the personnel and researchers' time on the project.

The first phase of the project will be to build data sets:

1. K-12 - to Postsecondary;
2. K-12 to Workforce; and
3. Postsecondary to Workforce.

These datasets will have practical day-to-day uses for the MLDS Center work. They can be efficiently used for research analysis and building dashboards. Currently, to work with the data, staff have to create a data set that pulls the needed variables from various tables. Having data sets ready to use will create efficiencies and improve production.

The next step will be to build synthetic versions of the three data sets and test those data sets for accuracy, research utility, and security (making sure the data sets cannot be reverse engineered to reveal identities). Once they are built they can be used to respond to data requests.

To illustrate how synthetic data works, Dr. Woolley provided a small data set displayed as points on a graph and noted that those data points have a linear relationship. Synthetic data that statistically models the actual data can be created. Then an analysis of the synthetic data would yield the same linear relationship. This can be done using more variables - but with increasing complexity. Dr. Woolley noted that the synthetic data will be population average data - you will not be able to identify a school district or a postsecondary institution. In other words, the synthetic data will be used for state level analysis. No one has created cluster specific nested hierarchical synthetic data and Dr. Woolley is not sure whether it can be accomplished, but its feasibility will be investigated.

A discussion followed on whether it would be possible to provide the equation that creates the synthetic data, as opposed to the synthetic data, in response to research requests. The advantage of providing the equation is that it would completely avoid the issue of providing data. Dr. Woolley stated that it was a fascinating idea and may be possible, but not a lot of people would be able to use the equation to establish a data set for research purposes.

Mr. Rizzi asked whether the Center is doing the synthetic data project because it is an interesting academic challenge or because it addresses a need. Dr. Woolley responded that it is both a fascinating project and it is a means to solve the challenge of providing data in a safe and efficient manner. Dr. Woolley believes that U.S. Department of Education, Institute for Education Statistics funded the project because it is a promising way for the Center and other state longitudinal data systems to make longitudinal data more accessible.

Mr. Rizzi asked whether it will be acceptable to a researcher to receive synthetic data, given that it may not be fully accurate. Dr. Stapleton explained that the Bureau of Labor Statistics addresses those

concerns by running the code that creates the researcher's final analysis against the real data to confirm the outcomes.

Ms. Mullinez asked how often the synthetic data set would be updated. Dr. Woolley stated that the update process hadn't been worked out, but estimated that it would be done annually.

Ms. Bjarekull asked whether the data will be clustered at the classroom or school level. Dr. Woolley responded that whether it is possible to synthesize clustered data is not known - that will be part of the investigation. If it is possible, it may be done at the school district or institution level only. Dr. Stapleton agreed that it is a theoretical question, and noted that great care would be taken to ensure clustered data could not be identified (such as using partial sampling).

Ms. Schulz concluded the discussion and noted that it would have been great to spend time on the research output of the Center so the Board could understand how the data is being used to answer research questions.

Ms. Schulz also recommended holding off on the creation of MOUs until the June meeting so it can be approved by the Board. Mr. Goldstein responded that holding off on MOUs until June is not feasible since the grant anticipates deliverables before June. Mr. Goldstein proposed having the executive committee of the Board review and approve the MOU before it is finalized and signed. Ms. Schulz stated that she was amenable to this proposal and asked if there were any questions from the Board. There were no questions.

Adjournment

Ms. Schulz reminded the members that the next meeting will be held on June 11th at 10 a.m. at the same location and then adjourned the meeting at 12:00 p.m.

Respectfully submitted,
Ross Goldstein
Executive Director

Approved: 5/17/2016