



**MLDS CENTER**

Maryland Longitudinal  
Data System

Better Data • Informed Choices • Improved Results

# Procedures for New Collections

**August 2023**

---

**Maryland Longitudinal Data System Center**  
**Procedures for New Collections**

## Contents

1. Overview .....	2
2. Definitions .....	2
3. Rationale for Collecting New Data .....	2
Required by Law .....	3
Request from Agency .....	3
Requests from the MLDS Governing Board .....	3
Resolutions of Data Gaps .....	3
4. New Data Collection Process Overview .....	3
New Data Collections Process Flowchart .....	4
Step 1: Initial Meeting with New Data Sharing Partner .....	5
1.1 Schedule Meeting .....	5
1.2 Meeting Content .....	5
Step 2: Determine if Identity Resolution is Possible .....	7
Step 3: Data Sharing Partner Completes Standard Data Elements Document .....	7
3.1 Optional Sub-step: Site Visits .....	8
Step 4: Determine what Data Elements to Collect .....	8
4.1 Schedule Meeting .....	8
4.2 Meeting Content .....	8
Step 5: Establish Data Sharing Agreement .....	9
5.1 Memorandum of Understanding .....	9
5.2 MOU Requirements .....	9
Step 6: Present Data Elements and Scope of Collection to Advisory Boards and Governing Board .....	10
Step 7: MLDS Center Drafts File Layout .....	10
Step 8: Data Sharing Partner Reviews File Layout .....	10
Step 9: MLDS Develops Data Loading Plan .....	11
Step 10: Data Sharing Partner Reviews Data Loading Plan .....	11
Step 11: File Layout and Data Loading Plan are Finalized .....	11
Step 12: Data Sharing Partner Sends Sample File Using MFT .....	11
Step 13: Results of Sample Reviewed Internally and with Data Sharing Partner .....	12
Step 14: Final File Requested .....	12
Step 15: Data Sharing Partner Submits Final File to MLDS Center .....	12
Appendix A: Data Elements for Identity Resolution .....	13
Appendix B: Data Elements for Research Purposes .....	15

**Maryland Longitudinal Data System Center**  
**Procedures for New Collections**

## **1. Overview**

Under Education Article §24-703, Annotated Code of Maryland, the Maryland Longitudinal Data System (MLDS) Center serves as Maryland’s central repository of student and workforce data. As the MLDS Center grows, new data collections will be submitted and incorporated into the MLDS. “New data” may be obtained through new data sharing agreements with state and federal government agencies or a third-party entity that does business with the state, new data collections initiated by existing data sharing partners, or new individual elements added to existing data collections. “New data” does not include “external data” that are collected for a limited program evaluation or study. This Procedures for New Collections document provides a comprehensive overview of the process for collecting new data from state and federal government agencies or a third-party entity that does business with the state.

The MLDS Center currently receives data sets provided by:

- Maryland State Department of Education (MSDE);
- Maryland Higher Education Commission (MHEC);
- Maryland Department of Labor (Labor);
- Maryland Department of Juvenile Services (DJS);
- Maryland Department of Human Services (DHS);
- Maryland Department of Health; and
- Other external data sources.

The new data collection process is reviewed annually through the Data Governance Advisory Board (Data GAB).

## **2. Definitions**

- Data Governance Advisory Board (Data GAB): The MLDS Center Executive Director’s advisory Board established to ensure ongoing input and participation from key stakeholders on the MLDS Data Inventory, data collection, quality, privacy, and security.
- Data Sharing Partner: The entity providing data to the MLDS Center.
- New Data: Data submitted to the MLDS Center that may be obtained through new data sharing agreements with state and federal government agencies or a third-party entity that does business with the state, new data collections initiated by existing data sharing partners, or new individual elements added to existing data collections.
- Research and Policy Advisory Board (RPB): The MLDS Executive Director’s advisory Board established to ensure ongoing input and participation from key stakeholders on the MLDS Center’s research and policy agenda, related output, and general operations.

## **3. Rationale for Collecting New Data**

There are four main reasons the MLDS Center will collect new data. These reasons are described below.

### **Required by Law**

The Maryland General Assembly may pass bills requiring the MLDS Center to collect new data. This may be due to a change in the law's definitions or an addition to the existing structure.

### **Request from Agency**

In some cases, a State agency may want to provide a new data source to the MLDS. For example, the agency may need assistance meeting specific reporting requirements or providing data for grant funding.

### **Requests from the MLDS Governing Board**

There may be instances where data that was not previously collected, but legally could be, is requested by the MLDS Governing Board. Internal and external stakeholders may ask Governing Board members to make these requests.

### **Resolutions of Data Gaps**

The MLDS Center catalogs known gaps in the data and reports them in an annual Data Gap Analysis document. Permissible resolutions for each gap are included in the Data Gap Analysis. The MLDS Center will explore collecting permissible data collections as they become available to help eliminate gaps.

## **4. New Data Collection Process Overview**

Each new data collection will include a number of steps as part of the entire process. Each step falls under one of the following sections, described in more detail in subsequent pages. Throughout this process the MLDS Center Assistant Attorney General (AAG) will be consulted.

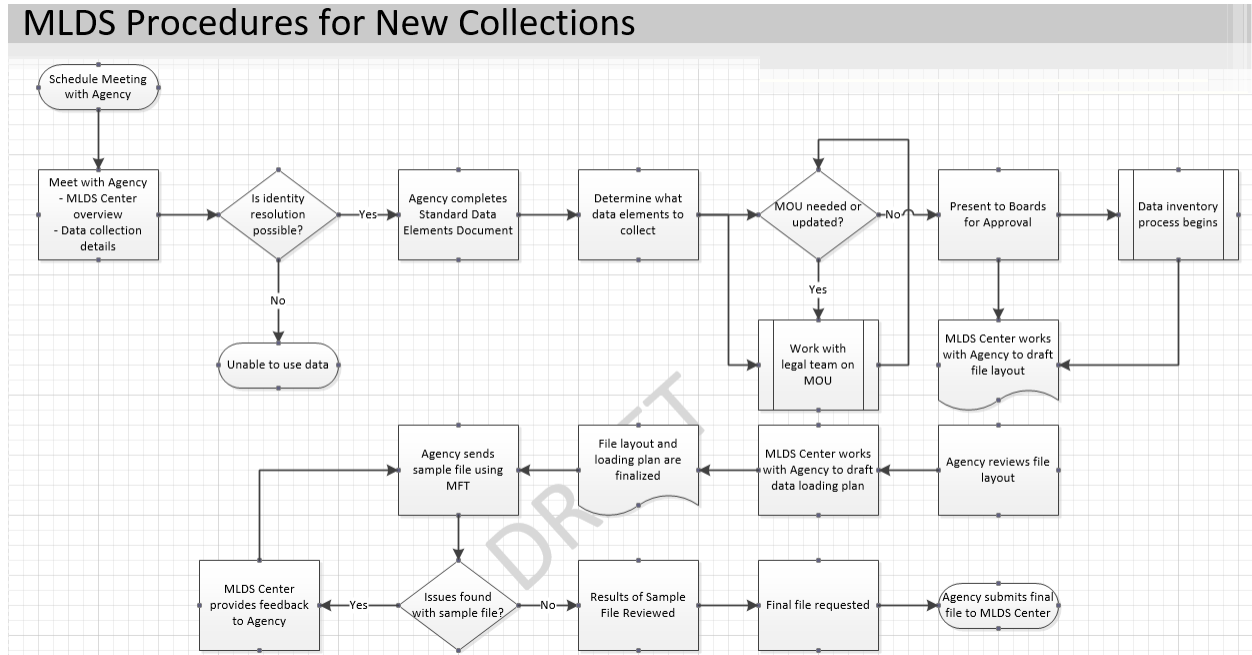
1. Initial Meeting with New Data Sharing Partner
2. Determine if Identity Resolution is Possible
3. Data Sharing Partner Completes Standard Data Elements Document
4. Determine What Data Elements to Collect
5. Establish Data Sharing Agreements
6. Present Data Elements and Scope of Collection to Advisory Boards and Governing Board
7. MLDS Center Drafts File Layout
8. Data Sharing Partner Reviews File Layout
9. MLDS Develops Data Loading Plan
10. Data Sharing Partner Reviews Data Loading Plan
11. File Layout and Loading Plan are Finalized
12. Data Sharing Partner Sends Sample File Using MFT

**Maryland Longitudinal Data System Center  
Procedures for New Collections**

- 13. Results of Sample File Reviewed Internally and with Data Sharing Partner
- 14. Final File Requested
- 15. Data Sharing Partner Submits Final File to MLDS Center

The figure below illustrates the process:

**New Data Collections Process Flowchart**



## **Step 1: Initial Meeting with New Data Sharing Partner**

### **1.1 Schedule Meeting**

- A. The MLDS Center will schedule a meeting with the data sharing partner.
- B. MLDS Center staff will include the Executive Director, Data Management Coordinator, Chief Information Officer, Director of Reporting Services, and Director of Research.
- C. The data sharing partner staff will include individuals who perform such roles as:
  - 1. Data Steward: Responsible for data content, context, and associated business rules.
  - 2. Data Custodian: Responsible for the safe custody, transport, storage of the data, and implementation of business rules.
  - 3. Subject Matter Expert: Has special knowledge and understanding of the content concerning the particular data collection. The Subject Matter Expert may be the same as the Data Steward or Data Custodian.
  - 4. Director of Research: Responsible for leading the research and development strategy.

These roles may also have the job title of Database Administrator, Data Manager, Chief Data Officer, Data Scientist, Data Analyst, etc.

### **1.2 Meeting Content**

The meeting will be structured to provide an overview of the MLDS Center and various questions pertaining to the data sharing partner's data collection. Specifically, the following detailed list of items will be addressed:

- A. Purpose for the meeting and the proposed data collection.
- B. Overview of the MLDS Center, including data governance policies, a description of the MLDS design, data privacy and security procedures, and procedures for new collections.
- C. Discussion on the scope of the data collection. The full population in the data sharing partner's data may not align with the populations of interest to the Center. Therefore, the MLDS Center and the data sharing partner need to discuss what segment of the population should be included in the data provided.
- D. Specific questions about the data sharing partner's data collection. These questions can be provided prior to the initial meeting.
  - a. Background/History
    - i. What is the purpose of the data collection?
    - ii. Is the data already being provided with an existing data sharing partner?
    - iii. Is the data sharing partner the originator of the data or serves as the collector of the data?
    - iv. What are the current reporting requirements associated with the collection/submission?

**Maryland Longitudinal Data System Center**  
**Procedures for New Collections**

- v. How long have the data been collected?
  - vi. Have there been any changes in the collection over time, such as new or removed items and how the data was collected. Is there an annual review of possible changes, such as change in goals/measures, scope of population, etc.?
- b. Data Collection Process
- i. What data systems are used by the data sharing partner for data collection?
  - ii. What are the dates the data are collected during a year?
  - iii. What are the dates the data could be provided to the MLDS Center?
  - iv. What is the estimate of the population size in each annual file?
    - 1. What is the data validation process for this data collection and is there a data quality and correction policy?
  - v. What technical documentation is available to accompany the collection? Are there certain crosswalk documents necessary to fully understand the data?
- c. Data Details
- i. What is the matching ability based on a persistent identifier already incorporated in the MLDS such as a State Assigned Student Identifier (SASID) or Social Security Number (SSN)?
  - ii. What person record level data are available?
    - 1. Is a process in place when multiple records for one identity exist as a result of receiving services from more than one provider? What are the deduplication practices?
    - 2. What are the identity resolution policies?
  - iii. Are there standard aggregation levels used?
    - 1. What are the suppression requirements?
  - iv. Are the data used for reports? If so, what are the main data elements or the elements most commonly used for reporting purposes?
  - v. Is a data dictionary and/or table with all variables available?
- d. Other
- i. Are there data the MLDS Center is prohibited from collecting (Adult criminal background records, child welfare (CINA) records, and medical and health records)?
  - ii. If multiple files will be provided, is there a particular order in which files should be loaded?

**Maryland Longitudinal Data System Center**  
**Procedures for New Collections**

- iii. Should records in this data collection establish an identity in the MLDS?
  - iv. Discuss the value and feasibility of a sit visit.
- E. How the MLDS Center can support the data sharing partner's work.
- a. What value can the MLDS Center data provide?
  - b. What data sharing partner data gaps can be filled using data from the MLDS?
  - c. What [MLDS Center Research Agenda](#) questions can the data help answer?

## **Step 2: Determine if Identity Resolution is Possible**

The ability to identify individuals across disparate data sources is key to the MLDS Center's ability to link records in the MLDS. The current linking strategy is both deterministic and probabilistic in nature. Multiple levels of matching algorithms are used and the matching is conducted in a sequential manner. The MLDS Center determines whether the data sharing partner's data has enough personally identifiable information to connect to a unique identity in the MLDS. If it is determined by the MLDS Center that there is insufficient data to properly match records to the MLDS data, data will not be loaded.

The MLDS Center creates a boilerplate list (see Appendix A) of the data elements needed for identity resolution, including:

- a. First Name,
- b. Last Name,
- c. Middle Name,
- d. Generational Suffix
- e. Date of Birth,
- f. Source Unique ID

The data sharing partner reviews the list and notes with a 'Y' or 'N' whether or not the data elements are available as part of the data sharing partner's data collection and if the data element is a required field at the time of data entry. There is also space for the data sharing partner to include any relevant notes about each specific data element. If an appropriate amount of these seven elements can be provided, matching can take place. Other data elements that are useful for identity resolution include Driver's License Number, Employee ID, Home Address, Sex/Gender, and Race/Ethnicity.

## **Step 3: Data Sharing Partner Completes Standard Data Elements Document**

The MLDS Center creates a boilerplate list of the data elements needed and desired from a data sharing partner for research purposes (see Appendix B). The list contains the expected data elements anticipated based on the type of data collection. The data sharing partner follows the same steps as they did for the identity resolution list. As this is a collaborative process, the data sharing partner should add any data elements they consider helpful to answering research questions.



**Maryland Longitudinal Data System Center**  
**Procedures for New Collections**

### **3.1 Optional Sub-step: Site Visits**

If appropriate, MLDS Center staff may participate in site visits to better understand the data partner's process for collecting data, including how data are input into their data system. Taken from a component of the Baltimore CitiStat model, this involvement enables MLDS Center staff to gain familiarity with how the data partner enters data, discern which fields are required or not required, identify which fields are self-reported or entered using documentation, etc. This data-driven process will also help MLDS Center staff develop detailed internal documentation. To be clear, this is an opportunity for MLDS Center staff to learn about the data collection process and not an evaluation of the data partners' current practices.

- A. MLDS Center staff will work to schedule site visits at convenient times for those involved. If possible, the visit will be scheduled at least 4 weeks in advance.
  - a. The MLDS Center Data Management Coordinator and data analyst assigned to load the data should plan to attend.
  - b. Additional members of the MLDS Center staff may plan to attend as needed.
- B. Prior to arriving for the site visit, the MLDS Center will discuss:
  - a) The essential questions needed to be addressed, including what additional documentation from the data partner is needed;
  - b) Observations to collect during the site visit;
  - c) Plans to collect observations, take notes, and compile findings after the visit; and
  - d) What documentation to bring and provide to the data sharing partner.
- C. After the site visit, MLDS Center staff will discuss the visit with internal staff and determine if any follow up questions or comments are necessary.

### **Step 4: Determine what Data Elements to Collect**

After confirmation that identity resolution is possible and all available data elements are documented, the MLDS Center and the data sharing partner will discuss the appropriate data elements to collect. The MLDS Center and the data sharing partner will discuss the data elements in detail, including the definition, permissible values for each element, imputed values, null values, and any intra-collection element relationships. All staff present at the initial meeting should be in attendance when deciding what elements to collect.

#### **4.1 Schedule Meeting**

- A. The MLDS Center will schedule a meeting with the data sharing partner.
  - 1. Attendees of the Initial Meeting from Step 1 will be included.
  - 2. Each data sharing partner's staff member who helps store and organize the data sharing partner's database will also be included. This role typically has the job title of IT Programmer, IT Systems Specialist, or Database Administrator.

#### **4.2 Meeting Content**

- A. Data Collection Process
  - a. How does the data sharing partner receive the data?
  - b. How does the data sharing partner validate and structure the data?

**Maryland Longitudinal Data System Center**  
**Procedures for New Collections**

- B. Data Loading Recommendations (The MLDS Center will explain current structure)
  - a. Identity establishing collection or subordinate collection
  - b. Order of files to load
  - c. Relationship to other files in the MLDS
- C. Data Elements and Population Scope
  - a. Element definitions
  - b. Element value ranges and outliers
  - c. Imputed values
  - d. Null values
  - e. Intra-collection element relationships
- D. Data Element Limitations/Gaps

## **Step 5: Establish Data Sharing Agreement**

### **5.1 Memorandum of Understanding**

1. The MLDS Center and data sharing partner must have an executed Data Sharing Agreement Memorandum of Understanding (MOU) in place prior to transfer of any data to the MLDS.
2. If the new data collection is from a data sharing partner with an existing MOU, a determination will be made whether the existing MOU covers the new data type. If it does not, an amendment to the existing agreement will need to be established.
3. If this is the first data collection from a data sharing partner then an initial MOU must be created. The development of the MOU should start early in the new data collection process to provide time for legal review.

### **5.2 MOU Requirements**

1. A Memorandum of Understanding will be drafted by the MLDS Center, with input and approval from the Center's AAG.
  - a. The MOU will describe the responsibilities of the data sharing partner to share data. The MOU should list categories of data, not specific data elements.
  - b. The MOU will establish requirements for the Center to receive and maintain the data in compliance with all relevant data privacy laws. The agreement will limit re-disclosure of the data by MLDS except as authorized by State law.
  - c. The MOU will specify security requirements that the Center must maintain.
  - d. The MOU will provide the terms of the agreement.
2. The MOU will be signed by the Chair of the MLDS Governing Board and the Secretary of the data sharing partner (or equivalent) providing the data. AAGs for both agencies will sign the MOU to acknowledge their approval as to its form and legal sufficiency.
3. The MOU should be completed prior to data being added to the Data Inventory.

## **Step 6: Present Data Elements and Scope of Collection to Advisory Boards and Governing Board**

In order to add new data collections and new data elements to the MLDS, the data must be presented and approved by the MLDS Center Governing Board for approval. Prior to presenting to the Governing Board, the Data Governance Advisory Board as well as the Research and Policy Advisory Board must review and comment on the proposed data collection and data elements. Part of the Advisory Boards' review will be to ensure the data is relevant to answering one of the questions in the [MLDS Center Research Agenda](#) and necessary for evaluating federal or State supported education programs.

The data sharing partner providing the new collection will include formal guidance on data usage. This includes background and scope of the collection, data limitations, appropriate use for research or data analysis, and any data governance factors. Meeting content should include the following:

- A. Background
  - 1. Purpose of the data collection
    - i. Reporting requirements
      - 1. Federal, state, local, agency, and/or other
    - ii. Collection scope
      - 1. Records included
      - 2. Records excluded
      - 3. Frequency of collection
      - 4. Calendar period for each file
- B. Limitations/Gaps
  - 1. Data collection
  - 2. Data elements
- C. Appropriate use for research or data analysis
  - 1. Current use by data sharing partner
  - 2. Recommendations for MLDS

## **Step 7: MLDS Center Drafts File Layout**

Once the data elements are approved by the Governing Board, the MLDS Center will create a file layout for data submissions based on the data elements agreed to be collected. The layout will include Data Element Name, Start, Length, End, Type, Definition, and Permitted Values. Agencies will also be instructed to save the file layout as a text file if possible.

## **Step 8: Data Sharing Partner Reviews File Layout**

The data sharing partner will review the drafted file layout and provide feedback and answer any questions the MLDS Center may still have. The data sharing partner's system may have certain capabilities and may require changes to the file layout to more accurately or efficiently transfer their

**Maryland Longitudinal Data System Center**  
**Procedures for New Collections**

data. If the file must be delivered in a different format, the MLDS Center staff will collaborate with the data sharing partner on how to receive the data.

### **Step 9: MLDS Develops Data Loading Plan**

The MLDS Center will develop a data loading plan in collaboration with the data sharing partner providing the data. The data loading plan may include but is not limited to the following recommendations:

1. What corresponding records must exist in the MLDS in order to load records from the new data collection to the Master Data Management (MDM) system and/or the Operational Data Store (ODS).
2. Whether or not the data collection should be used to establish an identity in the MDM.
3. If a required corresponding record is not currently in the MLDS, whether the record may be retained in the MDM.
4. The level at which the data should be loaded (e.g. student-level, school-level, course-level).
5. How corresponding records interact with each other and how this impacts whether a record should be loaded into the MDM or ODS.
6. Any values that may need to be populated based on the relationship between other data collections.

### **Step 10: Data Sharing Partner Reviews Data Loading Plan**

The data sharing partner reviews the drafted loading plan, provides feedback, and answers any remaining questions.

### **Step 11: File Layout and Data Loading Plan are Finalized**

The MLDS Center and data sharing partner both approve the final file layout and data loading plan. The file layout may be adjusted as data collections are altered and updated throughout the years. The data loading plan may also be updated if changes are made to future data collections.

### **Step 12: Data Sharing Partner Sends Sample File Using MFT**

The data sharing partner uses the file format to create a sample file and submits the file using Managed File Transfer (MFT). MLDS Center staff will provide technical assistance as needed. The sample file should consist of all fields and must provide the MLDS Center with a comprehensive understanding of how the data will appear.

Any issues found with the sample file are discussed and the MLDS Center provides feedback to the data sharing partner before they resubmit a sample file. Common issues may include connecting multiple sample files together, clarification on the difference (if any) between "No Value Entered" and those with no value (null), and extra spaces identified.

**Step 13: Results of Sample Reviewed Internally and with Data Sharing Partner**

The MLDS Center will compile and discuss the sample file review results at the next internal MLDS Center Data Meeting. Next, the MLDS Center will provide a review of the results to the data sharing partner.

**Step 14: Final File Requested**

Discussions about the data loading process and sample file are discussed until a final file is agreed upon. After agreement the MLDS Center will request a final file be submitted by the data sharing partner.

**Step 15: Data Sharing Partner Submits Final File to MLDS Center**

The data sharing partner uses the same Managed File Transfer (MFT) process as the sample file to submit their final file. The data sharing partner will confirm the file is final via email and the MLDS Center will confirm receipt of the file. If any issues with the final file are identified after receipt, the MLDS Center will work with the data sharing partner and, if necessary, determine a timeline for resubmission. Subsequent files will be submitted based on the [Data Collection Calendar](#) for each year.

**Maryland Longitudinal Data System Center**  
**Procedures for New Collections**

**Appendix A: Data Elements for Identity Resolution**

**Gathering DATA SHARING PARTNER TYPE Data for the MLDS Center**

Please review the table below and note with a 'Y' or 'N' whether or not the data element is available in DATA SHARING PARTNER's database. The table contains a list of data elements the MLDS Center needs to collect for identity resolution purposes.

Please also indicate with a 'Y' or 'N' whether the element is required during the data entry process. Include any relevant notes about the data such as whether or not it is a text field.

If possible, data starting from July 1, 2007 is requested.

<b>Primary Elements Used for Identity Resolution</b>				
<b>Data Element Name</b>	<b>Available (Y or N)</b>	<b>Required? (Y or N)</b>	<b>Notes</b>	<b>Load?</b>
First Name				
Last Name				
Middle Name				
Generational Suffix				
Date of Birth				
Social Security Number				
Source Unique ID				
<b>Additionally Useful Elements for Identity Resolution</b>				
Sex/Gender				
Race/Ethnicity				
Driver's License Number				
Employee ID				
License Number				
Home Address				
Home City				
Home Zip Code				
Home County				

**Maryland Longitudinal Data System Center**  
**Procedures for New Collections**

Home State				
School/Institution				

**Maryland Longitudinal Data System Center**  
**Procedures for New Collections**

**Appendix B: Data Elements for Research Purposes**

**Gathering DATA SHARING PARTNER TYPE Data for the MLDS Center**

Please review the table below and note with a 'Y' or 'N' whether or not the data element is available in DATA SHARING PARTNER's database. The table contains data elements desired for collection. Please add others that could be helpful in answering research questions.

Please also indicate with a 'Y' or 'N' whether the element is required during the data entry process. Include any relevant notes about the data, such as whether or not it is a text field.

If possible, data starting from July 1, 2007 is requested.

<b>Individual TYPE Useful / Research Information</b>				
<b>Data Element Name</b>	<b>Available (Y or N)</b>	<b>Required? (Y or N)</b>	<b>Notes</b>	<b>Load?</b>
Program Information				
Unique Program Identifier				
Program Name				
Program Type				
Program Level				
Admission Reason				
Program Start Date				
Program End Date				
Release Reason				
Program Status				
Program Length				
Total Weekly Hours				
Starting Wage				
Ending Wage				
Attainment				
Retention				



**Maryland Longitudinal Data System Center  
Procedures for New Collections**

Supervision Type				
Supervision Start Date				
Supervision End Date				
Behavioral Information				
Type of Behavior/Incident				
Behavior Code				
Date of Behavior				
Response to Behavior				
Length of Response				
Location of Incident				
Case Processing Date				
Case Processing Outcome				
Location/Geocode Information				
Census Tract				
Census Block				
Latitude				
Longitude				
Assessment Information				
Assessment Name				
Assessment Date				
Raw Score				
Scale Score				