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 Governing Board
# Meeting Minutes
# September 13, 2019

The meeting of the Maryland Longitudinal Data System (MLDS) Governing Board was held on September 13, 2019, in the Maryland State Department of Education Board Room, at the Nancy S. Grasmick Building. Mr. 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. Michael Harrison, Director of the Office of Policy Development (Designee for Ms. Tiffany Robinson, Secretary of the Department of Labor)
Mr. Chad Muntz, Assistant Vice Chancellor of Institutional Research, Data & Analytics (Designee for Dr. Robert Caret, Chancellor)
Dr. Sylvia Lawson, Deputy State Superintendent for School Effectiveness, & Chief Performance Officer (Designee for Superintendent Karen Salmon)
Dr. Farzad Moazzami, Acting Assistant Vice President for Academic Affairs and Associate Professor of Engineer, Morgan State University (Designee for Dr. David Wilson, President)
Ms. Tina Bjarekull, President of the Maryland Independent Colleges and Universities Association
Dr. Brad Phillips, Research and Policy Director, Maryland Association of Community Colleges (Designee for Dr. Bernard Sadusky, Executive Director, MACC)
Mr. Christopher J. Biggs, Information Assurance Manager, Raytheon Company
Mr. Steven Rizzi, Vice President, PAR Government

**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
Dr. David Blazar, Investigator, MLDS Center Research Branch and Assistant Professor of Education Policy and Economics at the University of Maryland
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 DLLR Liaison
Ms. Jamese Dixon-Bobbitt, Executive Associate, MLDS Center

**Approval of the June 14, 2019 Meeting Minutes**
Dr. Fielder asked for a motion to approve the minutes from the June 14, 2019 meeting. Dr. Lawson made a motion to approve the minutes, which was seconded by Mr. Rizzi. The motion was unanimously approved.

**Synthetic Data Project**

Mr. Goldstein noted that the Synthetic Data Project (SDP) is entering its fourth and final year. The research team is presenting an update and new plans for additional testing, for which they are requesting Governing Board approval. Dr. Michael Woolley began by reminding the Board that at the last presentation, the research team had reached a major milestone; successfully creating three synthetic data sets. Since that time the team has been studying the research utility of the synthetic data sets and whether they create any disclosure risk.

Dr. Woolley next provided an overview of the project to date. The first step of the project was to create three gold standard data sets. Second, they created the synthetic data sets out of the gold standard data sets. Third, they tested the synthetic data to determine if they are good for research and safe for release. The team now recommends conducting a field test for research utility and disclosure risk. Finally, they will be back to discuss the results of the field test and discuss whether and how the synthetic data should be publicly released.

Dr. Woolley provided definitions and explanations of various terms.

1. Research Utility - simply whether analyses using the synthetic data produce the "right" answer - one that is consistent with analyses on the real data.
2. Disclosure Risk - a determination of whether the synthetic data pose a risk of disclosure of confidential protected information.
3. "Gold Standard Data Sets" are simplified and streamlined versions of the data housed in the MLDSC.
4. Synthetic Data Set – data sets created from a computational model such that when statistically analyzed will act like the real data. There is no real data in the synthetic data sets. They are not synthesizing people, they are synthesizing the data set.
5. Fully Synthetic Data Sets – data sets comprised completely of synthesized variables.
6. Partially Synthetic Data Sets – a combination of synthesized and non-synthesized variables (i.e. variables that have the original or "real" values).

Next, Dr. Woolley provided the following facts about the synthetic data project:

1. There is no data from independent colleges or universities;
2. All variables in all synthetic data sets are synthesized (i.e. a fully synthetic model);
3. There are no ID numbers or identifying information carried over from GSDS to synthetic data;
4. There is no geographic information in the synthetic data - such as school districts, zip codes, or census tract; and
5. There is no identifying information about schools, colleges, universities, or employers in the synthetic data.

Turning to the beta test, Dr. Woolley explained that while the SDP team has done extensive testing, it is still beneficial to gain outside perspectives and input. Beta testers will be set up as affiliate researchers of the MLDSC through the standard procedures which include criminal history background checks, security training, and non-disclosure agreements. Beta Testers will only have access to the synthetic data sets.

Six to eight beta testers will conduct tests on research utility and two beta testers will conduct tests on disclosure risk. Up to four of the beta testers will be individuals who are already affiliated researchers. Dr. Mark Lachowicz, an SDP team member in charge of research utility testing, provided an overview of how research utility has been tested and the results so far. At a minimum, synthetic data should "look" similar to the gold standard data sets. Results of specific analyses conducted on both the gold and

synthetic data sets should be similar. The approach used was informed by the U.S. Census Bureau and involved analyzing all of the following factors:

- Variable distributions;
- Means and variability;
- Proportions of missing values;
- Group mean differences;
- Strength of relationship - bivariate;
- Strength of relationship - multivariate; and
- Longitudinal associations.

Each of the tests made comparisons against the gold standard data sets. Dr. Lachowicz also pointed out that the results of testing helped inform and refine the synthesization The beta test will continue the testing process and provide input on whether the data elements chosen for the gold standard data sets and synthesization are useful to a broader audience of actual researchers and policymakers.

Ms. Bjarekull asked what happens if the correlations aren't correct; are adjustments made to the synthetic data and could that lead to concerns that the synthetic data are too closely correlated to the real data? Dr. Lachowicz responded that if there are too many discrepancies, then the synthetic data models (i.e. the algorithms used to create the synthetic data) can be adjusted. However, it is a balancing act; if the correlation is too closely aligned the disclosure risk increases.

Dr. Terry Shaw, SDP team member responsible for leading the disclosure risk testing began by explaining the two types of possible disclosures:

1. Identification Disclosure - The potential for an intruder to match a given record with a specific individual; and
2. Attribute Disclosure - The potential for an intruder to identify sensitive characteristics of small subpopulations.

To analyze identification and attribute disclosure risk the assumption is the intruder has an external data set to help determine identifications or attributes. (For example, the external data might contain students and test scores - can that data be used with the synthetic data to learn the students' wages?) For this analysis, the team used the gold standard data set as the 'external dataset.' This assumes a worst-case situation where an intruder might know almost everything about specific individuals or subgroups. The testing focused on whether there is a case in the synthetic data that looks like an actual individual in the gold standard data set? And are there group attributes in the synthetic data that look like the attributes of actual groups in the gold standard data?

The conclusion is that there is an extremely low disclosure risk. Dr. Shaw was not able to find a record in the gold standard data set that matched a record in the synthetic data. To test attribute disclosure, Dr. Shaw looked across a number of demographic attributes and found a 1% chance of finding a series of matching attributes in both the gold standard data and synthetic data. However, Dr. Shaw was unable to find any matching grade or wage information between the two data sets. Dr. Shaw concluded by noting that he is very comfortable that the synthetic data does not pose a disclosure risk. Beta testing is still a good idea, especially with external users who can do in-depth machine learning type attempts to match

data and come up with sensitive information. Dr. Shaw is confident that they won't be able to find matches or sensitive information.

In response to a question from Dr. Fielder, Dr. Shaw confirmed that more variables in a data set will result in less chance of disclosure risk. Dr. Woolley noted that the U.S. Census Bureau, with whom he and Dr. Shaw met, states that their fully synthesized data set has no risk. Mr. Goldstein noted that in addition to the beta testing, the synthetic data will also be reviewed by the U.S. Department of Educations Privacy Technical Assistance Center. They will look at the synthetic data, the testing that has been done and provide their analysis of the disclosure risk.

In response to a question from Mr. Rizzi, Dr. Shaw clarified that synthetic data is not just a blurring of information. Every variable is created based on distributions in relation to every other variable. While small populations present challenges, the synthetic data works at a population level and the process of developing the synthetic data was structured to avoid small populations - in fact, any groups or nodes with less than 30 people were not included in the synthetic data.

In response to a question from Mr. Muntz, Dr. Shaw stated that identification disclosure was not based solely on a one-to-one match. The nearest neighbor approach was used, which means that if everything matches except one attribute, it would still be considered a match. Even using this standard, no matches were found.

Mr. Muntz asked whether someone could use the synthetic data to predict certain information. Dr. Shaw responded that it was possible, but noted that it would not reveal information about individuals that are private or confidential You get information about the population- not individuals. Mr. Muntz noted that the MLDS and now the synthetic data are providing information about Maryland citizens that weren't available before. Are there things that can be learned from the synthetic data that will make people uncomfortable? Dr. Woolley noted the Center has to provide aggregate data sets in response to public requests and there is no meaningful difference between what can be discovered by aggregate data as compared to what can be learned from access to the synthetic data.

In response to a question from Dr. Fielder, Dr. Woolley responded that the testing will cost around $20 thousand and will be fully paid for from the federal SLDS grant funds. The costs include a small stipend for the testers and funds to support efforts of the researchers to organize the test, train the testers, and study the results.

In response to a question from Dr. Moazzami, Dr. Shaw responded that access to the algorithms would not allow someone to reconstitute the actual data because there are random processes built into the synthetic data production process.

In response to a question from Mr. Biggs, Dr. Woolley explained that the beta testers are known professional associates with whom members of the research team have worked. All of the beta testers will go through the Center's standard vetting process.

Mr. Biggs next asked whether unique academic programs could result in identifying students. For example, only Morgan State University offers an architecture program. Ms. Kellogg responded that the synthetic data used truncated academic program information (2 digit CIP). As such, a program like architecture would fall under a broader category of applied engineering or applied arts. Dr. Woolley

agreed with Ms. Kellogg's assessment but also committed to further review and testing (including beta testing) to confirm that unique programs will not result in disclosure.

Mr. Harrison asked whether the synthetic data set that has been created is a one-time endeavor or will it be maintained and updated, including incorporating the new variables that the Center will begin collecting. Dr. Woolley responded by noting that they have built infrastructure to make this sustainable. Going forward, new values would only be added and synthesized if approved by the Governing Board. Dr. Woolley also noted that the availability of synthetic data may reduce the number of aggregate data requests - which require a lot of time and effort of the Center.

Ms. O'Croinin responded to a question from Ms. Bjarekull by confirming that the Center has a duty to provide aggregate data in response to data requests. When complete and if testing confirms that there are no disclosure risks the synthetic data will have to be provided upon request or can be made publically available.

In response to a question from Mr. Rizzi about the ownership of intellectual property of the synthetic data, Ms. O'Croinin explained that the MLDS data is the state's data and the derived product from that data, in this case, the synthetic data, is a state product. There are attribution requirements dictated by the federal government. Mr. Goldstein noted that currently there are no policies and procedures developed for administering the synthetic data - assuming that it is approved for public release. There is only one set of synthetic data created and there is no requirement to continue to create synthetic data. This current grant was a feasibility study. There are a lot of questions that have to be answered - for example, how do we distribute this one set of synthetic data and when do we allow researchers to run their code against the real data set to confirm and publish results? The utility of the synthetic data will depend on how the Center addresses those questions. Ms. O'Croinin suggested that additional issues and questions should be referred to the Center's advisory boards.

Mr. Biggs made a motion to approve a beta test of the synthetic data, which was seconded by Dr. Moazzami. The motion was unanimously approved.

**Data Inventory and Collection Calendar**

Additions to the Data Inventory

Ms. Molly Abend began by explaining that there were three new categories of data proposed for addition to the Data Inventory: Discipline data, Gifted and Talented Indicator; and LACES data.

Ms. Abend introduced Douglas Weimer, Adult Education Program Specialist with the Division of Workforce Development and Adult Learning (DWDAL) at the Department of Labor to provide an overview of the LACES program and the data elements proposed to be added. Ms. Abend also noted that the inclusion of LACES is already permitted under the data sharing agreement with the Department of Labor.

Mr. Weimer began by providing an overview of adult education in Maryland, which is overseen by the (DWDAL). The program is funded by the Workforce Innovation and Opportunity Act (WIOA Title 2 Funds) with federal oversight from the U.S. Department of Education, Office of Career Technical and Adult Education (OCTAE). The program funds 26 grantees throughout the state, including community colleges, local providers, and local school systems. The programs include adult literacy, GED® preparation, and English language acquisition. LACES, which stands for Literacy, Adult, Community Education System, is the database used to collect program information. It is owned and maintained by a

private vendor. The database collects information on students including demographics, attendance, courses, completion, the achievement of measurable skills gained, and transitions into workforce development or training program. LACES produce reports required for the national reporting system.

Mr. Weimer noted that improved data sharing will benefit research and strengthen reporting capabilities. Currently, federal reporting required by OCTAE includes reporting on wage records for the second quarter after exiting a program. This is currently difficult to do and requires DWDAL to either rely on survey data, which is unreliable or pay a third party to conduct the data matching to wage records. This partnership with the Center will also allow policymakers to gain insights into an important part of the educational system - adult students - which is not currently being told.

Ms. Abend noted that there are 109 new LACES data elements proposed for inclusion in the Data Inventory. The data elements have been categorized into five categories: Demographics, Class/Course Information, Assessments, Outcomes, and Intake/Enrollment.

In response to a question from Dr. Phillips, Mr. Weimer explained that LACES data is provided by the grantees who enter the information on the students who enroll in their programs.

In response to a question from Mr. Harrison, Ms. O'Croinin responded that the data-sharing agreement already includes LACES data. Mr. Goldstein also noted that the data sharing agreements specify the broad classes of data - not specific data elements. The Center has been receiving LACES data, just not to this level of detail.

Finally, in response to a question about apprenticeship data, Mr. Weimer responded that having both LACES data and the apprenticeship data (which was approved at the last meeting) will allow these interconnected programs to be studied and allow a better understanding of important transitions to different programs.

Next Ms. Abend presented the second new class of data - discipline data. There are eight discipline data elements being collected from MSDE. During the 2019 legislative session House Bill 704 passed, which removed the restriction against the Center collecting student discipline data. Ms. Abend explained that the 504 Indicator is a yes/no indicator of whether the student has a 504 plan (an accommodation for a disability). The Disability Code is a designation of the disability the student has but does not include information about services or treatment. The Offense Code provides information about the event that resulted in disciplinary action.

In response to a question from Mr. Rizzi, Ms. O'Croinin explained that the disability-related elements are part of the discipline data because they are required elements for federal reporting. The issue is whether students with disabilities are being disproportionately disciplined.

The final data element is the gifted and talented flag. MSDE is adding this indicator to the End of Year Attendance data collection in order to comply with accountability requirements. This will not go into effect until 2020-2021 academic year but is being added to the MLDS data inventory now to make sure it does not get overlooked once implemented.

Dr. Phillips made a motion to approve the inclusion of the proposed data elements in the MLDS Data Inventory. The motion was seconded by Mr. Biggs and unanimously approved.

Data Collection Calendar

Ms. Abend explained that the data collection calendar is being updated to include the collection schedule for the discipline data and the apprenticeship data. Both files are collected and will be provided to MLDSC twice a year - February and August for the apprenticeship data and April and October for the discipline data.

Mr. Muntz made a motion to approve the changes to the data collection calendar, which was seconded by Dr. Moazzami. The motion was unanimously approved.

**Dual Enrollment Report**

Ms. Kellogg began by noting that this is the Center's seventh *Annual Dual Enrollment Report.* Dual enrollment continues to increase, with an additional 1,500 students participating - which constitutes a 13% increase over the prior year.

Ms. Kellogg presented a series of graphs:

- The first graph, *High School Dual Enrollment Trends*, shows the four-fold increase in dual enrollment from the start of the reporting in 2010-2011 through 2017-2018.
- The second graph, *Dual Enrollment Trends: Race/Ethnicity*, shows the distributions of dually enrolled students among racial and ethnic groups. Dually enrolled students are predominately White. The number of Hispanic and Black students who are dually enrolled is increasing. However, the percentage of each group participation has remained flat. For example, Hispanics make up 16% of all students, but only 7% of dually enrolled students. Similarly, African Americans make up a third of all students, but only 26% percent of dually enrolled students.
- The third graph, *Dual Enrollment Trends: FARMS and Non-FARMS*, shows that the dual enrollment rate for FARMS students increased two percentage points in 2017-2018. However, like the race and ethnicity trends, the percentage of FARMS students has remained flat. FARMS students make up a third of the student population but only 21% of dually enrolled students.
- The fourth graph, *Dual Enrollment Trends: Course Subjects,* shows that the predominant courses are English and Social Sciences. Mathematics and Career and Technical Education Related courses are the next most popular courses.
- The fifth graph, *Dual Enrollment Trends: College Enrollment*, shows that 60% of students who were dually enrolled also enrolled in college compared to 42% of students who were never dually enrolled.

Finally, Ms. Kellogg discussed the future of the *Dual Enrollment Report*. The goal is to look at the report with stakeholders and make sure that the report is addressing their needs and answering the right questions. Statutorily, the report must address the number of students engaged in dual enrollment and the courses they are taking. After that, there is a lot of flexibility in what to report.

In response to a question from Dr. Phillips, Ms. Kellogg stated that the Center is working on reporting the number of total credits earned and is also interested in analyzing course completion versus just reporting on courses taken. Dr. Phillips commented that it would also be good to understand whether students are taking courses pursuant to the negotiated arrangements between school systems and community colleges. Those agreements specify who is eligible and those eligibility requirements may impact participation.

In response to a question from Mr. Harrison, Ms. Kellogg responded that the college-going rate represents first-time enrollees in the fall and that there is some evidence that students tend to enroll in the school in

which they were dually enrolled in high school. For some students, this may result in an undermatch, which would be considered a negative outcome.

Dr. Henneberger pointed out that in addition to the *Dual Enrollment Report*, there is also a research report on dual enrollment that used propensity score matching to confirm that there is a positive impact on college enrollment, graduation, and wages after graduation.

**External Research and Grant Funded Projects**
Mr. Goldstein began by noting that there are three presentations. The first two presentations are grant proposals that have been reviewed and approved by the Chairman and are being presented for review. The third grant proposal is new and requires Board approval.

Postsecondary and Labor Market Effects of Career and Technical Education
Mr. Goldstein introduced Dr. Marc Stein, who is one of the principal investigators on the proposed project and an Associate Professor at the Johns Hopkins University, School of Education and a researcher with the Baltimore Education Research Consortium (BERC). Also with Dr. Stein is Dr. Faith Connolly, the director of BERC. Mr. Goldstein also noted that Dr. Rachel Durham is also one of the principal investigators on the project and a frequent external researcher with the Center.

Dr. Stein began by noting that postsecondary and workforce outcomes of CTE is a growing area of research, but one with a lot of important questions still needing to be addressed. BERC has applied to the Institute for Education Sciences' specific call for studies on CTE.

In Baltimore City, most schools offer at least one or two CTE courses. Two recent reports on CTE have failed to explain why CTE programs get the results they do. This study seeks to address that issue.

Specifically, the study will address the following research questions:

1. Who participates in CTE - who applies and who completes?
2. How do the theorized mechanisms of the effect of CTE participation manifest themselves empirically in the Baltimore context? - Engagement and relevance, skill development, academic skill development.
3. What is the effect of CTE participation on student secondary and postsecondary outcomes, including graduation, postsecondary matriculation and persistence, labor market participation and wages?
4. To what extent do the outcomes vary as a function of student characteristics and CTE program of study?

To do this analysis, the researchers plan to take advantage of the CTE enrollment process in Baltimore City schools. A centralized enrollment process is used as part of a universal high school choice. Students who choose programs in selective CTE Centers are prioritized by a composite score and are offered enrollment until all seats in the program are filled. The project will use these admissions cutoffs to compare students who were selected for CTE participation to students with similar scores who were not selected. This will allow them to use a regression discontinuity design to look at the effect of the CTE Centers.

Dr. Stein noted that the project has a lot of potential to provide a causal estimate of the impact of CTE and will help provide an understanding of how CTE produces positive outcomes for kids and the differential outcomes of specific programs.

Dr. Fielder asked whether the CTE outcomes will include youth apprenticeship programs. Dr. Stein agreed that youth apprenticeship would be an interesting area to explore. Mr. Harrison noted that the youth apprenticeships are a certified CTE pathway and should be included with other CTE data.

Mr. Goldstein noted that the Center does not currently have data on Baltimore City student school choice preferences. The Center is developing procedures for how to manage requests to include specific data for a research project. The staff will bring procedures to the Board for consideration and the Board will have an opportunity to review and approve the new proposed data elements for inclusion.

Mr. Rizzi asked whether this was the first proposal to include a researcher from an out-of-state institution? Dr. Stein clarified that the out-of-state researcher will help with the project, but will not need access to the data system. Mr. Goldstein noted that this is the first external proposal to include an out-of-state researcher, but it is permissible under the procedures established by the Board. Specifically, an out-of-state researcher can become an external researcher as long as the project includes a Maryland sponsor.

Understanding How Modern Methods in Data Science Should be Used in Education

Mr. Goldstein stated that the second proposal is a project to understand how modern methods in data science should be used in education. Dr. Tracy Sweet is an Assistant Professor in the College of Education at the University of Maryland, a member of the Research Branch, and the principal investigator on the project.

Dr. Sweet began her presentation by noting that the project proposal was submitted to IES for a 3 yr methodological study. The project has four primary aims:

1. Exploration of data science techniques, including clustering, dimension reduction, and network analysis;
2. Prediction using machine learning - machine learning has been shown to be effective with large data sets - Dr. Sweet will explore whether these methods can be used with education data and if so, the project will train education researchers on how and when to use the methods;
3. Causal inference - use machine learning to inform covariate selection for propensity score matching;
4. Recommendation - develop guidelines for people to use to understand what methods to use and when to use the more innovative methods - and when should they stick to traditional methods.

Dr. Sweet went on to discuss the benefits that this project will have for Maryland. Specifically, it will improve the methodological capabilities of MLDS researchers. She intends to provide presentations and training on the use of the different data science tools being studied and help the researchers understand when to use different methodologies. In addition, Dr. Sweet explained that the researchers will employ the methods to address complex research questions. Each question will be submitted separately to the MLDSC for review and approval.

In response to a question from Dr. Fielder, Dr. Sweet explained that the project will not be working to develop new methodologies, instead it will focus on testing methodologies that are currently out there. In response to a question from Mr. Rizzi, Dr. Sweet stated that the use of the new data science tools are for retrospective analysis and not to create prospective predictive models.

2019 State Longitudinal Data System Grant

Dr. Henneberger began by thanking the Board for their support for grant applications. This research proposal is being submitted to the National Center for Education Statistics at the U.S. Department of Education. MSDE is the lead applicant and decided to focus the application on the *Equity Priority*. If funded, there will be a sub-award of approximately $1 million to the MLDS Center Research Branch. (MLDSC, UMB, and UMCP). There are three projects proposed by the Research Branch.

Project 1 - Conduct research to help inform policy, programs, and practices on educational equity as applied to the allocation of resources to Maryland students. To achieve the goal, the researchers will host two policy academies in the first and final year of the project. The first convening will include education stakeholders and will be roughly organized around the Kirwan Commission objectives (early childhood education, high-quality teachers, college and career readiness pathways, and resources for at-risk students). The goal of the meeting will be to have stakeholders ensure researchers are answering the right questions to inform policy down the road. Then during the second and third year of the project, researchers will conduct the research identified in year one. In the final year, another academy will be assembled to help disseminate the results.

The research questions will be developed collaboratively, but may include: (1) What was the impact of additional funds for schools with high concentrations of students living in poverty? and (2) What is the effect of diverse teachers and diverse peer groups on long term student outcomes?

Project 2 - Exploration of technical enhancement to support equity-oriented research. This will include exploring two initiatives: linking data from local school systems to extend the panel of data to examine the full transition from early elementary school into college and the workforce; and participation in interstate data exchange. The outcomes for Project 2 are two annual meetings to discuss the potential for an interstate workforce data exchange and meetings with local school systems to determine whether collaboration is mutually beneficial (and if so, to discuss necessary data-sharing agreements, collaborative research agendas, and an exploration of the reliability of the historical data).

In response to questions, Dr. Henneberger explained that the interstate data exchange would ideally be designed to allow the MLDS to link Maryland student records to out-of-state wage records.

Project 3 - Increasing equitable research access to state longitudinal data through piloting the use of synthetic data sets. Dr. Henneberger noted that graduate students represent an under-utilized resource for expanding the Center's research capacity. However, MLDS also needs to balance access and security. So the goal is to explore the use of synthetic data as a solution. The outcome for this project is to train students from historically under-represented backgrounds (i.e. women, students of color, and students in under-represented disciplines). Dr. Henneberger said that Morgan State University has expressed interest in working with MLDS data. The project would develop procedures for access to SDP; develop a course syllabus; and train 40 graduate students in advanced statistical analyses for policy-oriented research. At the end of the class, students would provide poster presentations of their research and findings would be made available on the Center's website.

Ms. Bjarekull asked whether it is premature to approve the third proposal since the synthetic data is still being tested. Dr. Henneberger said that regardless of whether the synthetic data is approved for wide release, it could still be a teaching tool. It would simply limit the number of people accessing the actual data.

Dr. Fielder asked why only MSU was recruited to participate in this program and suggested that participation should be broader. Dr. Henneberger agreed that further expansion and inclusion of students from other institutions are necessary and part of the goal. MSU was a good place to start as they had expressed interest in this type of collaboration.

In light of the plan to give 40 additional students access to the synthetic data, Ms. Bjarekull asked how many people currently have access to the synthetic data? Mr. Goldstein provided a listing of all of the staff that have access to the MLDS. The list includes members of the Research Branch, Synthetic Data Project researchers, MLDS Center staff, and external researchers and each individuals' level of access. Ms. Bjarekull noted that there are currently only eight individuals with access to synthetic data. Due to the significant increase in synthetic data access and the fact that testing is still ongoing, Ms. Bjarekull stated that she did not think that this part of the project should be approved.

Mr. Rizzi asked who would have control over Maryland data if an interstate data sharing agreement is created. Mr. Goldstein responded that the Center cannot currently participate in an interstate data sharing agreement since State law strictly prohibits the Center from redisclosing unit record data. The proposal is only about exploration - the Center wants a seat at the table while these discussions are going on to learn about the process and determine whether it is something worth pursuing.

In response to the issues raised by Ms. Bjarekull, Dr. Henneberger suggested a modification to the third project as follows: Increase equitable access to state longitudinal data through an exploration of coursework. In this revised project, students would get access to real data to do their research project and would be no different than what students working with the Research Branch do now. Ms. O'Croinin clarified that the project focus is increasing equitable access to the MLDS for students - which data they will use will depend on the result of synthetic data testing. In response to Ms. Bjarekull, Dr. Henneberger clarified that the 40 students would be added over four years, not all at once and not for the entire project duration. Additional faculty may be needed to supervise the students.

Ms. Kellogg pointed out that a benefit of the project would be developing strategies for the Center to engage with students in a manner that supports their academic needs while ensuring products for the Center.

In response to a question from Dr. Fielder, Dr. Henneberger stated that students would pay for the course and the MLDS Center would pay for part of the professor's time.

Finally, Mr. Muntz summarized the changes to project three, as removing "piloting the use of synthetic data sets" and adding "providing secure access to MLDS data." Based on the upcoming results of beta testing on research utility and disclosure risk, the project may in fact use synthetic data as the secure access method.

The Board voted on each project of the SLDS grant proposal separately as follows:

- Mr. Harrison made a motion to approve Project 1 - conduct research to help inform policy, programs and practices on educational equity. Ms. Bjarekull seconded the motion which was unanimously approved.
- Dr. Phillips made a motion to approve Project 2 - explore technical enhancement to support equity-oriented research. The motion was seconded by Mr. Biggs and unanimously approved.

- Dr. Moazzami made a motion to approve Project 3 - increase equitable research access to state longitudinal data through secure access to the MLDS. Ms. Bjarekull seconded the motion which was unanimously approved, with Dr. Fielder abstaining.

Dr. Fielder explained his abstention by noting that he wanted to further study the implications of the Center expanding its authority into teaching. Dr. Henneberger clarified that the proposal essentially creates a research apprenticeship course, under faculty supervision, where students will learn how to build a research question, operationalize that question, and apply the most appropriate statistical analysis around policy-relevant research for the MLDS. The goal is to expand the Center's research capacity by providing these for-credit courses instead of paying graduate students. Dr. Fielder reiterated that the project will be paying faculty to provide these for-credit courses and that he wants to study this expansion of the Center's role.

**New Business**

Mr. Goldstein stated that the new business is a review of the *MLDS Authorized Staff Access List*. The list was provided and reviewed during the prior discussion on the SLDS Grant proposal. Mr. Goldstein noted that the list will be provided in all future meeting folders so the Board can stay apprised of who has been granted access to the system.

**Closed Session**

Ms. O'Croinin stated that pursuant to General Provision Article § 3-305(b) the body will move into a closed session to discuss the appointment, employment, assignment, promotion, discipline, demotion, compensation, removal, resignation, or performance evaluation of possible future appointees or current employees, and any matters related to and stemming from the discussion that is permissible under Md. Code Ann., Gen. Prov. Art. § 3-305(b)(1), which must take place in closed session to allow the Governing Board to discuss individual compensation and protect the privacy of the confidential, personal information of the individual discussed, and/or to prevent any harm to the reputation of any individual discussed; (b) An information technology security briefing and discussion of the MLDS System Hosting Analysis Assessment prepared by Securance Consulting. This discussion will include information which, if made public, would jeopardize the cybersecurity of the agency. See MD. Code.Ann., Gen. Prov. Art. § 3-305(b)(15); and (c) Legal advice regarding the matters referenced in subsections (a) and (b) above. Ann., Gen. Prov. Art. § 3-305(b)(7). Mr. Rizzi made a motion to close the meeting, which was seconded by Mr. Harrison. The motion was unanimously approved.

Summary

The Closed session began with the following individuals in attendance: members of the Governing Board (see the list on page 1), Ross Goldstein, Tejal Cherry, Dawn O'Croinin, and Paul Ashe from Securance Consulting. Mr. Goldstein provided a summary for why the consultant was hired and the objective of the analysis. Mr. Ashe then provided an overview of the written report he and his firm provided for the Center. The report considered different database platforms and hosting options for the Center. Mr. Ashe provided the cost of each option, the relative operational and security strengths and weaknesses, the report's recommendation, and the rationale for the recommendation.

Mr. Goldstein, Ms. Cherry, and Mr. Ashe left the meeting for the Board to discuss personnel matters. The Board considered a request for a salary increase for Mr. Goldstein. Ms. Bjarekull

made a motion to approve the request, which was seconded by Dr. Phillips. The motion was unanimously approved.

**Closing**

The open meeting reconvened at 12:45 p.m. Mr. Biggs made a motion to adjourn the meeting, which was seconded by Mr. Rizzi.

Respectfully submitted,
Ross Goldstein
Executive Director

Approved: [pending]