



Maryland Longitudinal Data System Governing Board
Meeting Minutes
June 12, 2020

The meeting of the Maryland Longitudinal Data System (MLDS) Governing Board was held on June 12, 2020, via conference call. Dr. James Fielder, Chairman of the Governing Board, called the meeting to order at 9:00 a.m. and noted that a quorum was present.

The following Governing Board members were in attendance:

- Dr. James Fielder, Secretary of Higher Education
- Mr. Sam Abed, Secretary, Department of Juvenile Services
- Ms. Tiffany Robinson, Secretary of the Department of Labor
- Dr. Nancy Shapiro, Associate Vice-Chancellor for Education & Outreach (Designee for Dr. Robert Caret, Chancellor)
- Dr. Sylvia Lawson, Deputy State Superintendent for School Effectiveness (Designee for Superintendent Karen Salmon)
- Dr. Farzad Moazzami, Assistant Vice President for Academic Affairs, Morgan State University (Designee for Dr. David Wilson, President)
- Dr. Brad Phillips, Director of Research, Maryland Association of Community Colleges, (Designee for Bernard Sadusky, Executive Director)
- Ms. Sara Fidler, President of the Maryland Independent Colleges and Universities Association
- Mr. Steven Rizzi, Vice President, PAR Government
- Mr. Jason Dykstra, Executive Director, Instructional Data Division, Anne Arundel County Public Schools
- Dr. Susan Sterett, Director of the School of Public Policy, University of Maryland, Baltimore County
- Mr. Christopher Biggs, Information Assurance Manager, Raytheon Company

The following MLDS Center staff were in attendance:

- Mr. Ross Goldstein, Executive Director, MLDS Center
- Ms. Tejal Cherry, Director of System Management Branch, MLDS Center
- Dr. Angela Henneberger, Director of Research, MLDS Center and Research Assistant Professor, University of Maryland, School of Social Work
- Ms. Ann Kellogg, Director of Reporting Services, MLDS Center and MHEC Liaison
- Ms. Molly Abend, Data Management Coordinator and MSDE Liaison
- Ms. Dawn O’Croinin, Assistant Attorney General
- Mr. Roy Enehiroana, Data Analyst and Department of Labor Liaison
- Ms. James Dixon-Bobbitt, Executive Associate, MLDS Center

Approval of the December 13, 2019 Meeting Minutes

Dr. Fielder asked for a motion to approve the minutes from the December 13, 2019 meeting. Mr. Dykstra made a motion to approve the minutes, which was seconded by Dr. Shapiro. The motion was unanimously approved.

Review of Bylaws and Appointment of Vice-Chair

Mr. Goldstein stated that at the first meeting of each year the Governing Board Bylaws are reviewed and a new Vice-Chair is appointed. The Vice-Chair is on a rotating schedule among the leadership of the partner agencies (MHEC, MSDE, Labor, and USM). This year the Chancellor is scheduled to serve as the Vice-Chair. Accordingly, the Chancellor's designee, Nancy Shapiro will assume the role of Vice-Chair. Ms. O'Croinin clarified that the Vice-Chair appointment did not require a vote. However, the Board still needed to review and vote on the bylaws. Dr. Fielder, noting that the bylaws authorize the Chairman to sign contracts and approve certain expenditures, inquired whether the Board was comfortable with the level of executive privilege he was exercising. No objections were noted. Ms. O'Croinin asked whether the Secretary of the Department of Juvenile Services should be added to the Vice-Chairman rotation. Dr. Fielder noted his agreement with the recommendation. Mr. Abed also agreed and noted his willingness to share in the responsibilities. Mr. Rizzi made a motion to include the Secretary of Juvenile Services in the Vice-Chair rotation, which was seconded by Mr. Biggs. The motion was unanimously approved.

Dr. Fielder asked if there were other recommendations. Mr. Goldstein proposed working with the Research and Policy Advisory Board to review the bylaws and propose revisions and updates to the Governing Board at a subsequent meeting. Dr. Fielder agreed with the proposal.

Center Output

Research and Reporting Priorities - 2019-2020 - Year End Review

Ann Kellogg began by noting that the Center had a very productive year (2019-2020). The focus of much of the efforts was on three critical workforce areas: teaching, nursing, and computer science. For nursing, the Center conducted exploratory analysis around the Nursing Support Program funded by the Health Service Cost Review Commission. Work was also done collaboratively with various stakeholders to establish criteria for teacher supply and retention dashboards. These dashboards will be supported by new teacher education data collections initiated by MHEC. Finally, the Center continues to support the work of the Computer Science for All initiative.

The Center completed its mandatory annual reports including the *Dual Enrollment Report*, the *Career Preparation Expansion Act (CPEA) Report*, and the *Annual Report on the Maryland Longitudinal Data System and Center*. The CPEA Report added new supplemental content including the impact of student loans and their relationship to wages; enrollment patterns and duration of time in college of non-completers; and how earnings vary for high school graduates who do not go to college based on the type of high school completion (i.e. CTE, USM, etc). The Center also published several interactive reports and held its first webinar in conjunction with the Governor's Workforce Development Board to share highlights of the CPEA Report.

The Center completed two new reports: one on workforce outcomes five years after earning a high school diploma through the GED® or NEDP® programs; and another on workforce outcomes five years after earning an Associate's Degree. The Center also produced several interactive reports and reports at the request of the Department of Legislative Services, partner agencies, and other stakeholders. In addition to reports, the Center updated or expanded 45 dashboards. This productivity was aided by a new comprehensive reporting table, which includes students of interest and all dashboards generated from that population. The dashboards also embed a filtering feature that, in addition to statewide trends, allows the user to compare outcomes for a specific local school system or a specific school. Finally, the Center responded to 30 data requests, a few of which were cancelled, including two that were cancelled because there was already a dashboard responsive to the request.

In response to a question from Dr. Shapiro, Ms. Kellogg responded that the Center can and will address the impact of the Covid-19 on educational outcomes. However, the Center has to wait for the data, which is submitted to the agencies and then to the Center. Certainly the dashboards will reflect trends created by Covid-19 and the Center will provide information and context in the dashboards to help explain those trends. In response to Mr. Rizzi's question about reporting on the broader impact of Covid-19, Ms. Kellogg responded that staff is exploring what data are needed and exploring whether partner agencies are collecting additional data that may be relevant for the Center.

In response to a question from Dr. Fielder, Mr. Goldstein and Dr. Henneberger provided an overview of a grant that Dr. Fielder gave approval for the Center to conditionally support. Dr. Jane Lincove, Assistant Professor with the School of Public Policy, University of Maryland, Baltimore County and Dr. David Blazar, Assistant Professor with the College of Education, University of Maryland, College Park applied for a Spencer quick turnaround grant of \$50,000. The research would study the impact of Covid-19 on African American youth. The grant required a narrow focus, but the Center intends to conduct similar analyses on different student groups.

Next, Dr. Angela Henneberger provided an overview of the work completed by the Research Branch over the past year. First, Dr. Henneberger noted that five research reports were published in peer-review journals. The Research Branch also received a prestigious research award for advances in methodology in applied education research. There were also seven Research Series presentations, one full research report on the effects of Career and Technical Education, seven grant submissions, and six local and national presentations.

Dr. Shapiro congratulated Dr. Henneberger on the accomplishments of the Research Branch, noting the significance of the published articles and how that will help the Center gain a national standing. Dr. Shapiro asked if there was a key takeaway from the Dual Enrollment Report. Dr. Henneberger responded that the report found that dually enrolled students did better in both college enrollment and, on average, earned slightly higher wages. In addition, the study showed stronger outcomes for youth who are underrepresented in higher education and the workforce.

Research and Reporting Priorities - 2020-2021 - Proposed Priorities

Ms. Kellogg explained that staff worked with the Research and Policy Advisory Board to help develop the following priorities for the coming year:

1. Continue to focus on the critical workforce areas of nursing, teaching, and computer science;
2. Additional analysis on Associate's degrees with a specific focus on the outcomes for students who earn a career degree versus transfer degree;
3. Work on outcomes for Bachelor's degree earners, including developing measures to help provide information for the graduate follow-up survey;
4. Begin analysis of apprenticeship data, which is new to the system;
5. Continue to build out the dashboards on the high school to college series, including adding wages, and focusing on different enrollment patterns such as in-state vs. out-of-state and whether students are taking classes online;
6. Continue to work with agency partners to assist with reporting and analysis on various projects, such as analysis for the Cyber Warrior Grant Program.

In response to a question from Dr. Phillips about licensure certification at the community colleges, Ms. Kellogg responded that MHEC is working with the Community Colleges on reporting data on non-degree

seeking students. Those data will make their way to the Center and allow us to more broadly understand the contributions to the Community Colleges with helping students achieve different credentials.

Dr. Henneberger presented the research priorities developed with stakeholders. The priorities include completing in-depth statistical analyses and being able to support causal inferences where appropriate on the following topics:

1. Pathways from high school into college and the workforce;
2. Disentangling the roles of student and school poverty;
3. Teacher preparation/characteristics (e.g., demographic match to the student) and student outcomes;
4. Critical workforce areas;
5. Incorporating wage data when missing not at random;
6. The causal effect of attending a 4-year college;
7. Early elementary predictors of long-term outcomes;
8. K-12 discipline and DJS overlap; and
9. Production of high school and college students that meet workforce demands.

The second priority is to develop technical documentation to guide staff on the use of MLDS data, appropriate methods, and best practices. One example is research on the nested nature of the MLDS data and how applied researchers can statistically model students nested in schools. The final priority is to submit applications for external funding to support the first two research priorities.

Data Inventory

Ms. Abend began by noting that there are 10 proposed new data elements from the Department of Juvenile Services.

- | | |
|--------------------|---------------------------|
| 1. Youth ASSIST ID | 6. Sex |
| 2. First Name | 7. Race/Ethnicity |
| 3. Middle Name | 8. Social Security Number |
| 4. Last Name | 9. Address Type |
| 5. Date of Birth | 10. County of Residence |

This addition is a result of the 2019 legislation (HB 704) which removed the restriction against including juvenile delinquency data as part of student information. The DJS records include a Youth Assist Identification number, but no State Assigned Student ID (SASID). Mr. Dykstra asked about the need for the Social Security Number. Ms. Abend responded that only about 20% of the records have an SSN. Even though the number of SSNs is small, Mr. Goldstein noted that SSN is critical for matching data to the Unemployment Insurance wage data. Mr. Goldstein also noted that the Center has no control over whether SSN gets collected, but the Center does have the authority to collect it and that it is maintained in a highly secure and confidential manner. Mr. Abend noted that DJS is very interested in having wage information for the children that they serve. DJS wants to develop other outcome measures beyond just recidivism, which is a negative outcome. Stable employment is a good stability measure and a positive outcome that can be measured.

In response to a question from Mr. Rizzi, Ms. Abend responded that the Center does not have occupation codes for workers - only the industry in which they work. This is an important gap and is something the staff continually explores to determine if a source can be found.

Ms. Abend responded to a question in the chat feature by confirming that DJS does not collect SASID. Mr. Dykstra asked whether the Governing Board could request DJS to consider collecting it. Ms. Abend responded that currently the Center is working with the Juvenile Services Education (JSE) Division at MSDE, which does assign a SASID to the students, to determine whether the Center can use JSE data to bridge the gap.

Mr. Rizzi made a motion to approve the incorporation of the new data elements, which was seconded by Dr. Moazzami. The motion was unanimously approved.

Data Collection Calendar

Ms. Abend began by noting that the Data Collection Calendars are revised annually in consultation with the Data Governance Advisory Board. The calendars are substantially the same as last year with a few additions: the MHEC calendar now includes the Course Information System data collection and Student Registration System data collection. The calendar also notes that DJS data collection is under development.

Ms. Abend also called the Board's attention to a memorandum entitled Special Considerations for MLDSC Data Collection Calendars. The memorandum notes the potential impact on the data collection schedule due to Covid-19 and recognizes that partner agencies may require extensions or that certain data may not be collected at all.

Mr. Biggs made a motion to approve the Data Collection Calendar, which was seconded by Dr. Shapiro. The motion was unanimously approved.

Data Gap Analysis

Ms. Abend began by noting that the Data Gap Analysis was first created in 2016 at the request of the Governor's P-20 Leadership Council and is updated annually in consultation with the Research and Policy Advisory Board. The Data Gap Analysis identifies the sector, the reason for the gap, the impact of the gap, and whether there is a feasible resolution to the gap. Currently, there are 28 gaps identified. For five of the gaps, data is available and the Center is making progress in collecting the data. For six of the gaps, there is no data available. For another six gaps, data is available and under consideration. For four of the gaps, a legislative change is required before data can be collected. Finally, for the remaining seven gaps, data is available, but obtaining it is not feasible.

Next, Ms. Abend identified the data gap priorities:

1. Juvenile delinquency records
2. Licensure from Labor and other license granting agencies
3. Socioeconomic status for PK12 students
4. Unemployment applications and benefits
5. Social determinants
6. Federal and military employees

Dr. Shapiro noted that one of the research priorities is to study teacher preparation characteristics and student outcomes and suggested that the priority should be expanded to focus on teachers who are prepared in Maryland and are teaching in Maryland. Specifically, there is a need to be able to link teacher preparation programs with certification, placement of teachers, who is teaching in the schools, and understand what percent are from out of state. However, there is a teacher certification data gap, which limits the ability to identify licensure status and whether teachers are teaching in the field in which they

studied and were certified. Dr. Shapiro recommended that this be listed as a separate gap (from labor licensure) and be listed as the second-highest priority after the Juvenile Services data. Dr. Phillips agreed, noting that this data will help address the workforce shortages and gain a better handle on the subject. Dr. Fielder added that a better understanding of the teacher pipeline could help us develop better scholarship programs that require more years of service. Dr. Henneberger noted that the Research Branch is analyzing what is happening with Maryland teachers and teacher attrition, including which teachers are likely to attrite from teaching and what sectors do they go into when they leave.

Dr. Shapiro clarified that there are three things that policymakers need to know: the level of certification that each teacher has; how the teacher became certified; and the area of certification. Once this information is available, policymakers will know a lot more about what the workforce looks like and what policies are needed to address workforce demands.

Ms. Kellogg noted that there was a chat question about the business license gap. Ms. Abend responded that proposed legislation (HB 850) would have removed MHEC as the data collector and had MLDS Center collect it directly. It also clarified that the Department of Health is authorized to send PII. In response to a question from Dr. Fielder, Ms. Kellogg noted that the licensure data will be used to understand high school student outcomes. Specifically, it would provide an understanding of the effectiveness of vocational programs for high school graduates who do not attend college. Ms. Abend noted that the licensure data would also provide insights into whether CTE programs lead to certification in related fields.

Policy on External Researcher and Grant Funded Projects - Proposed Revisions

Mr. Goldstein began by noting that the Center has a detailed process for approving research that is being funded by a grant or being proposed by an external researcher (i.e. not a member of our research branch). The current process has been in place for two years and provides a lot of transparency and accountability. However, it is also very time consuming for the Center, the researchers, and the reviewers. Because most of the projects have grant funding attached, the majority do not come to fruition because funding is not received. In addition, the process requires Governing Board approval. There is an exception that allows the Chairman to approve the application, which happens more often than not due to the quarterly Board meeting schedule.

The goal was to find some ways to streamline the process while providing the same level of scrutiny and accountability to stakeholders. Mr. Goldstein noted that he provided a memorandum to the Board in advance of the meeting that detailed each specific change to the procedures and therefore would only review the highlights.

The application process eliminates the formal research proposal, which was a 10-15 page narrative of the project. The online application requires the applicant to provide all the necessary information to make a determination. Further, applicants make a presentation to the Research and Policy Advisory Board, which provides an opportunity to supplement information and respond in depth to questions as needed.

The current review process requires an internal review, Research and Policy Advisory Board review, partner agency notification, and final determination by the Governing Board. The proposed procedures would create an expedited review process - which would allow me to approve an application instead of the Board. The expedited review process can be used if certain conditions are met. Those conditions are:

1. The principal investigator is a member of the Research Branch or a MLDS Collaborator (which is defined as faculty at Maryland College or University and frequent staff appointment);

2. There is no grant funding or the funder is on a pre-approved funder list;
3. The research is a *standard topic*, which is a topic that is responsive to the Center's research priorities, similar to a topic already approved, or builds upon prior work of the Center (In comparison, there are also *unique topics*, which present novel questions, methodological studies, or address sensitive topics or plans to analyze and report on data in a way that may be contrary to the manner in which stakeholders generally report such data.);
4. A unique topic may be subject to the expedited review process only if it is being done at the request of a stakeholder.

During discussions with the Research and Policy Advisory Board, an off-ramp was created for the expedited review. Specifically, a full review will be used at the request of one of the partner agencies. All other proposals that don't meet the exception for an expedited review would go through the full review process.

Another step to the procedures would be to allow the executive director to approve the resubmission of a previously approved project as long as (i.e. if the grant is not approved)

1. The subject is substantially the same;
2. The same principal investigator is applying; and
3. The same funder or another funder on the pre-approved list is being used.

The revisions would also allow the executive director to send a conditional letter of support that clarifies that Governing Board approval is required.

Finally, Mr. Goldstein provided an overview of the approved funders list, which includes:

1. Federal Government
2. State and local government, including local school systems;
3. Maryland public colleges; and
4. Private foundations.

The proposal is for funders to remain on the list for five years. Of course, the Board can review the list at any time and add to it or remove funders from it.

In response to a question from Dr. Fielder, Mr. Goldstein noted that the current procedures have a provision that allows the Center to seek reimbursement for costs associated with providing access to the system and data. Dr. Sterett commented that the expedited review process was a good idea, especially in light of some of the very short grant application deadlines. She also commented on the fact that the expedited process will signal to the research community that the Center is a good partner for research. Finally, Dr. Sterett noted that the funders on the approved funders list represent the elite funders in the field.

Mr. Goldstein noted that in general, the focus of the procedures is to ensure that all research has a value to Maryland and the Center, including the requirement that there must always be a Center product.

Ms. O'Croinin announced that Ms. Robinson left the meeting, but there was still a quorum.

Dr. Phillips made a motion to adopt the proposed revisions, which was seconded by Mr. Biggs. The motion was unanimously approved.

Policy on External Data - New Proposed Policy

Mr. Goldstein stated that the Center has developed a set of proposed procedures for adding new external data to the Center's Data Inventory. External data are data provided for a unique study or program evaluation. This is not data from a partner agency. For example, the Baltimore Education Research Consortium has proposed a research project that would use Baltimore City high school choice data.

The procedures outline the legal requirements. First, Governing Board approval is required for all external data. Second, the data must meet the definition of workforce and student data. Third, the data could be subject to a Public Information Act (PIA) request - with the same limitations that the request must be for cross-sector data and can only be fulfilled with aggregate, de-identified data. Finally, the access and use restrictions still apply. Specifically, once the data is provided to the Center, the Center staff load it and the researcher only has access to de-identified data.

The procedures also outline the request process, which requires the applicant to provide information about:

1. The number of records;
2. How the data is collected and maintained;
3. Whether there are unique identifiers to allow matching; and
4. Other information to allow the Center to assess the scope and quality of the data.

Next, the procedures establish a review process and some of the factors that will be considered. The review process includes:

1. Staff review;
2. Data Governance Advisory Board review;
3. Notification to Partner Agencies; and
4. Governing Board approval.

The policy establishes two types of data. Permanent data for when the intention is for the data to remain in the MLDS - perhaps for a follow-up study or to allow future use by other researchers. Temporary data for when data are only for one project and then removed from data inventory. The policy also establishes Institutional Review Board requirements for the data provider. Specifically, the data provider must provide documentation that an institutional review board has approved the inclusion of the data into the MLDS and acknowledges that MLDS staff will have access to the personally identifiable data and that data may be subject to a PIA request. The procedures also specify the requirement for a data sharing agreement. Finally, the procedures specify how the data will be included in the data inventory. Permanent data will be incorporated by sector into the data inventory and temporary data will be listed in a supplemental section with the dates it will be added and removed.

In response to a question from Mr. Biggs, Mr. Goldstein clarified that MLDS staff have access to the personally identifiable information in the external data. Mr. Goldstein also noted that the procedures include a reimbursement section, which is important given the time consuming and technically complicated process of adding new data into the system.

Mr. Biggs made a motion to approve the Procedures for External data, which was seconded by Dr. Shapiro. The procedures were unanimously approved.

External Research and Grant Funded Projects

Dr. Henneberger presented updates on a few external research proposals. First, is a grant submission by

herself and Dr. Mathew Uretsky. The project was going to be presented in March, but due to the cancellation of that meeting, approval was provided by the Chairman. A \$25,000 grant was awarded from the Institute for Research on Poverty. The research will study patterns and classes of people who are participating in school and the workforce in Maryland. The central hypothesis is that there are distinct classes of students whose patterns of student academic participation and Maryland workforce participation during high school relate to later academic and workforce outcomes. The study will look at the likelihood of membership in the different classes by key student subgroups: race, ethnicity, and socio-economic status. The third hypothesis is that membership in classes will predict academic (such as degree completion) and workforce outcomes (such as wages) six years after ninth grade.

Dr. Henneberger went on to note that the project is cross-sector and will use data from K-12, postsecondary, and the workforce. The project is also responsive to the Center's Research Agenda. The project will be beneficial to the state by allowing policymakers to better understand outcomes for specific subgroups of youth to help improve college and career readiness. The study can also help policymakers better understand realistic expectations about academic and workforce demands, especially when they occur at the same time. Finally, the research can ultimately help improve high school and workforce outcomes in Maryland.

Dr. Phillips asked when they could see the report? Dr. Henneberger responded that there is not a report at this time - the research is first getting started. However, this is part of a long line of research that they have been focused on and there are older research reports that can be provided. The grant is for one year, therefore Dr. Henneberger stated that a Research Series presentation in the fall or early spring is planned to get feedback.

Next, Dr. Henneberger gave updates on a few other projects and noted her thanks to the Board for their support for this work.

1. Institute for Education Sciences (IES) - A proposal will be submitted to conduct a long-term follow-up examining outcomes for students involved in PBIS Tiers 1 and 2. The submission is due August 20, 2020. In response to a question from Dr. Fielder, Dr. Henneberger clarified that PBIS - is the Positive Behavioral Interventions and Supports program, which is a tiered public health approach to the prevention of problem behavior. Tier 1 is a general prevention model and Tier 2 is a targeted model.
2. IES - Effects of Career and Technical Education on Student Outcomes - A proposal will be submitted by Dr. Marc Stein and Dr. Rachel Durham (Johns Hopkins University and the Baltimore Education Research Consortium) on August 20, 2020. This proposal was approved last year and is being re-submitted after very favorable reviews from the first submission.
3. Spencer Foundation - Covid 19 Request for Proposals - This was a rapid turnaround submission. The project will be brought before the Board if a grant is awarded.

MLDS Center FY 21 Budget Plan

Mr. Goldstein began by noting that the FY 21 Operating Budget is \$2.4 million, which represents an increase of \$500,000 from last year. However, like all state agencies, MLDS Center received a 10% mandatory budget reduction. The Center can accommodate the reduction and still fully fund our operations because of the overall increase in funding and because the Center's Oracle license was paid this year and will last through most of FY21. Last year the Center could not fully fund the research branch budget, which was reduced by \$40,000. This year, the Center was able to restore that additional funding. However, after the memorandum was written, DBM informed the Center that Interagency

Agreements could not include an increase in total funding. Accordingly, the research budget will be \$345,000 instead of \$365,000. It is possible that the Center can still use the remaining \$20,000 to fund research - including funding a research project with one of Maryland's Historically Black Colleges and Universities (HBCU). DBM has a HBCU goal that agencies are supposed to work towards.

Finally, Mr. Goldstein presented the breakdown of the agency's budget:

- a. 61% - Salaries
- b. 12% - IT and operations
- c. 15% - Research
- d. 10% - Reduction

In response to a question from Dr. Phillips, Mr. Goldstein stated the reduction was dictated by DBM as a result of the economic downturn caused by Covid-19. More cuts may be coming.

Mr. Dykstra made a motion to approve the budget plan, which was seconded by Dr. Shapiro. The motion was unanimously approved.

Synthetic Data Project - Final Report

Dr. Michael Woolley thanked the Governing Board for the opportunity to provide this final update. This marks the end of a four-year project, which brought together a terrific team. Today's presentation will focus on a broad project overview and some final thoughts on using synthetic data as a method for sharing data.

The Synthetic Data Project was a result of MSDE's invitation to the MLDS Center to include a research proposal as part of MSDE's application for a State Longitudinal Data System grant from the U.S. Department of Education. MSDE was funded \$7 million, including the MLDS Center proposed \$2.6 million project entitled, *Expanding MLDS Data Access and Research Capacity with Synthetic Data Sets* (referred to as the Synthetic Data Project - SDP - for short).

The central aim of SDP was to explore the feasibility of synthetic data as a data-sharing strategy for the MLDS Center. Dr. Woolley stated that the project has successfully demonstrated that synthetic data are a feasible strategy for the MLDS Center.

The project was divided into four main tasks:

1. Creating Gold Standard Datasets (GSDS);
2. Synthesizing those GSDS;
3. Evaluating the research utility of the synthesized GSDS; and
4. Determining whether they present a risk of disclosing sensitive student or worker information.

Currently, the project has initiated a final step of handing off the project, including the infrastructure, procedures, code, and processes to the MLDS Center staff.

Dr. Woolley noted that the Center's operational data store is very large and complex and therefore they needed to create smaller and less complex data sets to synthesize. They created the GSDS based on a review of MLDS related research and consultation with end-users, key stakeholders, and experts.

Ultimately, they created three GSDS using data from 2010-11 through 2015-16. The data sets include:

1. High School to Workforce
2. High School to Postsecondary (PS)

3. Postsecondary to Workforce

Creating the synthetic data required a review of the different synthesization models and identifying a model that could address the Center's data, which include both categorical variables and numerical values. The researchers chose the Classification and Regression Trees (CART) Model. Using the CART model they were able to impute synthetic versions of the GSDS. Assessing research utility asks the question of how well does the synthetic data reflect the characteristics of the GSDS. The findings were that the variable characteristics were well preserved and weaker effects were also well preserved. Stronger effects were slightly attenuated. For example, factors that have a direct impact on wages were still observable and in the same proportion - but to a lesser degree.

Dr. Shapiro asked whether synthetic data could be used to analyze the impact of Covid-19 on education outcomes. Dr. Woolley and Dr. Stapleton responded that if the data needed to conduct the analyses are synthesized, then it can be used. It may be possible to use data to conduct predictive analyses, but that would be dependent on assumptions and simulated outcomes. Generally, synthetic data can be used in the same ways as the real data. Dr. Lawson recommended that before discussion on what to study and analyze using the synthetic data, MSDE's research director should be consulted.

Ms. O'Croinin noted that Mr. Abed left the meeting, but a quorum was still present.

The final phase of the work was assessing the disclosure risk. The data was determined to be quite safe. The synthetic data that they created are fully synthetic (as opposed to keeping certain attributes). There are two types of disclosure risk that have to be considered. First, is Attribute Disclosure Risk. The researchers were able to determine with confidence that sensitive values for any one member of a rare group are protected. Rare groups were defined as a group with ten or few individuals who share seven demographic characteristics. Even within those small groups, across all 20 synthetic data sets, there is zero ability to identify sensitive values for any one individual. Second, is Identification Disclosure Risk. The question is whether you can take synthetic data and find an actual individual in the synthetic data. The analysis confirmed that there are no actual individuals in the synthetic data. The data is very safe and, Dr. Woolley asserted, slightly safer than aggregated data sets.

Dr. Woolley next reiterated the steps to hand-off the synthetic data project to Center staff. There were a series of meetings on the data warehouse, the process of creating the GSDS, the process of creating synthetic data, and the process of assessing the data.

Informing the governing board - what do you need to know

1. Safe and could be used to respond to a data request. You could give it out to end researchers and analysts for research purposes. Employers could use it to learn about students
2. Robust for research - things you learn from the data are the same.
3. Leverage Center Data - increase what we learn from Center data - more information developed to better inform policy
4. Resources are required to maintain and continue to develop - even with infrastructure already created - it would take resources (maybe through a grant)

Ms. O'Croinin noted that Mr. Biggs left the meeting, but a quorum was still present.

Mr. Rizzi thanked the team of researchers for their great work and the outcomes they produced. Mr. Rizzi noted that the ongoing concern is that making synthetic data widely and easily available would remove

the scrutiny and stewardship over the data and the research that the Center and Board currently employ. Dr. Woolley noted that this is a legitimate concern, but noted that people can take real data and draw wrong conclusions. Dr. Woolley noted that research based on synthetic data could not be published - the analyses would have to be conducted on the real data. That would, therefore, allow the Center to continue to maintain some control over the research process. Dr. Woolley also noted that researchers are subject to rigorous ethics requirements. Finally, Dr. Woolley and Dr. Terry Shaw met with the U.S. Census Bureau, which has been using synthetic data for several years. Representatives from the Bureau did not report any issues with problematic or irresponsible research using the synthetic data.

Mr. Goldstein pointed out that under the Public Information Act, the Center must fulfill data requests, which could create the same set of concerns about irresponsible use of data. Mr. Goldstein went on to state that now that the project is complete and the feasibility of synthetic data has been proven, the next question is what the Center will do with the data. There are a lot of options, including taking an incremental approach. For example, as proposed by Dr. Henneberger, the synthetic data could be used as a teaching tool for students. The Center could build from there and find other opportunities to integrate the synthetic data into Center operations. As uses for the synthetic data come up, staff will come back to the Governing Board for approval.

Dr. Phillips asked to see more comparisons between gold standard data and synthetic data. Dr. Mark Lachowicz noted that the analysis included:

- Group and mean differences
- Strength of relationship - bivariate
- Strength of relationship - multivariate
- Longitudinal associations

Dr. Lachowicz then showed a graph with results of wages for a postsecondary cohort. The graph demonstrated the overlapping values between the GSDS and the synthetic data. It also showed one variable where results did not overlap. In that instance, the team was able to adjust the synthesization model to improve the performance of the synthetic data.

Ms. O'Croinin noted that Dr. Phillips left the meeting, but a quorum was still present.

Closing

After confirming there were no more questions for Dr. Woolley and the team, Dr. Fielder thanked them for their work and announced that the next meeting would be held on September 11, 2020, hopefully in person in the Boardroom. A motion to adjourn was made by Mr. Dykstra and seconded by Dr. Shapiro. The motion was unanimously approved.

Respectfully submitted,
Ross Goldstein
Executive Director

Approved: September 11, 2020