MLDS Center
Maryland Longitudinal Data System

Better Data • Informed Choices • Improved Results

December 2017

2017 Annual Report on the Maryland Longitudinal Data System and Center

Submitted by:
Governing Board, Maryland Longitudinal Data System

Prepared by:
Ross Goldstein, Executive Director

Maryland Longitudinal Data System Center
525 West Redwood Street
Baltimore, MD 2101
410-706-2087
Mlds.center@maryland.gov
http://mldscenter.maryland.gov

Larry Hogan
Governor

Governing Board

Dr. James D. Fielder, Jr.
Secretary of Higher Education
Chair of MLDS Governing Board

Ms. Kelly M. Schulz
Secretary of the Department of Labor, Licensing and Regulation

Dr. Robert Caret
Chancellor of the University System of Maryland

Dr. Karen Salmon
State Superintendent of Schools

Dr. David Wilson
President of Morgan State University

Ms. Tina Bjarekull
President, Maryland Independent Colleges and Universities Association

Dr. Bernard Sandusky
Executive Director Maryland Association of Community Colleges

Dr. Jack Smith
Superintendent, Montgomery County Public Schools

Dr. Scott Tingle
Assistant Principal, Snow Hill High School

Mr. Steven Rizzi
Vice President, PAR Government

Ms. A.J. Brooks
Privacy Analyst, U.S. Department of Health and Human Services

Mr. Christopher J. Biggs
Information Assurance Manager, Raytheon Company

Ross Goldstein
Executive Director

This Annual Report of the Governing Board of the Maryland Longitudinal Data System Center is submitted to the Governor and the Maryland General Assembly in compliance with Education Article § 24-705 of the Annotated Code of Maryland.

Table of Contents

Highlights 3

Introduction 4

Section 1. Implementation of the MLDS 4

1.1 System Management 4

A. DoIT Enterprise System 4
B. Security 4
C. Records Count 5
D. Identity Resolution 6

1.2 Activities of the Center 7

A. Staffing 7
B. Stakeholder Engagement 7
C. Data Requests 9
D. Legislation 10
E. Development of Communications Plan 11
F. State Workforce and Education Alignment Project (SWEAP) 11
G. Grants 12
H. Research Series 13
I. Predefined Population Groups 14
J. Development of External Researcher Protocol 14
K. STATS-DC Presentations 15
L. Synthetic Data Project (SDP) 16

Section 2. List of all studies performed by the Center 17

2.1 Policy Report 17

2.2 Dashboards 17

2.3 Dual Enrollment Report 17

2.4 Maryland Public School Student Pathways Results 18

2.5 More Jobs for Marylanders Analysis 18

2.6 Planned Research and Reporting 18

Section 3. Data Determined to be Unnecessary 20

Section 4. Proposed or Planned Expansion of Data 21

1.1 Data Inventory 21

1.2 Additions to the Data Inventory 21

Section 5. Recommendations to the Governing Board 22

Attachment A – Maryland Public Schools Pathway Analysis 23

Attachment B – Data Requests for 2017 25

Attachment C. SLDS Issue Brief. 27

Highlights

The Maryland Longitudinal Data System (MLDS) Center is pleased to report on the significant progress it has made over the past year to refine and improve the System and increase output.

- Provided actionable data for key state policy makers through various collaborative engagements, such as:
  - Supporting the State Department of Education's required report under the *Teacher Induction and Retention Act*;
  - Providing data and analysis to the Higher Education Commission for the *Student Outcome Achievement Report*;
  - Supporting the work of University System of Maryland's Maryland Center for Computer Education on various analyses of computer science education trends in Maryland public high schools and postsecondary outcomes;
  - Providing a student pathway analysis for the Department of Legislative Services; and
  - Working with the Governor's Workforce Development Board on reporting requirements in the *More Jobs for Marylanders Act*.

- Responded to 34 data requests – a threefold increase from the previous year.

- Improved the cross-sector match rate by an additional 3% over the previous year. The Center is able to match 95% of all Maryland public high school 12th graders to a higher education record, a workforce record, or both.

- Became the only state agency to add an additional layer of security by proactively implementing an intrusion detection system (IDS). The IDS provides 24/7 monitoring services to protect the Center's most sensitive personally identifiable student and worker data.

- Several Center projects and initiatives received important national recognition this year. Two proposals were selected for presentation at the National Center for Education Statistics annual STATS-DC conference and the Center's Synthetic Data Project was the subject of an Institute of Education Sciences Issue Brief.

Introduction

This Annual Report is submitted in fulfillment of the requirement under Education Article, §24-705, Annotated Code of Maryland. The Governing Board must provide information to the Governor and General Assembly annually on the following:

1. An update on the implementation of the MLDS and activities of the MLDS Center;
2. List of all studies performed by the Center during the reporting period;
3. List of all currently warehoused data that are determined to be no longer necessary to carry out the mission of the Center;
4. Any proposed or planned expansion of data maintained in the database; and
5. Any other recommendation made by the Governing Board.

The following sections of the report will address of each of the five statutorily required topics.

Section 1. Implementation of the MLDS

1.1 System Management

A. DoIT Enterprise System

As reported last year, the Center has been working with the Department of Information Technology (DoIT) on migrating to the enterprise system. This year, the Center continued to work with DoIT on implementing additional components of the enterprise system, including hard drive encryption, migrating to DoIT's virtual private network solution, and a labor-intensive project to convert all Center system IP addresses to be compliant with the statewide government intranet, which is part of networkMaryland™ - the private network connection among state government entities and partners.

The final piece of the migration to the enterprise system requires physically moving the Center's data system from its current environment at the MSDE data center to the DoIT managed data center. Center staff is still working to determine whether this migration is in the Center's best interest. Outstanding issues include whether the enterprise system meets the same security standards established by the Center, whether Center staff will have sufficient visibility into the operations and management of the enterprise system to fulfill their obligations to the Governing Board and partner agencies to securely maintain and manage the Center's highly confidential student and workforce data, and whether the costs of the enterprise system are within the Center's budget. The Governing Board has asked for an analysis and recommendation from the Center's chief information officer in early 2018 on how to proceed.

B. Security

Last year, the Center completed a voluntary, independent audit of the System. The audit contained two primary recommended actions. First, was to comprehensively train all Center staff on security awareness. To fulfill this recommendation, all staff were required to attend *Cyber Security Awareness Training*, a three-hour training session tailored for the Center staff by Cyber Engineering Service, Inc.

The training included an in-depth understanding of the type and nature of different threats, an understanding of how phishing attacks operate and how to protect against them, a review of how to protect credentials, an understanding of mobile services and security, the need to be vigilant about one's work environment, and general best practices to maintain a secure environment.

The second recommendation from the voluntary, independent audit was to implement an *intrusion detection system* (IDS). The IDS includes hardware, software, and monitoring services for the MLDSC network to protect the system server containing sensitive personally identifiable student and worker data. The IDS monitoring is both automated and routinely reviewed by security experts. The IDS reviews server access to determine whether it is from an expected source at an expected time or whether it may represent an unauthorized intrusion or threat. The information gathered is assessed and reported to the Center weekly or sooner if necessary. The IDS cost is over $40,000 annually and represents a significant expenditure for Center.

The MLDS Governing Board approved the IDS implementation at the June 9, 2017 Governing Board Meeting. At that meeting, Charles Ames, State Director of Cyber Security for DoIT, commended the MLDS Center staff for putting together a strong security program. He noted that continuous monitoring, like the IDS, is a goal for any system. Mr. Ames stated that the first step in security is to create a strong defense, which he noted the Center has done. The next step is to have a more proactive stance, which the IDS implementation provides. Mr. Ames notes that the cost of the IDS is about a fifth of the cost of an equivalent full-time employee, which is reasonable to protect the Center's valuable data. Finally, Mr. Ames also stated that the Center's implementation of the IDS will hopefully plant a seed that will expand this type of protection throughout the State.

The IDS was fully implemented in September 2017 and the Center has been receiving and reviewing weekly reports. Updates on the implementation and output from the IDS have been presented at all subsequent Governing Board meetings.

C. Records Count

The following table shows the summary individual person counts from the data loaded into the System from the partner agencies.

Table 1: Number of individual records in the MLDS by Data Source

Data Source	Count as of6/29/2016	Count as of 11/16/2017	Percent Change
MSDE	1,673,949	1,980,714	18%
MHEC	1,203,673	1,389,867	15%
DLLR	1,074,724	1,381,175	29%
Net Total	**2,559,477**	**3,029,122**	**18%**

D. Identity Resolution

The Center's ability to match K-12 student data to higher education and workforce data remains at a high level as demonstrated in the following analysis. The Center assesses the match rate by determining the number of Maryland public school 12th graders for whom one or more cross-sector matches can be found. This year, the overall match rate has increased to 95%, a three percent increase from last year.

Table 2: Counts of Cross-Sector Linkages for 12th Grade Cohorts in the MLDS by Total and Cohort Year

Maryland Public School 12th Graders	Total for all 12th Grade Cohorts (8)	12th Grade Cohort Year								
		2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016
Total count for 12th grade cohorts	577,427	66,326	64,406	64,653	64,531	64,022	63,682	62,820	63,534	63,453
K12 only (no links to other sectors)	29,412	6,998	4,344	3,830	3,983	3,884	4,261	4,568	5,894	10,805
K12 and College (no Workforce)	47,347	7,494	4,423	4,078	4,047	4,271	4,749	5,638	6,770	5,877
K12 and Workforce (no College)	133,142	11,203	12,493	13,257	13,978	14,080	14,652	14,333	15,589	23,557
K12, College, and Workforce	348,371	40,631	43,146	43,488	42,523	41,787	40,020	38,281	35,281	23,214
Percent matched across all 3 sectors	**60%**	**61%**	**67%**	**67%**	**66%**	**65%**	**63%**	**61%**	**56%**	**37%**

Table 3: Percent of 12th Grade Cohorts in the MLDS with one or more cross-sector matches

Academic Year	Total for all 12th Grade Cohorts (8)	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016
2017	95%	89%	93%	94%	94%	94%	93%	93%	91%	83%
2016	92%	87%	92%	93%	93%	93%	92%	90%	85%	N/A
2015	88%	87%	91%	91%	90%	89%	87%	82%	88%	N/A
Difference	**↑ 3%**	**↑ 2%**	**↑ 1%**	**↑ 1%**	**↑ 1%**	**↑ 1%**	**↑ 1%**	**↑ 3%**	**↑ 6%**	**N/A**

1.2 Activities of the Center

A. Staffing

The Center filled its last two vacant positions and is now fully staffed. The two positions were a data analyst and business analyst. The data analyst added additional capacity for data loading, developing, and analyzing data sets for reports, data requests, and dashboards. The business analyst develops formalized procedures and documentation on various Center operations, a necessary component, as the Center and System continues to mature.

B. Stakeholder Engagement

Over the course of the past year the Center has made a concerted effort to collaborate with stakeholders. This collaboration has been most effective when stakeholders approach the Center with a policy question. The Center collaborates with the stakeholder to provide actionable and relevant information that informs the stakeholder's decision.

1. Maryland Higher Education Commission

Replication of the MHEC Graduate Follow-Up Survey Reports

Every three years, the MHEC conducts a follow-up survey of postsecondary graduates after they have completed their program. These surveys are a valuable tool that help the State and institutions better understand student outcomes. There was an interest in determining if the data available in the System could enhance the existing report, including providing more robust employment data. . Working with the MHEC Data Advisory Group, staff of the Center's Research Branch have begun replicating the Follow-Up Survey Reports conducted by MHEC. Researchers will share their findings and conclusions with MHEC later this year.

SOAR Report

The Center is collaborating with MHEC to produce the Student Outcome and Achievement Studies report (SOAR). SOAR is a periodic report produced by MHEC in response to the General Assembly's 1988 charge to improve information to high schools and local school systems concerning the performance of their graduates in college. The first data table required for SOAR has been created and validated.

2. Maryland State Department of Education

CTE Federal Reporting

The MLDS Center is assisting MSDE, Division of Career and Technical Education, Student and Assessment Services Branch by providing data for federal reporting requirements on Career and Technical Education (CTE). This is the second year the MLDS will be used to meet this federal reporting requirement. MSDE is required to submit two aggregate data files to the U.S Department of Education, EDFacts. Both of these federal reporting requirements relate to the outcomes of CTE students after they leave high school, which also aligns to the MLDS Research Agenda. [1]

[1] https://mldscenter.maryland.gov/ResearchAgenda.html

Teacher Induction

Under the Teacher Induction and Retention Act (Chapter 740, Laws of 2016), MSDE was required to convene a workgroup and provide a report that determines how to incorporate various strategies to recruit, retain, and promote quality teachers. MSDE research staff requested the Center provide information to help them assess whether Maryland high school students are entering teaching programs in college, whether students are diverted from teaching and entering other professions instead, and whether graduates of teaching programs are teaching in Maryland. MSDE reported that they were able to glean policy relevant information from the Center data to support the work of the workgroup and its analyses.

3. Governor's Workforce Development Board

The Center has been working closely with the Governor's Workforce Development Board (GWDB) and the Governor's P20 Leadership Council on compiling data to support the reporting requirements of the More Jobs for Marylanders Act. The Center presented preliminary information at the September Governor's P20 Leadership Council Meeting and has subsequently met with the GWDB to complete the report. Additional information on the final report can be found in Section 2.5 More Jobs for Marylanders.

4. Department of Legislative Services

In August, Center staff met with education policy and budget analysts at the Department of Legislative Services (DLS) to discuss their upcoming priorities and areas where MLDS could provide information. The following two projects are a result of that meeting.

Maryland Public School Student Pathways Results

DLS staff requested a high-level overview of the outcomes of the high school class of 2008. Using data in the System, pathways were mapped for 12th grade high school exiters who graduated with a high school diploma, entered college, and either persisted, graduated, transferred, or were no longer enrolled within a specified amount of time. The Student Pathways project will be further developed to include additional subgroup analysis and median earnings.

Teacher Preparation Study

DLS staff in their work to support the Commission on Innovation & Excellence in Education, requested information from the Center on Maryland's teacher preparation programs. The Center's response to this request was composed of two parts. First, the Center provided information related the Center's novice teacher dashboards. Second, the Center convened a group of experts from MSDE, MHEC, USM, and MICUA to help define the scope of an analysis and select outcome indicators for a thorough analysis of the performance of teacher preparation programs in Maryland. The MLDSC Research Branch is currently working on this analysis.

5. University System of Maryland

Maryland Center for Computing Education

In 2017, USM established the Maryland Center for Computing Education (MCCE) to expand access to high-quality computing education for all Maryland public school students (Pre-kindergarten through

college). Earlier this year, Dr. Marie desJardins, computer science professor at UMBC and a key contributor to MCCE, requested that the Center support research efforts on a variety of topics involving the computer science education pipeline. Specifically, the Center developed a series of data sets that explored the relationship between taking computing and math courses in high school and college major selection. The Center plans to continue collaborating with MCCE by providing data on the college training of computer science teachers and the workforce outcomes of college students who graduated with degrees in computer science programs.

Health Care Workforce Writing Group

USM has convened a writing group to analyze health care workforce disparities, issues of clinical placement needs across various health occupations, and whether P-20 partnerships are having an impact on the number of students enrolling in health science programs. The Center has been asked to serve as a resource to the writing group and provide needed data analysis to support the work.

C. Data Requests

Regulations

The Governing Board adopted proposed amendments to its regulations governing the Center's implementation of the Public Information Act. Before the amendments, the regulations stated that a longitudinal data request under COMAR 14.36.04 are separate from, and are not subject to, PIA. The rationale for that conclusion was based on the language in the PIA which does not obligate an agency to create a record in response to a request if the record does not already exist. In the case of a longitudinal data request, the requested aggregate data set does not exist prior to the request and therefore requires the Center to compile the requested data. The issue was whether this compilation constituted "creating a record" for purposes of PIA.

According to Dawn O'Croinin, Assistant Attorney General, and counsel for the Governing Board and the Center, compiling an electronic record does not constitute creating a new public record. Ms. O'Croinin's view is supported by state law, a 2014 opinion of the Maryland Court of Appeals, and the recently revised *Maryland Public Information Act Manual (14th Ed. – October 2015).* The Manual states that an agency should comply with a request if it has staff available who routinely perform the type of data extraction requested, but need not do so if it would call for expertise outside the agency's existing capabilities. The Manual also clarifies that the agency is not required to comply with requests that call for it to generate new data or analyze or summarize data.

Since the Center's staff and System have the capability to extract aggregate data sets (which does not require analysis), the regulations were amended to clarify that data requests fall under PIA and must be processed accordingly.

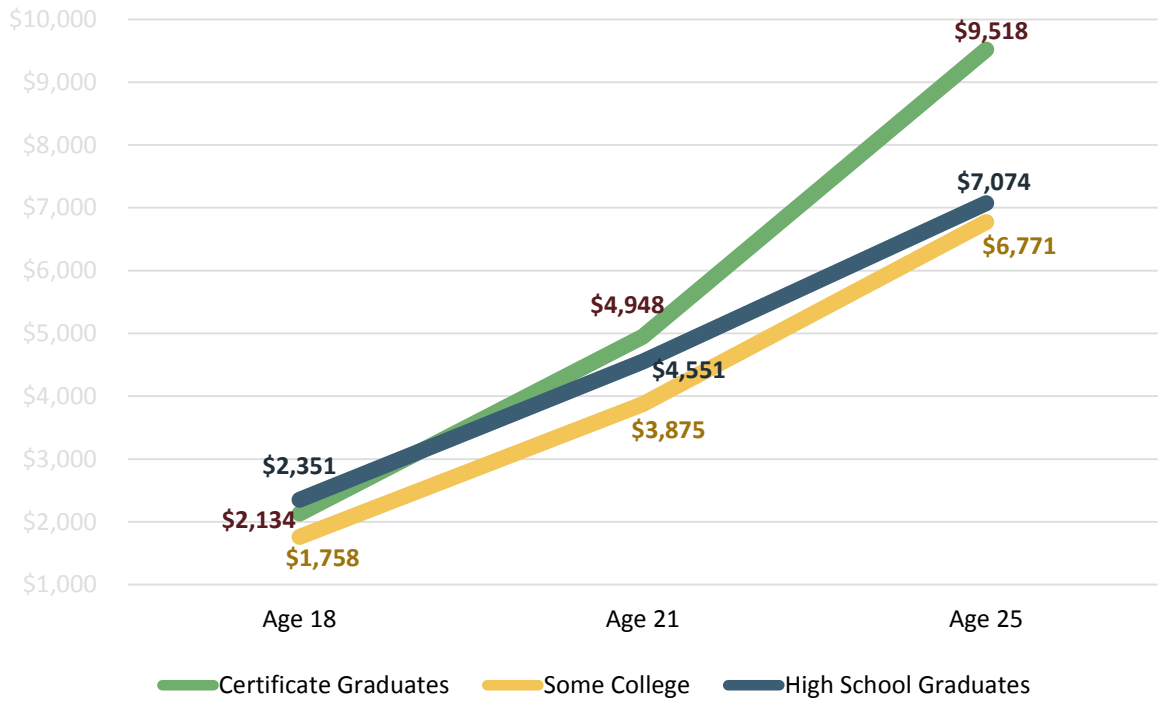
Requests

This past year, the number of requests that the Center fulfilled tripled from 11 in 2016 to 34 in 2017. Attachment B contains a complete list of each of the data requests.

D. Legislation

Last year, the General Assembly passed HB 680/SB 1165 - *Maryland Longitudinal Data System - Student and Workforce Data Linkage - Extension of Time Limit* (Chapters 790, 791 of 2017). The new law changes, from five-years to 20-years, the amount of time that a student record can remain linked to a workforce record following the student's last enrollment in a Maryland education institution. The change has been fully implemented in the MLDS. The positive impact of the change was immediate in that it allowed the Center to conduct the analysis required under the *More Jobs for Marylanders Act* (Chapter 148 of 2017). The Act required an analysis of the workforce outcomes for non-college graduates at age 25. Since a substantial proportion of that population never attended college[2], their last enrollment in a Maryland education institution was high school, which is more than five years prior to turning 25. Accordingly, without the legislative change, the Center could not analyze the workforce outcomes of a substantial number high school graduates at age 25. Further, as the chart from the analysis demonstrates (below), more years of workforce data better illuminate the workforce wage trajectory following a student's formal education.

Chart 1. Median Quarterly Wage Trajectory from Ages 18 to 25 for Individuals with Continuous Wage Visibility by Educational Attainment



[2] Roughly a fifth (22%) of Maryland high school graduates between 2008 and 2011 have no record of any college attendance.

E. Development of Communications Plan

One of the Governing Board's recommendations in last year's Annual Report was to, "develop and implement a communications plan that creates awareness of MLDS output and information, establishes a better understanding of how longitudinal data can be used to help drive decision making, and generally provides a systematic approach to engaging stakeholders."

In response to the recommendation, Center staff created its *Communications Plan*. The plan is organized around a multi-phased approach to improving communications. The first phases assessed current communications and needed improvements, identified the Center's audiences, and reviewed the resources available for communications and outreach. Part of the initial assessment also reviewed and refined the Center's mission and vision and established clearly defined outcomes goals that seek to develop an audience that:

1. Uses the MLDS Center as the resource for longitudinal education and workforce data in Maryland;
2. Engages the MLDS for analysis of longitudinal student outcomes in Maryland;
3. Provides input on the direction for the continuation of MLDS Center output; and
4. Champions the role of the MLDS Center across the state of Maryland.

To help facilitate communications and implement the Communications Plan, a Communications Advisory Board (CAB) was established and held its first meeting in May. The meeting included policy and communications staff from the Maryland Higher Education Commission, Maryland Department of Labor, Licensing and Regulation, the Maryland State Department of Education, University of Maryland, School of Social Work and the University System of Maryland.

The Communications Advisory Board encouraged the Center to focus its efforts on the first tier of its targeted audience: the General Assembly and executive branch agencies. Specifically, the CAB recommended working closely with these key stakeholders to deliver content and research that will directly assist in policy decisions. The CAB suggested that output and demonstration of direct value would form a basis for communications to broader audiences. As discussed above in the Stakeholder Engagement section, the staff has followed the recommendation by finding opportunities to work with stakeholders and provide information to help inform their work.

F. State Workforce and Education Alignment Project (SWEAP)

The Center, in collaboration with the Workforce Data Quality Campaign[3] (WDQC), held a meeting in April 2017 to learn about and discuss the SWEAP data tools, an initiative of the National Skills Coalition[4]. Those tools include:

1. Dashboards that use a small number of common metrics to report on education and employment outcomes across workforce development programs;

[3] More information available at Workforce Data Quality Campaign website: http://www.workforcedqc.org/
[4] More information available at National Skills Coalition website: https://www.nationalskillscoalition.org/

2. Pathway evaluators that show the best pathways to gain skills for work in particular industries; and
3. Supply and demand reports that compare the number of trained workers in a state to the number of workers those employers need in order to help align training with employer demand.

The WDQC hosted Domenico "Mimmo" Parisi and Zack Krampf, of the National Strategic Planning and Analysis Research Center (nSPARC) at Mississippi State University. Dr. Parisi and Mr. Krampf developed these tools for Mississippi and shared their process and the impact those tools have had.

Thirty-five people attended the meeting, including representatives from all three partner agencies, MICUA, MACC, researchers, local government, and local foundations. The meeting also included a discussion about how the tools could be used to further Maryland's policy priorities. The Center continues to work with its various stakeholders to explore the tools, determine which tool to pursue, whether the Center has the necessary data for the tool, and whether the tool aligns with the Center's Research Agenda.

G. Grants

Listed below are the grant opportunities for which the Center has issued letters of support and has committed to providing research access to System data to facilitate the analyses required under the grant projects.

1. *Growing and Sustaining CS for All in Maryland*
   The goal of this National Science Foundation (NSF) grant is to expand computer science (CS) education throughout the state as a critical component of education and workforce development. The grant is being sought by USM (lead) and MSDE. The Center's participation will focus on evaluating the effectiveness of state and federally funded programs to increase computer science education in Maryland.

2. *Minority Youth Violence Prevention II: Integrating Social Determinants of Health and Community Policing Approaches*
   The goal of this Department of Health and Human Services grant is to identify innovative approaches to significantly reduce the prevalence and impact of youth violence among racial and ethnic minority and disadvantaged at-risk youth. The grant is being sought by the University of Maryland, Baltimore (UMB) School of Social Work. The Center's participation will focus on evaluating the federally funded education program established by the grant.

3. *Exploring Links between Arts Education and Academic Outcomes among International Baccalaureate Students*
   This project proposes to use data from the MLDS to understand how postsecondary enrollment, persistence, and degree attainment vary among International Baccalaureate (IB) Diploma Program students who do and do not elect to include an arts education course in their IB course of studies. The grant is being sought by Dr. Kenneth Elpus, Professor of Music Education, University of Maryland, College Park, for his grant application to the Institute of Education

Sciences (IES) for a research project under the Education Research Grants program (CFDA 84.305A).

4. *Promise Neighborhoods Grant FY 2017*
   Last year, the Center Partnered with the University of Maryland, School of Social Work (SSW) on a proposal for a federal Promise Neighborhood Implementation Grant. SSW was not awarded a grant. However, a new round of funding is available and SSW, with support from the Center, has submitted a new proposal to the U.S. Department of Education, Office of Innovation and Improvement. If awarded, the Center will receive funding to examine the postsecondary and workforce outcomes associated with student participation in the Promise Heights programs.

H. Research Series

During the academic year, the Center holds a monthly Research Series open to the public that provides presentations on the Center's research or areas of interest. The presentation materials and additional information can be found on the MLDS website.[5]

1. *Dual Enrollment in Maryland: Highlights from the 2016 Dual Enrollment Report. (February)*
   Dr. Angela K. Henneberger, Ph.D., Research Director, MLDS Center used data from the Maryland Longitudinal Data System (MLDS to identify students who (1) had overlapping enrollment dates in a Maryland public high school and a Maryland college and (2) were enrolled in the college for at least 30 days. Findings on the rates and trends of dual enrollment, the courses in which students were dually enrolled, and the college enrollment outcomes of dually enrolled students were presented.

2. *Need-based Grant Aid in Higher Education: The Effects of the Howard P. Rawlings Educational Assistance (EA) Grant on Financial Aid, Academic Persistence, and Working while in School. (March)*
   Heath Witzen, Doctoral Student, University of Maryland, College Park, presented preliminary findings on the effects of the Howard P. Rawlings Educational Assistance (EA) Grant, the State of Maryland's largest need-based grant program, on several student outcomes including overall financial aid, earnings while in school, and persistence and degree attainment.

3. *Investigating the "Brain-Drain" Phenomenon in Maryland. (April)*
   Dr. Amber Bloomfield, University of Maryland, College Park, presented her research on whether Maryland students who attend out-of-state postsecondary institutions are less likely to return to Maryland seek employment in their original state of residence following graduation, and thus are less likely to contribute to the state's economy and workforce quality.

4. *Value Added Modeling and Alternate Approaches for Ranking Institutions. (May)*
   Dr. Laura M. Stapleton and Yating Zheng, University of Maryland, College Park presented their

[5] Research Series presentation materials available at: https://mldscenter.maryland.gov/ResearchSeries.html

work on the MLDS research agenda "which 4-year institutions are graduating students most effectively and in the timeliest fashion?" The presentation reviewed common approaches to this type of analysis, the shortcomings of those approaches and a proposal for an alternative approach.

5. *Applying Causal Inference Techniques to Strengthen Dual Enrollment Program Evaluation in Maryland. (October)*
   Dr. Angela K. Henneberger, Ph.D., Research Director, MLDS Center and Mr. Heath Witzen, Research Fellow presented on how modern causal inference techniques, such as propensity score methods, can be used to strengthen causal inferences in the absence of randomization in education sciences and how the approach to data from the MLDS to evaluate dual enrollment program participation in high school. Outcomes examined will include college enrollment and persistence.

6. *Validating Teacher Effects on Students' Attitudes and Behaviors: Evidence from Random Assignment of Teachers to Students (November)*
   The MLDS Research Branch invited Dr. David Blazar, an Assistant Professor of Education Policy and Economics at the University of Maryland College Park, to discuss his research on teacher quality, with a focus on professional learning, organizational contexts of schools and districts, and accountability policy.

I. Predefined Population Groups

Currently, the Center defines unique population groups for each dashboard as part of the dashboard development process. Rather than develop unique populations that can only be used for a single dashboard, the Center is working towards developing predefined population groups that align to key sector transitions that can be repeatedly queried for dashboard development. A preliminary flowchart has been developed to capture the major population groups and subgroups. The predefined population group development has been shared with and reviewed by the Center's Research and Policy Advisory Board. The development of the population groups takes time away from developing new dashboards. However, the expectation is that the investment in this development effort will ultimately streamline and standardize the dashboard development process and allow the Center to rapidly respond to data requests and other inquiries.

J. Development of External Researcher Protocol

Direct access to data in the Maryland Longitudinal Data System is restricted to authorized staff of the Center (Md. Code, Ed. Art. 24-703(g)). Without staff access, a researcher is limited to only de-identified, aggregate data (Md. Code, Ed. Art., §24-703(f)(4) and (5)). To adequately conduct research specified in state statute and outlined in the MLDS Research Agenda, researchers need access to unit-record level data.

To support the research directives and the mission of the MLDS Center, the Executive Director may designate researchers as the authorized staff of the Center. Under current regulations, there are two ways researchers can have access to the MLDS:

1. *Member of the Research Branch (COMAR 14.36.06.01C2a)* - A faculty member within the University System of Maryland (USM), may be assigned to the Research Branch of the Center through the MOU with the University of Maryland, School of Social Work and College of Education. This includes the Principal Investigator, Research Director, Statistician, and Associate Research Director as well as other researchers from within USM who work under the auspices of the Research Branch.
2. *Temporary Researcher (COMAR 14.36.06.01C2b)* - Additional researchers identified by the Director of the Research Branch may become authorized staff and receive research access to the MLDS. In 2015, the Governing Board considered procedures that would limit the number of temporary research appointments to two and require a referral from a member of the Governing Board. To date, one temporary researcher appointment has been granted (for Baltimore's Promise).

Over the past year, there have been more researchers interested in having access to the MLDS; either as a member of the Research Branch or as a temporary researcher. Accordingly, the Center needs a clearly articulated process for selecting qualified researchers who are proposing work on topics that fit within the Research Agenda and have the potential to provide meaningful policy insights. A documented process will allow the Center to leverage its collaborative relationship with the research community throughout Maryland and better utilize the System's rich store of data to inform stakeholders and policymakers, in a way that is controlled, clear, and transparent.

Center staff reviewed other state policies and procedures to develop a draft document. Input was solicited from the MLDS Advisory Groups and other stakeholders to ensure the procedures are sufficiently comprehensive and protective of the state's interests while still providing meaningful opportunities for rigorous academic research on the MLDS. A final draft will be presented to the MLDS Governing Board at the December 2017 meeting.

K. STATS-DC Presentations

The National Center for Education Statistics (NCES), in the U.S. Department of Education's Institute of Education Sciences (IES), held their annual NCES STATS-DC Data Conference August 1st through August 3rd. The conference provides important updates on federal and national activities affecting data collection and reporting, and information about the best new approaches in collecting, reporting, and using education statistics. This year, MLDS Center staff provided two presentations during the conference.

1. Laia Tiderman and Ann Kellogg - *Which Came First? The Dashboard or the Population?* Abstract - Maryland Longitudinal Data System Center staff found that the current dashboard production process was cumbersome and slow. Center staff began rethinking the development process and asked, "What if we start from the population and not the question?" The Center is piloting an alternative process that defines key population groups and that can be used repeatedly to address multiple research questions. This approach requires the Center to work with stakeholders to define criteria for "key populations" as well as redesign its dashboard

processes. This session discussed the Center's experience with this alternative process and the criteria it used for defining key population groups.

2. Angela Henneberger - *Dual Enrollment in Maryland: Using Propensity Scores to Strengthen Program Evaluation with State Longitudinal Data.* Abstract - The Center is legislatively required to submit an annual report on dual-enrollment program participation (simultaneous enrollment in both a public high school and college). Programs such as dual enrollment are often difficult to evaluate causally due to the absence of randomization. Modern causal inference techniques, such as propensity score methods, can be used to strengthen causal inferences in the absence of randomization in education sciences. The presenters applied this approach to program evaluation using data from the MLDS and discussed the strengths and limitations of using this approach with state longitudinal data systems.

L. Synthetic Data Project (SDP)

Project Overview

The Center continues its work on the development of synthetic data, a project being funded by a grant from the U.S. Department of Education, Institute of Education Science.[6] The Center's project, which is being conducted by the Research Branch, involves evaluating the feasibility of developing synthetic data from the Center's actual data. Synthetic data is essentially a set (or multiple sets) of hypothetical students. If successful the hypothetical data set would allow a researcher to test a research question about actual students against the hypothetical students (i.e. synthetic data). Because the synthetic data does not represent actual students, broader access can be granted to allow more research while ensuring student privacy is not compromised.

This past year, the SDP researchers have been working to create the data sets to be synthesized. To determine what data to include in the data sets, an end user panel of 15 education and workforce researchers was convened to get input on likely research inquiries and needed data. Creating the data sets required careful analysis of the data and has led to discovery and resolution of specific data quality issues.

SDP researchers are also convened a methodological expert consultants meeting in late October. The meeting provided an opportunity to show the team of experts the results of the synthetic data model development, and to receive feedback on how to proceed.

National Recognition

The Institute of Education Sciences (IES) Statewide Longitudinal Data Systems Grant Program produces a monthly publication - *SLDS Issue Brief. (See Attachment C)* The August edition is devoted to *Maryland's Synthetic Data Project*. The *Issue Brief* provides an overview of what synthetic data are, the goals of the Center's Synthetic Data project, and a discussion of the benefits of using synthetic data.

[6] The Center was one of several agencies to partner with MSDE in applying for the 2015 Statewide Longitudinal Data Systems Program Grant. In total, MSDE was awarded $6.9 million over four years, $2.6 million of which is for the Center's project.

Section 2. List of all studies performed by the Center

2.1 Policy Report

Workforce Outcomes in Maryland for Students Who Do Not Attend College: Patterns among Students Who Earn a High School Diploma, Certificate of Completion, Diploma via GED, and High School Non-Completers[7]

This report focused on the annual wages earned by the following categories of high school students who did not attend college: regular high school diploma earners, certificate of completion earners, high school diploma via GED earners (ages 16-21), and high school non-completers in the academic year 2009-2010. The report used three different methodologies for calculating annual wages. The report found that the GED earners had the highest wages (possibly because they are older and may have more work experience), followed by the high school diploma earners (who narrow the gap with the diploma via GED earners after five years). The non-completers had the next highest wages and the certificate of completion earners had the lowest wages. The report also found differences in the industries in which the different groups primarily worked. Finally, the report explored the policy implications of this analysis by providing in-depth background information on the financial and social burdens for individuals who do not complete high school and for those who complete high school but are not fully engaged in the workforce.

2.2 Dashboards[8]

The Center added 10 new dashboards to the website this year. The first three dashboards, under the category of college transitions to workforce, examine the workforce outcomes of novice teachers.

1. Bachelor's Degree Graduates Employed as Public School Teachers within 1 Year of Graduation
2. Time to Employment for Bachelor's Degree Graduates Employed as Public School Teachers
3. Maryland Novice Teachers who Attended Maryland Public Schools

The next seven dashboards provide information on Dual Enrollment Trends. The first set includes three dashboards on overall statewide dual enrollment trends and then disaggregates the trends by grade level and by demographic trends. The second set includes four dashboards on overall dual enrollment trends for each county and then disaggregates the trends by gender, race and ethnicity, and grade level.

2.3 Dual Enrollment Report

The Center is required under Education Article, § 24-703.1 of the Annotated Code of Maryland, to provide a report to the Governor and General Assembly on the number of students who are dually enrolled and the number and course name of the courses in which the students dually enroll. The Dual Enrollment Report has been submitted each year starting in 2013. The 2015 report was the first report written using cross sector data from the MLDS. The 2016 report was the first to include high school

[7] MLDS Center Policy Report: https://mldscenter.maryland.gov/egov/Publications/HS_to_Workforce_FinalMerged_May2017.pdf
[8] MLDS Center Dashboards: https://mldscenter.maryland.gov/Dashboards.html

course information. The 2017 takes advantage of the increased years of data to provide additional analyses on dual enrollment trends over time.

2.4 Maryland Public School Student Pathways Results

At the request of the staff of the Department of Legislative Service, the Center created an analysis of *Maryland Public School Pathways Results.* The analysis starts with a cohort of 2008 high school exiters and reports on their graduation from high, college enrollment patterns, and college persistence and graduation patterns at 150% of normal program length. For a 2-year program, 150% of normal program length is 3 years and for a 4-year program, 150% of normal program length is 6 years. The analysis is included in Attachment A.

2.5 More Jobs for Marylanders Analysis

The *More Jobs for Marylanders Act* (Chapter 148 of 2017), requires the MLDS Center and the Governor's Workforce Development Board to, "develop annual income earnings goals for high school graduates who have not earned at least a 2-year college degree by age 25."[9] The Center's role in establishing the initial annual income earnings goal was to provide an analysis of the current earnings of 25-year-olds who graduated from high school in 2008 through 2010 and did not earn a college degree. In addition to reporting on actual wages, the analysis also provides comparison points between the actual wage data and cost of living calculations[10] and reported median wages[11] for all workers.

2.6 Planned Research and Reporting

The following research projects that are currently being worked on by the Research Team for release later this year or next year.

Dual Enrollment in Maryland: What is the Effect on College and Workforce Outcomes?

The goal of this study is to examine the causal relationship between participation in dual enrollment programs and postsecondary outcomes using data from the MLDS. Propensity score matching was used to match students who participated in dual enrollment programs to similar students who did not participate in dual enrollment programs. The matched sample is then used to examine the relationship between dual enrollment program participation and college enrollment outcomes over time, college degree attainment, and earnings six years after high school. Results indicated that dual enrollment program participation had a significant positive causal effect on enrollment in a 2-year college, but not on enrollment in a 4-year college one year after high school. In years 2-4, dual enrollment program participation had a significant positive causal effect on enrollment in both a 2-year and 4-year college, suggesting that students who participate in dual enrollment are beginning their college careers at 2-year institutions and moving to 4-year institutions. Dual enrollment program participation also had a

[9] Corporations and Associations Article, § 21-204(b), Annotated Code of Maryland

[10] The Living Wage Calculator is a product of the Massachusetts Institute of Technology and is available online at http://livingwage.mit.edu/.

[11] The source for median wages is from the American Community Survey 5 Year Estimates, available online at http://planning.maryland.gov/msdc/American_Community_Survey/2011-2015/County/County.shtml

significant positive effect on degree attainment and wages six years after high school when compared to similar students who did not participate in dual enrollment programs.

The Relationship between School-level Concentrated Poverty and Individual Student Poverty

Senator Ferguson requested that the MLDS Center conduct a research study examining the effects of school-level concentrated poverty and individual student poverty on outcomes such as high school graduation, entry and persistence in post-secondary education, entry into the workforce, and wages earned. Specifically, the research questions include: How does the socio-economic status (SES) of a school's student population impact student outcomes? Do student outcomes change as the SES of the student population at a school changes? Using MLDS data the Research Branch plans to examine the unique associations between individual- level poverty and student outcomes and school-level poverty and student outcomes including high school graduation, post-secondary enrollment and persistence, and workforce wages earned in Maryland.

Pathways to Successful Teaching in Maryland

Department of Legislative Services staff in their work to support the Commission on Innovation & Excellence in Education asked the MLDS Center for research and analyses to help them understand the pathways to successful teaching in Maryland. Specific research questions included:

- Research question 1: Where do individuals who graduate from Maryland teacher preparation programs go on to teach?
- Research question 2: Are teachers who graduate from Maryland teacher preparation programs successful, particularly as measured by student outcomes?

The first research question asks about the teacher pipeline in Maryland. The teacher "pipeline" is the set of key component steps through which individuals must pass in order to become part of the teacher supply. Center staff are currently working on completing work on the pipeline. The second research question asks about teachers' success, particularly as measured by student outcomes. The first step in answering this research question is to examine the pipeline. The research analyses associated with research question 2 are planned to begin after the pipeline information (research question 1) is completed.

Section 3. Data Determined to be Unnecessary

Data elements identified for removal in the System are presented to the Governing Board for approval. The following data elements have been approved for removal from the data inventory in the 2017 calendar year:

1. *Correctional Education*

   The Governing Board approved the removal of the Correctional Education data provided by DLLR. This data collection was determined not to comply with Md. Code, Ed. Art. 24-701(f)(3)(ii) as it discloses information regarding a person's criminal record. Correctional education data includes educational, vocational, and training data pertaining to correctional education participants who are incarcerated in State Correctional Facilities captured in the DLLR/DWDAL database. The specific data elements include:

   - Address Line 1
   - Address Line 2
   - American Indian or Alaska Native
   - Asian
   - Award Date
   - Birthdate
   - Black or African American
   - City
   - Demographic Race Two or More Races
   - First Name
   - Gender
   - Generation Code or Suffix
   - Hispanic or Latino Ethnicity
   - Last Name
   - Middle Name
   - Native Hawaiian or Other Pacific Islander
   - Social Security Number (SSN)
   - State
   - Test Component Name (Subject/Module)
   - Test Component Score
   - Test Date
   - White
   - Zip code
   - Zip code +4

2. *Highly Qualified Teacher Indicator*

   The Governing Board approved the removal of the highly qualified teacher indicator provided by MSDE. Under the Elementary and Secondary Education Act of 1965 as amended by the Every Student Succeeds Act, MSDE no longer collects this data element. Beginning with the 2016-2017 academic year, this data element will no longer be provided from MSDE.

Section 4. Proposed or Planned Expansion of Data

1.1 Data Inventory

Md. Ed. Art. §24-701(f) defines the permissible types of student and workforce data that the MLDS may collect. Data that falls under that definition and are determined to be necessary to carry out the mission of the Center are presented to the Governing Board for approval to be included in the inventory. The Data Inventory represents the complete list of data that the MLDS Center collects.

The development of the Data Inventory is a collaborative process between DLLR, MSDE and MHEC, and the MLDS Data Advisory Group. The MLDS Data Advisory Group annually reviews the Data Inventory and identifies what data elements to propose for inclusion or remove from the System. All data must be relevant to answering one of the questions in the Research Agenda or generally identified as necessary for evaluating federal or State supported education programs. Data elements identified for inclusion or removal in the System are presented to the Governing Board for approval. Once approved, the requirements are developed with the agency and included in the data collection calendar.

The MLDS Data Inventory is available online through the MLDS Center website.[12] This change allows the inventory to be more dynamic, by presenting the inventory in a format that allows it to be organized according to the user's needs. A document listing the *Substantive Changes to the MLDS Data Inventory* has been created to catalogue the specific changes made each year.

1.2 Additions to the Data Inventory

There are eight new data elements collected from PK-12 Attendance that the Governing Board approved for inclusion in the MLDS. These data elements provide additional details on PK-12 students including program participation and funding information.

- Kindergarten Program Full Time Participation
- Evening High Program Participation Status
- Part-Time Student Status
- Number of College Courses
- Number of Part-time Courses
- Number of Evening High Courses
- State Aid Eligibility Code
- Number of Days in Membership

[12] https://mldscenter.maryland.gov/DataInventory.html

Section 5. Recommendations to the Governing Board

The Maryland Longitudinal Data System Governing Board recommends that Center output should include a section on any limitations in the analysis due to the lack of data collected and reported to the Center. Included in the discussion of limitations should be any necessary cautions about how the noted limitations may impact the interpretation of the output. Prior to each June meeting of the Governing Board, staff should analyze these limitations and make recommendations to the Governing Board on actions needed to address these data deficiencies.

Attachment A – Maryland Public Schools Pathway Analysis

Overview

In 2008, nearly 61,000 students[1] exited 12th grade from Maryland public high schools. Of those exiters, 25% graduated from college and an additional 8% of them remain enrolled (persisted) at 150% of normal time to completion (three years for a 2-year program and six years for a 4-year program).

High School Graduation

The majority, 96% of the exiting class, earned a high school diploma. Of those earning a diploma, 21% fulfilled the requirements for a Career Technology Education program.

College Enrollment Patterns

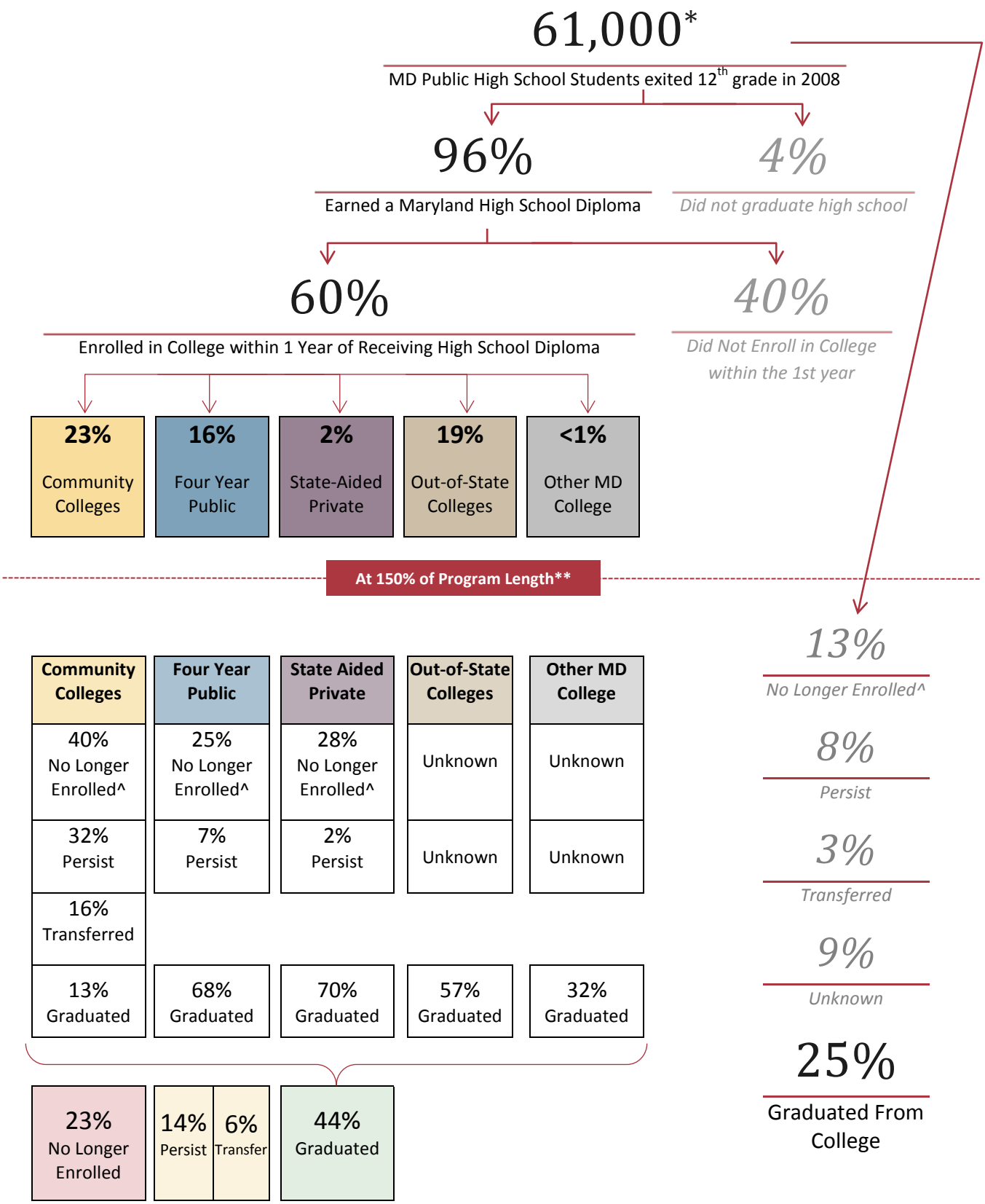
The majority of this 12th grade exiting class, enrolled in college within 1 year of exiting high school[2]. More students selected a Maryland college (41%) for their initial enrollment, rather than an out-of-state college (19%). The type of Maryland college selected for initial enrollment varied widely. Of the students that enrolled within Maryland, 23% selected a community college, 16% selected 4-year public college, 2% selected a state-aided private college, and less than 1% selected another type of private college.

College Persistence and Graduation

Distinct differences can be seen in college persistence and graduation[3] for each of the types of Maryland colleges. A small percentage of the college-going high school students enrolled in state-aided private colleges but, those students graduated at a rate of 70% within six years. Maryland 4-year public colleges enrolled a larger percentage (16%) of the college-going high school students, and graduated students at a rate of 68% within six years.

Community colleges enrolled the largest percentage of college-going high school students, and graduated 13% within three years. Another important measure for community colleges is the rate at which students transfer to a 4-year program. Of the college-going high school students that initially enrolled in a community college, 16% transferred to a 4-year college. As community colleges students often enroll part-time and work full-time, 32% of the college-going high school students enrolled in a community college were still enrolled three years later.

[1] The MLDS may only report aggregate, de-identified data. All the numbers and percentages reported are rounded.
[2] Students were considered enrolled if they entered an in-state or out-of-state college in summer 2008, fall 2008, or spring 2009.
[3] Persistence, graduation, transfer, and no longer enrolled were calculated at 150% of normal program length. For a 2-year program, 150% of normal program length is 3 years. For a 4-year program, 150% of normal program length is 6 years.



*The MLDS may only report aggregate, de-identified data. All the numbers and percentages reported are rounded.
**For a 2-year program, 150% of program length is 3 years; for a 4-year program, 150% of program length is 6 years.
^ "No Longer Enrolled" means the student was not enrolled at the institution in which he or she first enrolled. The student may have enrolled in another institution and graduated or persisted.

Attachment B – Data Requests for 2017

January

1. Towson University requested the number of Towson graduates who completed initial licensure who accepted a teaching position in 2016; # grads who completed initial licensure at Towson in 2014 and accepted a teaching position in 2014; 3 grads who completed initial licensure at Towson in 2014 and are still teaching in 2016. STATUS-completed.
2. *Great Schools*, requested a listing data on high school graduate enrollment, college persistence, and college remediation aggregated by high school. STATUS-request withdrawn.
3. Johns Hopkins University requested records of JHU recent graduates who are employed in Maryland. STATUS-closed (request required unit record data and therefore could not be fulfilled).
4. *Research for Action*, requested student-level Pell grant recipient data. STATUS-closed (request required unit record data and therefore could not be fulfilled).

February

5. *I am O'Kah!*, requested dual enrollment data for Baltimore City students grades 9 through 12. STATUS-completed.
6. Howard County Community College, requested a correction to the MLDS (which had omitted Howard County Community College from the list of Maryland Community Colleges). STATUS-correction completed.

March

7. Towson University requested the number of 2016 graduates employed as teachers in Maryland public schools. STATUS-completed.
8. Doctoral Candidate at Howard University, requested access to high school transcript data and data related to student coursework and standardized test scores. STATUS-request required unit record data and single sector data and therefore could not be fulfilled. Referred to MSDE.
9. Morgan State University, requested data on employment of Morgan bachelor's recipients, employment of current students, and graduates employed as teachers in Maryland. The request also included data for first time undergraduates showing high school math classes taken and grades earned. STATUS-completed.

May

10. MSDE, requested data further disaggregating the MLDS dashboard on postsecondary enrollment by year to provide disaggregation by: (a) by FARMs status and (b) by type of receiving institution (two and four year). STATUS-completed.
11. MSDE, requested data on public school teachers that attended any Maryland public school and public school teachers that are teaching within the same LEA of the K-12 school they attended. STATUS-completed.
12. Morgan State University, requested employment data for students completing a BS in Nursing for the years 2014- 2016. STATUS-ongoing (in review).
13. MHEC, requested a report on the time to degree and workforce outcomes of Community College students who earned an Associate's Degree versus those who earned a certificate. STATUS-completed.
14. UMBC, Data on outcomes of students who take at least one computer science class in high school. Outcomes of interest are whether students continue to take a second computer science class in high school and whether students take at least one CS class in college. STATUS-completed.
15. *Blue Media Inc*., requested a list of MLDS staff with contact information. STATUS-completed.

16. MSDE, Questions related to graduates of teacher preparation programs, to inform policy recommendations required by SB493. STATUS-completed.

June

17. MSDE, data on the average college GPA of the students who became public school teachers as compared to private school teachers and other employment. STATUS-complete.
18. MSDE, for students who graduated with a major in education but did not initially enroll with that major, data on their initial major. STATUS-completed.
19. UMBC, data showing the declared college majors of students who graduated high school having taken high school courses in a computing, engineering pathway. STATUS-completed.
20. USM, requested updated tables from MLDS Center Report, *Assessing the Workforce Outcomes of Maryland Science, Technology, Engineering, and Math (STEM) Postsecondary Graduates.* STATUS-ongoing.
21. MSDE, data comparing the college entry and employment status for CTE concentrators versus non-CTE concentrators. STATUS-completed
22. UMBC, data on computer science teachers including their degree attainment and major course of study at each degree level. The request also seeks information to determine whether there is a correlation between the numbers of certified or endorsed computer science teachers and the number of students taking courses in computer science. STATUS-ongoing.
23. Towson University, request to refresh a prior data request that provided employment outcome information on Towson degree recipients.
24. UMBC, data on the highest level of high school math achieved by students who are in Computing/Engineering pathways, students who have taken any computing in high school, and students majoring in a computing/engineering field in college. STATUS-done.
25. MSDE, the number of novice teachers who attended a Maryland public school and the number who go on to teach in the district in which they attended. STATUS-complete.

August

26. UMBC, data on the part time jobs of students in high school or college and the time to degree. STATUS-ongoing.
27. University of Baltimore, data on the employment outcomes of graduates of the University of Baltimore. STATUS-ongoing.
28. UMBC, data on students who take at least one computer science class in high school the number that continue to take a second computer science class in high school and what majors to the students have in college. STATUS-done.

September

29. MSDE, the number of students who enrolled in postsecondary among students who either met graduation requirements using: world language, advanced technical education, or CTE completers. STATUS-ongoing.

November

30. UMBC, data on the postsecondary enrollment of high school students who took Advanced Placement computer science courses. STATUS-ongoing.
31. Morgan State University, data on the employment status of fall 2016 full-time undergraduate students at Morgan State disaggregated by student major and workforce classification in which the student is employed. STATUS-ongoing.
32. Department of Legislative Services – the number of students who initially enroll in an education program and subsequently graduated from an education program. STATUS-done.
33. Department of Legislative Services – for the teacher dashboards provided on the MLDS website, provide that same information disaggregated by institution and provide the same information for state aided institutions. STATUS-ongoing.

Attachment C. SLDS Issue Brief.

SLDS Issue Brief

Maryland's Synthetic Data Project

The Maryland Longitudinal Data System Center (MLDS Center) is investigating the use of a synthetic data method to increase the amount of rigorous policy research conducted with MLDS data while protecting confidential individual data.

The method would allow policy analysts and researchers to use synthetic data without going through the lengthy approval process required to use confidential data. In addition to increasing access to MLDS data and the data's impact on policy and practice, the project could be a model for states seeking to protect confidential data while encouraging statewide longitudinal data system (SLDS) use for research, training, and evaluation.

Maryland's synthetic data project is the work of the MLDS Center and the Maryland State Department of Education (MSDE) as part of a 2015 SLDS grant awarded by the U.S. Department of Education. The MLDS Center partners with Maryland Higher Education Commission; Maryland Department of Labor, Licensing, and Regulation; MSDE, the University of Maryland, Baltimore; and the University of Maryland, College Park.

What Are Synthetic Data?

The concept of synthetic data was first proposed by Harvard University Professor Donald Rubin in 2012 in response to the access constraints of sensitive individual-level data.[1] The goal of developing synthetic data is to provide publicly available datasets that can be used for valid research analyses in place of the confidential data.

Producing synthetic data requires identifying variables of interest and creating "gold-standard" files that contain the original confidential information. The gold-standard files serve as the basis for creating and evaluating synthetic datasets. Borrowing from imputation methods, or the process of replacing missing data with substituted values, MLDS Center staff members would construct joint distributions of the original variables. Then, they would randomly select values from the joint distributions to create multiple sets of new, or synthetic, data that mimic the actual data.

The synthetic datasets would then be evaluated to verify that their statistical characteristics were sufficiently similar to those of the original data. Before the synthetic datasets would be released, a disclosure risk assessment would be conducted. That assessment would ensure negligible risk of linking synthetic data records to the students, workers, schools, or employers represented in the gold-standard files.

To help verify results, external researchers completing analyses with synthetic datasets could request that the Center replicate the analysis with actual data. Currently, the Survey of Income and Program Participation (SIPP) synthetic data project of the U.S. Census Bureau provides such an option for external researchers.

Maryland's Synthetic Data Project

The MLDS Center serves as a central repository of data from all levels of the state's education and workforce programs. Because such data could be linked to individual students, workers, schools, and employers, the Center treats the data as confidential.

[1] Donald Rubin curriculum vitae, *https://statistics.fas.harvard.edu/files/statistics/files/rubin-cv-june2017.pdf*

This product of the Institute of Education Sciences (IES) SLDS Grant Program was developed with the help of knowledgeable staff from state education agencies and partner organizations. The information presented does not necessarily represent the opinions of the IES SLDS Grant Program. We thank the following people for their valuable contributions:

Chandra Haislet
Maryland State Department of Education

Laura Stapleton
Maryland Longitudinal Data System Center and the University of Maryland, College Park

Michael Woolley
Maryland Longitudinal Data System Center and the University of Maryland, Baltimore

Xiaying Zheng
Maryland Longitudinal Data System Center and the University of Maryland, College Park

Corey Chatis
SLDS Grant Program, State Support Team

Carla Howe, Ph.D.
SLDS Grant Program, State Support Team

For more information on the IES SLDS Grant Program or for support with system development, please visit http://nces.ed.gov/programs/SLDS.

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

Notes on Data Privacy

The federal Family Educational Rights and Privacy Act (FERPA) allows schools to share student data with the SLDS even if families have "opted out" of releasing directory information for other uses. Properly de-identified data can be released from the SLDS without violating FERPA and should be the preferred option for fulfilling research requests.

Review applicable state, federal, and local privacy laws, as well as current privacy policies in each agency involved with data integration. Check with the state attorney to determine state and local requirements for using or releasing student data.

The Privacy Technical Assistance Center (PTAC) team is available to answer questions about data privacy. Contact PTAC at PrivacyTA@ed.gov.

State law[2] limits direct access to data in the MLDS Center to authorized MLDS Center staff members. Additionally, the MLDS Governing Board's policy and MLDS Center regulations strictly limit access to MLDS data, which in turn limits researcher access to unit-record-level data.

The synthetic data method would expand access to such data while protecting confidential data.

The MLDS Center's synthetic data project has four goals:

1. *Create three gold-standard files, which cover K12 to postsecondary education, postsecondary education to workforce, and K12 education to workforce.* The specific variables to be included in the gold-standard files are being determined, in part, by an end-user panel convened in April 2017 to define the needs of interested education and workforce researchers from across Maryland.
2. *Generate multiple sets of synthetic data based on the gold-standard files.* The utility of the multiple synthetic datasets and the potential disclosure risk will be extensively examined before the decision to release the data is brought to the MLDS Governing Board. MLDS Center staff are examining the possibility of taking non-parametric approaches to the synthesis process—specifically, a classification and regression tree (CART) approach—and using information about the research questions of interest to end users.
3. *Disseminate information about the Center's synthetic data via a summit for education and workforce researchers.* An online access portal and training materials will be developed to ensure the utility of synthetic files for interested users.
4. *Examine the feasibility of using synthetic datasets for cluster-level inference analysis.* Education data typically have a hierarchical or multi-level structure in which students are clustered within classrooms and schools, and schools are clustered within local education agencies. The question of whether synthetic data can be developed and used to capture cluster-specific deviations (i.e., random effects) has not been explored previously, but the synthetic data project provides the opportunity to further understand this possibility.

What are the benefits of using synthetic data?

There are several benefits of using synthetic data.

1. Synthetic data allow external researchers to access data at a granular, individual level to allow for more nuanced analyses. Traditionally, a common strategy for protecting sensitive or confidential data was to provide aggregated data, which cannot be used to answer detailed questions and do not allow for many types of analysis.
2. When properly generated, synthetic data can yield comparable results to those from the original data without violating confidentiality.
3. Creating gold-standard datasets and re-running analyses on the original data could help the data-hosting agency better understand its own data and improve the analytic validity of the synthetic datasets.
4. Synthetic data greatly expand the number of researchers—each bringing different backgrounds, expertise, and orientations—to important education and workforce issues in Maryland.

The synthetic data method has great potential to help SLDS agencies efficiently use limited resources. Data stored in the SLDS are expected to become exponentially more useful with the size of the database; the demand from researchers would predictably also increase. Allowing external researchers direct access to individual-level synthetic data through the SLDS could alleviate the burden of handling research requests.

[2] Education Article, §24-707, Annotated Code of Maryland, *http://mgaleg.maryland.gov/webmga/frmStatutesText.aspx?article=ged&section=24-707&ext=html&session=2017RS&tab=subject5*

Additional Resources

Maryland Longitudinal Data System Center
https://mldscenter.maryland.gov

Maryland Longitudinal Data System Center Policy Reports
https://mldscenter.maryland.gov/PolicyReports.html

Maryland State Department of Education
https://www.marylandpublicschools.org/Pages/default.aspx

Nowok B., Raab, G.M., Dibben, C. (2016). synthpop: Bespoke Creation of Synthetic Data in R. *Journal of Statistical Software, 74(11)*, 1-26.
https://www.jstatsoft.org/article/view/v074i11

Raghunathan, T. E., Reiter, J. P., & Rubin, D. B. (2003). Multiple imputation for statistical disclosure limitation. *Journal of Official Statistics, 19(1)*, 1-17.

Reiter, J. P. (2004). Inference for partially synthetic, public use microdata sets. *Survey Methodology, 29(2)*, 181-188.

Reiter, J. P., & Raghunathan, T. E. (2007). The multiple adaptations of multiple imputation. *Journal of the American Statistical Association, 102(480)*, 1462-1471.

Rubin D. B. (1993). Discussion: statistical disclosure limitation. *Journal of Official Statistics, 9(2)*, 461-468.

U.S. Census Bureau's Survey of Income and Program Participation (SIPP) synthetic data project
https://www.census.gov/programs-surveys/sipp/guidance/sipp-synthetic-beta-data-product.html