Enterprise Information Management Program

Pilot Outcome Report

Report from the EIM Team: Len Silverston, Keith Glenn, Stacie Demchak, and Dee Abueldahab

Version: 1.3

Date: 9/2/11
# Revision Chart

<table>
<thead>
<tr>
<th>Version</th>
<th>Primary Author(s)</th>
<th>Description of Version</th>
<th>Date Completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>Len Silverston/Stacie Demchak</td>
<td>Document creation</td>
<td>6/27/11</td>
</tr>
<tr>
<td>1.1</td>
<td>Len Silverston/Stacie Demchak</td>
<td>Document Draft Version</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>Len Silverston/Stacie Demchak</td>
<td>Draft Version</td>
<td>7/26/11</td>
</tr>
<tr>
<td>1.3</td>
<td>Len Silverston/Stacie Demchak</td>
<td>Final Version</td>
<td>9/2/11</td>
</tr>
</tbody>
</table>
# Contents

1. Executive Summary .................................................................................................................. 1  
   1.1 Summary ............................................................................................................................. 1  
2. Pilot Team .................................................................................................................................. 2  
   2.1 Pilot participants ................................................................................................................... 2  
   2.2 Pilot Core Team Members and Their Respective Roles ....................................................... 2  
3. Project Overview ........................................................................................................................ 3  
   3.1 Project Description ................................................................................................................. 3  
   3.2 Pilot Project Business Requirement Status ........................................................................... 5  
   3.3 Pilot Results .......................................................................................................................... 7  
   3.4 Tools used .............................................................................................................................. 8  
   3.5 Major Project Milestones ....................................................................................................... 9  
4. Business Requirement Detail ........................................................................................................ 10  
   4.1 Logical Data Model .............................................................................................................. 10  
   4.2 Conceptual Data Model ....................................................................................................... 11  
   4.3 Business Data Model ......................................................................................................... 12  
   4.4 Physical Database Design .................................................................................................... 13  
   4.5 Data Definition Sample ....................................................................................................... 14  
   4.6 Data Quality Report Format ................................................................................................ 15  
   4.7 RACI Documentation .......................................................................................................... 16  
   4.8 Data Issue Documentation .................................................................................................... 19  
5. New Processes Suggested ............................................................................................................ 22  
   Current Process ...................................................................................................................... 22  
   Suggested Processes ............................................................................................................... 23  
6. Lessons Learned .......................................................................................................................... 25  
7. Recommendations ....................................................................................................................... 26  
8. Next steps .................................................................................................................................... 27
1. EXECUTIVE SUMMARY

1.1 Summary
This report documents the results of a pilot project conducted by the enterprise information management (EIM) program for the Colorado Department of Education.

The EIM program will define and facilitate implementation of governance processes, architectural designs, data models, strategies, data quality improvements, data security administration and effective information delivery.

The pilot project's objective was to kick off the EIM program and specifically to establish processes for improving the collection, aggregation and dissemination of data at CDE. The pilot focused on a small amount of core data from our enterprise data model (eight fields from four data collections) to help us refine and expand our approach to our Enterprise Information Management.
2. PILOT TEAM

2.1 Pilot participants

The following people participated and contributed to this pilot

<table>
<thead>
<tr>
<th>Participants</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EIM Core Team consisting of Keith Glenn, Dee Abueldahab, Stacie Demchak, Mani Bhoraniya, Jerry Taylor, Len Silverston</td>
<td>Define and implement Enterprise Information Pilot Project</td>
</tr>
<tr>
<td>Yoon Shin, June Maginnis, Debbie Puccetti, Jan Rose Petro, Kristi Gleason, Annette Severson, Sonia Vargas, Brenda Meyer</td>
<td>Data governance team who worked on data definitions, common rules, defining data issues and suggesting new processes</td>
</tr>
</tbody>
</table>

2.2 Pilot Core Team Members and Their Respective Roles

<table>
<thead>
<tr>
<th>Role</th>
<th>Name/Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program Sponsor</td>
<td>Dan Domagala, CDE CIO</td>
</tr>
<tr>
<td>Program Manager</td>
<td>Keith Glenn, Data Warehouse Manager</td>
</tr>
<tr>
<td>Enterprise Architect</td>
<td>Dee Abueldahab</td>
</tr>
<tr>
<td>Pilot Data Steward/Business Analyst</td>
<td>Stacie Demchak</td>
</tr>
<tr>
<td>Data Analyst</td>
<td>Mani Bhoraniya</td>
</tr>
<tr>
<td>Subject Matter Expert</td>
<td>Jerry Taylor</td>
</tr>
<tr>
<td>Consultant</td>
<td>Len Silverston</td>
</tr>
</tbody>
</table>
3. PROJECT OVERVIEW

3.1 Project Description

The EIM pilot focused on defining the data issues, definitions, rules, processes, responsibilities, and accountabilities and conducted data governance tasks for the targeted student demographics elements. The team defined common CDE enterprise-wide edit rules (as opposed to the current edit rules, which vary, by collection), identified data issues in the current processes/systems and suggested new processes for the future. This will pave the way for bringing this data into a new standardized structure that has common definitions and rules, and is built based upon CDE’s enterprise data model. The specific attributes of this pilot are the first name, middle name, last name, suffix (added during the pilot), race, ethnicity, gender, date of birth, and SASID (State assigned student identifier). The four collections in the pilot are Student October, Student End of Year, Special Education December Count, and Special Education End of Year. In this pilot, we identified data inconsistencies across collections, and recommend next steps such as improved processes and ongoing rules for handling inconsistencies.

Some key outcomes of this pilot are:

- Common, enterprise-wide definitions and business rules for the pilot data, independent of particular collections
- Current data issues were identified, existing processes were documented, and future processes suggested
- Established a benchmark of the percentage of data inconsistencies across the pilot collections
- Suggested decision rights and accountabilities were documented
- Decision rights and accountabilities for the pilot data were assigned to specific CDE staff using a RACI (Responsible, Accountable, Consulted, Informed) matrix. A SharePoint site was established to document data issues, data definitions, rules, the RACI chart, suggested processes and other data governance documents.
- 16 data issues were identified and suggested solutions were documented.
The diagram below shows the functionality that the EIM (Enterprise Information Management) team envisioned for the future. We worked towards this vision during this pilot. Based upon what we learned from the pilot as well as our research on the new CODEX (ADE 2.0) solution, this diagram will evolve.
## 3.2 Pilot Project Business Requirement Status

<table>
<thead>
<tr>
<th>Goal ID</th>
<th>Business Goal/Objective or Business Requirement</th>
<th>Type</th>
<th>Status</th>
<th>Actual Deliverable Produced</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A subject data area model that is part of the conceptual data model and shows how this pilot fits in with the CDE enterprise data model</td>
<td>Data Model</td>
<td>Completed</td>
<td>Subject data area model (high level data model known as a Conceptual Data Model)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>A business data model</td>
<td>Data Model</td>
<td>Completed</td>
<td>Business Data Model</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>A logical data model</td>
<td>Data Model</td>
<td>Completed</td>
<td>Logical Data Model</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>A physical database design</td>
<td>Data Model</td>
<td>Completed</td>
<td>Physical Database Design</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Develop functional requirements regarding CDE enterprise wide definitions. Specify what will be included in the definition and provide examples. For example, will we use CDE definitions, NEDM, NIEM, etc and where will this be maintained, what types of metadata such as format, valid values, etc. For each entity and attribute in the logical and business data model, we will maintain a definition. This will be the CDE data definition from the CDE dictionary. If there is no definition from the dictionary, the data steward will develop and socialize a definition. Included in the definition, there will be examples.</td>
<td>Data Definition</td>
<td>Completed</td>
<td>Suggested enterprise-wide data definitions and rules were produced during this pilot. Sample Data Definition</td>
<td></td>
</tr>
<tr>
<td>Goal ID</td>
<td>Business Goal/Objective or Business Requirement</td>
<td>Type</td>
<td>Status</td>
<td>Actual Deliverable Produced</td>
<td>Comment</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------------</td>
<td>-----------------------</td>
<td>------------</td>
<td>----------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>7</td>
<td>For each attribute, we will maintain the standard format, valid values, and examples for that attribute.</td>
<td>Data Definition</td>
<td>Completed</td>
<td>For each attribute in the pilot, the team developed enterprise wide and consistent standard format, valid values, and examples of data.</td>
<td>Sample Data Definition</td>
</tr>
<tr>
<td>8</td>
<td>Initial Data Quality Report - the number of inconsistencies that currently exists in data between collections, for example, when the gender is different from one collection to another.</td>
<td>Data Quality</td>
<td>Completed</td>
<td>Data Quality Report</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Report(s) showing benchmark of existing data quality. In the first report we will show data inconsistencies (at a detailed level and at a summary). In the second report, we will categorize the types of edits and errors.</td>
<td>Data Quality</td>
<td>Eliminated from Scope</td>
<td>This was eliminated from scope because it is not possible to determine the root cause of why the data was different. Different business rules and edits for each pilot collection and collecting/storing demographic data with every collection are the major reasons for the data inconsistencies.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>We will use a RACI (Responsible/Accountable/Consulted/Informed) matrix to document who is responsible, accountable, consulted, and informed regarding these data attributes in the pilot.</td>
<td>Data Governance</td>
<td>Completed</td>
<td>RACI Documentation</td>
<td>This was completed for the pilot data elements but was not utilized as intended. The process for the pilot worked differently than was envisioned.</td>
</tr>
<tr>
<td>Goal ID</td>
<td>Business Goal/Objective or Business Requirement</td>
<td>Type</td>
<td>Status</td>
<td>Actual Deliverable Produced</td>
<td>Comment</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------</td>
<td>------</td>
<td>--------</td>
<td>-----------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>11</td>
<td>We will document data issues - the issue number, issue date of initial documentation, issue name, issue description, root cause of issue, issue severity, identification if this issue should be escalated to a data governance board, issue possibilities, and issue recommended solution.</td>
<td>Data Issues</td>
<td>Completed</td>
<td>Data Issue Documentation</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Data Collection/ETL – Define the requirements for developing ETL, which will extract data from the collections and load to the IDS as well as to the issues database, when there are data issues.</td>
<td>Data Collection</td>
<td>Eliminated from scope</td>
<td>Deferred pending CODEX (ADE 2.0)</td>
<td></td>
</tr>
</tbody>
</table>

### 3.3 Pilot Results

*This describes outcomes from the pilot*

<table>
<thead>
<tr>
<th>The pilot helped CDE move towards the following outcomes:</th>
<th>Pilot Results/Learnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improve Data Quality</td>
<td>The pilot resulted in standardized, consistent edit rules. This will eliminate 95% of the data inconsistencies (for example, null versus NMN (no middle name)). The team identified and documented key data issues and recommended solutions. The team developed common definitions, valid values, examples and other metadata for the pilot data elements on a CDE enterprise wide basis, eliminating discrepancies between collections.</td>
</tr>
<tr>
<td>Process Improvement</td>
<td>New processes were recommended by data stewards to improve quality, such as communications to districts when there are inconsistencies.</td>
</tr>
</tbody>
</table>
The pilot helped CDE move towards the following outcomes:

<table>
<thead>
<tr>
<th>Pilot Results/Learnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>The suggested processes will result in better data quality via notifications to districts and more efficient data collections.</td>
</tr>
</tbody>
</table>

Specifically, this pilot:

- Took a step towards improving the quality of data for student demographics by:
  - Getting a baseline measurement of the data quality and consistency of student demographic data that exists in many current collections (initially we looked at Student October, Student December count (special ed), Student EOY, Special Ed EOY)
  - Defining common CDE enterprise wide rules and processes that can improve data quality, especially inconsistencies between various collections e.g., Student October may report as Hispanic and Student EOY may report this as non-Hispanic when a student changes districts – what is the business rule and process?
- Documented a cross collection, consistent enterprise-wide view of the definitions (and other metadata) of the attributes in this pilot.
- Created suggested processes that can increase data quality and improve efficiencies in data collections for the following student demographic attributes:
  - Last name, middle name, first name, suffix, race, ethnicity, date of birth, gender, SASID,
- Provided the opportunity to gather experience and feedback on the data governance program.

### 3.4 Tools used

*The software used for this project, included the following:*

<table>
<thead>
<tr>
<th>Software Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER Studio Business Architect</td>
<td>Maintained the information requirements for the pilot (as well as the CDE enterprise conceptual data model that provided context for this effort)</td>
</tr>
<tr>
<td>ER Studio Data Architect</td>
<td>Maintained the logical and physical data models for the pilot (as well as the CDE enterprise data model that provided context for this effort)</td>
</tr>
<tr>
<td>SharePoint</td>
<td>This was utilized for documenting the data definitions, rules, issues and suggested processes for the pilot.</td>
</tr>
</tbody>
</table>
3.5 Major Project Milestones

The following table shows when major milestones for the project were completed:

<table>
<thead>
<tr>
<th>Milestone</th>
<th>Completion Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall requirements for pilot, implementation of IDS for 8 elements and CDE-enterprise wide definitions</td>
<td>Nov 30, 2010</td>
</tr>
<tr>
<td>Creation of data inconsistencies report</td>
<td>Jan 1, 2011</td>
</tr>
<tr>
<td>Creation of business, logical, and physical data models (for IDS and for issues database)</td>
<td>Feb 1, 2011</td>
</tr>
<tr>
<td>RACI matrix</td>
<td>Mar 15, 2011</td>
</tr>
<tr>
<td>Issues, rules and processes suggested from pilot as well as enterprise wide data definitions (and other metadata)</td>
<td>June 1, 2011</td>
</tr>
<tr>
<td>Pilot report</td>
<td>September 2, 2011</td>
</tr>
</tbody>
</table>
4. BUSINESS REQUIREMENT DETAIL

4.1 Logical Data Model

The following data model was developed after several iterations and forms the basis for the intended database design.
4.2 Conceptual Data Model

We used the following Conceptual Data model as a contextual, integrated model to demonstrate how the pilot data elements fit within the CDE’s whole data model.
4.3 Business Data Model

STUDENT

SASID
First Name
Middle Name
Last Name
Suffix Name
Gender
Date Of Birth
Ethnicity
Race: American Indian or Alaskan Native
Race: Asian
Race: Black or African American
Race: White
Race: Native Hawaiian or Other Pacific Islander
Federal Race/Ethnicity Reporting Category
### 4.4 Physical Database Design

**ETHNICITY TYPES**
- **ETHNICITY_TYPES_ID (PK)**: NUMBER NOT NULL
- **NAME**: VARCHAR2(60) NOT NULL
- **DESCRIPTION**: VARCHAR2(255) NOT NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**PERSON_RACES**
- **PERSON_RACES_ID (PK)**: NUMBER NOT NULL
- **PERSONS_ID (FK)**: NUMBER NULL
- **RACE_TYPES_ID (FK)**: NUMBER NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **PREVIOUS_UPDATE_TIMESTAMP**: TIMESTAMP(6) NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**PERSON_RACES_HISTORY**
- **PERSON_RACES_HISTORY_ID (PK)**: NUMBER NOT NULL
- **PERSONS_HISTORY_ID (FK)**: NUMBER NULL
- **PERSON_RACES_ID (FK)**: NUMBER NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **PREVIOUS_UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**PERSONS**
- **PERSONS_ID (PK)**: NUMBER NOT NULL
- **GENDER_TYPES_ID (FK)**: NUMBER NULL
- **ETHNICITY_TYPES_ID (FK)**: NUMBER NULL
- **FIRST_NAME**: VARCHAR2(60) NOT NULL
- **MIDDLE_NAME**: VARCHAR2(60) NOT NULL
- **LAST_NAME**: VARCHAR2(60) NOT NULL
- **BIRTH_DATE**: TIMESTAMP(6) NOT NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**PERSONS_HISTORY**
- **PERSONS_HISTORY_ID (PK)**: NUMBER NOT NULL
- **PERSONS_ID (FK)**: NUMBER NULL
- **GENDER_TYPES_ID (FK)**: NUMBER NULL
- **ETHNICITY_TYPES_ID (FK)**: NUMBER NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **PREVIOUS_UPDATE_TIMESTAMP**: TIMESTAMP(6) NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**PERSONS_HISTORY_ID (PK)**
- **PERSONS_HISTORY_ID**: NUMBER NOT NULL
- **PERSONS_ID**: NUMBER NULL
- **GENDER_TYPES_ID**: NUMBER NULL
- **ETHNICITY_TYPES_ID**: NUMBER NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **PREVIOUS_UPDATE_TIMESTAMP**: TIMESTAMP(6) NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**PERSON_RACES_HISTORY_ID (PK)**
- **PERSON_RACES_HISTORY_ID**: NUMBER NOT NULL
- **PERSONS_HISTORY_ID**: NUMBER NULL
- **PERSON_RACES_ID**: NUMBER NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **PREVIOUS_UPDATE_TIMESTAMP**: TIMESTAMP(6) NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**APPLICATIONS**
- **APP_ID (PK)**: NUMBER NOT NULL
- **NAME**: VARCHAR2(60) NOT NULL
- **URL**: VARCHAR2(255) NOT NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**PERSONS_ID (PK)**
- **PERSONS_ID**: NUMBER NOT NULL
- **FIRST_NAME**: VARCHAR2(60) NOT NULL
- **MIDDLE_NAME**: VARCHAR2(60) NOT NULL
- **LAST_NAME**: VARCHAR2(60) NOT NULL
- **BIRTH_DATE**: TIMESTAMP(6) NOT NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**GENDER_TYPES**
- **GENDER_TYPES_ID (PK)**: NUMBER NOT NULL
- **NAME**: VARCHAR2(60) NOT NULL
- **DESCRIPTION**: VARCHAR2(255) NOT NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**ETHNICITY_TYPES_ID (PK)**
- **ETHNICITY_TYPES_ID**: NUMBER NOT NULL
- **NAME**: VARCHAR2(60) NOT NULL
- **DESCRIPTION**: VARCHAR2(255) NOT NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**PERSON_RACES**
- **PERSON_RACES_ID (PK)**: NUMBER NOT NULL
- **PERSONS_ID (FK)**: NUMBER NULL
- **RACE_TYPES_ID (FK)**: NUMBER NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **PREVIOUS_UPDATE_TIMESTAMP**: TIMESTAMP(6) NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**PERSON_RACES_HISTORY**
- **PERSON_RACES_HISTORY_ID (PK)**: NUMBER NOT NULL
- **PERSONS_HISTORY_ID (FK)**: NUMBER NULL
- **PERSON_RACES_ID (FK)**: NUMBER NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **PREVIOUS_UPDATE_TIMESTAMP**: TIMESTAMP(6) NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**PERSONS_ID (PK)**
- **PERSONS_ID**: NUMBER NOT NULL
- **GENDER_TYPES_ID**: NUMBER NULL
- **ETHNICITY_TYPES_ID**: NUMBER NULL
- **FIRST_NAME**: VARCHAR2(60) NOT NULL
- **MIDDLE_NAME**: VARCHAR2(60) NOT NULL
- **LAST_NAME**: VARCHAR2(60) NOT NULL
- **BIRTH_DATE**: TIMESTAMP(6) NOT NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**PERSONS_RACES**
- **PERSONS_RACES_ID (PK)**: NUMBER NOT NULL
- **PERSONS_ID (FK)**: NUMBER NULL
- **RACE_TYPES_ID (FK)**: NUMBER NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL

**PERSON_SASID**
- **PERSONS_ID (PK)**: NUMBER NOT NULL
- **SASID**: NUMBER NOT NULL
- **UPDATE_TIMESTAMP**: TIMESTAMP(6) NOT NULL
- **UPDATE_USER_ID**: VARCHAR2(10) NULL
4.5 Data Definition Sample

A surname indicated in the legal documentation such as a birth certificate or passport. Note: this is not necessarily the name by which the person goes by. This usually is the name borne in common by members of a family but it may not be.

- Alpha numeric, could have punctuation such as dashes, hyphens, and spaces.
- Examples: Martinez, Hernandez, Martinez

No risk name(s) or alternative spelling.
- If the last name is a field in a collection and the form is not allowing a risk name, then the field is required.
- When data is collected to do a “count” or other statistical calculation, the last name is optional and depends on the collection.

Forms of valid legal documentation to determine a person’s last name include birth certificates, passports, driver’s licenses, immigration documents, or adoption paperwork.
- If there is no legal documentation (e.g., a strong preference is evident), consider asking the parent or guardian.
- Other documentation that may be available include baptismal record, entries in a family Bible, hospital certificate of birth, previous verification, school record, life insurance policies, or physician certificate.
- If there is no documentation then the parent or guardian’s name may be used.

Allow numbers (may special characters using standard query technique) (but not allowing the entire name in a field label of English language).
- Initials are only allowed if represented on the legal documentation.
- If a person has multiple surnames, therefore all should be included in this attribute in the order as defined by the documentation provided.
4.6  Data Quality Report Format

<table>
<thead>
<tr>
<th>Number of Unique SASIDs in all four collections (2009-2010 Data):</th>
<th>889,759</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Unique SASIDs in Student October:</td>
<td>824,587</td>
</tr>
<tr>
<td>Number of Unique SASIDs in Special Ed. December count:</td>
<td>83,662</td>
</tr>
<tr>
<td>Number of Unique SASIDs in Student End of Year</td>
<td>877,512</td>
</tr>
<tr>
<td>Number of Unique SASIDs in Special Ed. End of Year</td>
<td>106,397</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of SASIDs with more than one attribute</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of SASIDs with more than one First Name</td>
<td>3,160</td>
</tr>
<tr>
<td>Number of SASIDs with more than one Middle Name</td>
<td>31,893</td>
</tr>
<tr>
<td>Number of SASIDs with more than one Last Name</td>
<td>10,208</td>
</tr>
<tr>
<td>Number of SASIDs with more than one Birthdates</td>
<td>2,434</td>
</tr>
<tr>
<td>Number of SASIDs with more than one Gender</td>
<td>705</td>
</tr>
<tr>
<td>Number of SASIDs with more than one Race/Ethnicity</td>
<td>10,533</td>
</tr>
</tbody>
</table>

Total # of errors: 58,933

Number of Students with at least one attribute is different: 51,859 5.83%

*Note: Report does not include null/invalid SASIDs*
4.7 RACI Documentation

This is the RACI matrix that was created by the whole team (core team and data governance team) showing suggested responsibilities, accountabilities, who should be consulted and informed regarding the attributes in this pilot.

<table>
<thead>
<tr>
<th>Accountability Matrix - Student Demographics</th>
<th>Last Name</th>
<th>First Name</th>
<th>Middle Name</th>
<th>SASI D</th>
<th>Birthdate</th>
<th>Ethnicity</th>
<th>Race</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAR Manager</td>
<td>Jan Petro</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Enterprise Wide Data Steward (Person and Student)</td>
<td>Yoon Shin</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>EOY Data Custodian</td>
<td>Yoon Shin</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Student October Data Custodian</td>
<td>Annette Severs on</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>RITS Data Owner (Person and Student)</td>
<td>Jan Petro</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>RITS Data Expert</td>
<td>Debbie Puccetti</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>RITS Data Custodian</td>
<td>Debbie Puccetti</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Special Education Business Data Owner</td>
<td>Mary Greenwood</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Role</td>
<td>Name</td>
<td>Expertise</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>---------------------------------------------------------------------</td>
<td>----------------------</td>
<td>------------------------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Special Education Data Expert</td>
<td>Sonia Vargas</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Special Education Data Custodian</td>
<td>Kristi Gleason</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>EIM Team EIM Program Sponsor</td>
<td>Dan Domagala</td>
<td>I/C</td>
<td>I/C</td>
<td>I/C</td>
<td>I/C</td>
<td>I/C</td>
<td>I/C</td>
<td>I/C</td>
</tr>
<tr>
<td>EIM Team EIM Program Manager</td>
<td>Keith Glenn</td>
<td>I/C</td>
<td>I/C</td>
<td>I/C</td>
<td>I/C</td>
<td>I/C</td>
<td>I/C</td>
<td>I/C</td>
</tr>
<tr>
<td>EIM Team Enterprize-wide Data Business Analyst (Person and Student)</td>
<td>Stacie Demchak</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>EIM Team Enterprize-wide Data Modeler (Person and Student)</td>
<td>Dee Abueldahab</td>
<td>R (logical data model)</td>
<td>R</td>
<td>R (logical data model)</td>
<td>R (logical data model)</td>
<td>R (logical data model)</td>
<td>R (logical data model)</td>
<td>R (logical data model)</td>
</tr>
<tr>
<td>EIM Team Enterprize-wide Data Modeler (Person and Student)</td>
<td>Keith Glenn</td>
<td>R (logical data model)</td>
<td>R</td>
<td>R (logical data model)</td>
<td>R (logical data model)</td>
<td>R (logical data model)</td>
<td>R (logical data model)</td>
<td>R (logical data model)</td>
</tr>
</tbody>
</table>
– R - Responsible:

**PMO definition:**

– Individuals who perform a task/activity; the doer, responsible for action/implementation.

– The degree of responsibility is defined by the Accountable person.

– Responsibility can be shared.

– Responsibility can be delegated.

**Definition as applied to data governance:**

– Person/organization who is responsible for the enterprise wide definition, the quality of the data, the ongoing processes (and suggestions to improve these processes), and the access rights and appropriately use of this data.

– A - Accountable:

**PMO definition:**

– The individual who has ultimate accountability and authority.

– There is only one accountable (A) to each task/activity.

– Accountability is assigned at the lowest level and implied at higher levels.

– Accountability cannot be delegated person who is ultimately accountable for this data and to whom the person responsible is accountable.

**Definition as applied to data governance:**

– This person/organization has bottom line responsibility for the structure and content of this data.

– C - Consulted:

**PMO definition:**

– The individuals to be consulted prior to a final decision or action is taken.

– Two-way communication.
This role will be involved in working group meetings.

**Definition as applied to data governance:**

- Individual that needs to provide feedback and contribute to the data definitions, data quality, and/or data processes.

- **I - Informed:**

  **PMO definition:**

  - The individuals that need to be informed after a decision or action is taken.

  **Definition as applied to data governance:**

  - The individuals that need to be informed after a decision or action is taken regarding data definitions, rules, processes, or data issues.

### 4.8 Data Issue Documentation

The team identified, documented and worked on possible solutions for 16 data issues regarding the data attributes in the pilot and they are summarized below.

<table>
<thead>
<tr>
<th>ID</th>
<th>Field</th>
<th>Issue Type</th>
<th>Issue Name</th>
<th>Status</th>
<th>Priority</th>
<th>Description</th>
<th>Escalate?</th>
<th>Impact</th>
<th>Data Steward</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Middle Name</td>
<td>Data</td>
<td>“NMN” in Middle Name Field</td>
<td>Proposed</td>
<td>Medium</td>
<td>Should NMN be allowed, or should the field just be left blank if there's no middle name?</td>
<td>0</td>
<td>Students are counted twice, funding and reporting statistics.</td>
<td>Yoon Shin</td>
</tr>
<tr>
<td>3</td>
<td>Process</td>
<td>Different information for the same SASID collected during the same collection</td>
<td>Proposed</td>
<td>High</td>
<td>A different value is reported for the same SASID during the same collection, causing duplicate counts. Currently, the last reported data is what is kept in RITS and the data warehouse as the current data. From Yoon - This is issue. We will need to build the edits in place (business rules) to prohibit this. Currently, Post collection process associated with 4 collections may causing this issue.</td>
<td>0</td>
<td>Students are counted twice, funding and reporting statistics.</td>
<td>Yoon Shin</td>
<td></td>
</tr>
</tbody>
</table>
4 | Process | Inconsistent Data Across Collections | (3) Proposed | (2) High | Yoon - This is valid issue. Students' information may update during school years. We need business rules (e.g. CDE send notification if 45% of their students information change within 1 year). This will require new business rules and process with new application. | 0 | Yoon Shin |

5 | Data | Should Suffix/Title be a separate field? | (3) Proposed | (3) Medium | For ease of data-matching should the suffix/title be a separate field? | 0 | Yoon Shin |

6 | Last Name | Need to discuss what to do with people who have two last names. | (1) Pending | (3) Medium | There is a problem with data-matching for people who have two last names because there's no consistent way to record the data. | 0 | Lack of consistency makes data matching complex and time-consuming. | Yoon Shin |

7 | Storing upper and lowercase letters | (3) Resolved | (3) Medium | The name field definitions recommend storing names with the first initial capitalized. Is it feasible to collect and store the data in this format? | 0 | Yoon Shin |

8 | Race and Ethnicity | Race and Ethnicity - switch to alpha codes | (3) Proposed | (3) Medium | Should we switch from using numeric to alpha codes? Need to discuss impacts. | 0 | There is a financial impact to districts. With any of these changes, please set the implementation date well into the future so LEAs can plan and give vendors plenty of notice. | Yoon Shin |

9 | Policy | Authoritative source for last name | (3) Proposed | (2) High | What can we do to encourage the districts to follow the rules for submission of legal last name, especially given the current situation of not having consequences for doing so, since this is locally controlled? This is a potential data governance council issue since it may required executive policy. | 0 | Yoon Shin |
| 10 | Data | Home school and special ed and non-public - special circumstances | (3) Proposed | (2) High | There are some collections where the demographic student data is not collected and it is not required. CDE has no authority to collect this data. Non-public school provides count only. Home school and special circumstance - we get individual records but no SASID or name. | 0 | Yoon Shin |
| 13 | Federal Reporting Code | District Feedback | Do we need to capture federal reporting code (type = district feedback) | (3) Proposed | (2) High | Do we need to capture federal reporting code? Since we can derive this information is it necessary to record this. This is an issue that we need to discuss with the district. | 0 | Yoon Shin |
| 14 | First name | No first name - NFN | (3) Proposed | (2) High | Need to get more clarity on the exact rules for when there is not a first name - NFN (Home school and special circumstance) | 0 | Yoon Shin |
| 15 | Any date | Data | Date format | (3) Proposed | (2) High | Need to inquire to districts of the feasibility/value of switching how they provide date information from month/day/year to year/month/day | 0 | There is a financial impact to districts. With any of these changes, please set the implementation date well into the future so LEAs can plan and give vendors plenty of notice. | Yoon Shin |
| 16 | District Number | Data | How to handle students reported in two districts | (3) Proposed | (2) High | This may be because the student switched mid-year or because they are validly dually enrolled. - need more details here. | 0 | Yoon Shin |
| 17 | Race and Ethnicity | District Feedback | Send Warning to Districts when data is different | (1) Pending | (3) Medium | Should the current and last prior district get a warning report on a nightly basis when the race and/or ethnicity is reported differently. Jan will get feedback from the districts. | 0 | Yoon Shin |
5. **NEW PROCESSES SUGGESTED**

**Current Process**

The team (core team and data governance team together) documented our existing data collection process and drafted new suggested processes for collecting data in a more efficient and effective way to lead to better data quality and consistency across collections. The following diagram documents the existing data collections process:

![Current Process for Collecting Data](image)
**Suggested Processes**

The following diagrams document the suggested new process for data collections:

---

**GO TO 'Future Process for collecting and updating when no SASID is required'**

- **COLLECT DATA** - We will collect data once in the future (except for identifying data such as date of birth, race, gender, etc.), so that the SASID is correct.

- **IDENTIFICATION DATA**:
  - Race: White
  - Gender: Male
  - Birth date: 5/23/2000

- **LAST NAME**: Jones
- **FIRST NAME**: Larry

- **District**: Douglas County

- **Enterprise Information Management Program Pilot**
  - Colorado Department of Education

---

**UPDATE INTEGRATED DATA STORE**

- If the fields match, but partial match, have partial match be sent to CDE, warn the districts and pointing out warnings.

- More communications to districts showing changes in data to districts when there are changes in data.

- New process benefits:
  - Allows districts to have more control over validations and time up their own CDE for most ID management processes.
  - More communications to districts showing changes in data to districts pointing out warnings.

---

**NOTE**: Do we want to notify the districts when a change is made? Should we add the capability of being able to set what warnings for what fields do they want.

---

**New Process for Collecting and updating Data (such as Race and Ethnicity)**

- Include the last 100 days of the student's attendance, with the possibility of it being updated on a daily basis.

---

**New Process for Matching and Updating Identification Data**

- Is a partial match, yes partial match be sent to CDE, warn the districts and pointing out warnings.

---

**Update the fact that there is a new collection was done with a timestamp for it, ideally with a timestamp for it, ideally on a daily basis.**

---

**UPDATE INTEGRATED DATA STORE**

- Update identification data once in the future, and keep their history of the change.

---

**STATE GO WITH OLD INFORMATION**

- Go to 'Future Process for collecting and updating when no SASID is required'.
COLLECT DATA – We will only collect data once in the future (except for identifying data such as date of birth, name, etc).

Is the data (i.e. race and ethnicity) different than the current record?

No

Yes

NOTIFY DISTRICT OF THE CHANGES IN DATA (e.g., GENERATE A REPORT SHOWING THE DISTRICT THE DIFFERENCE IN DATA) (ideally on a daily basis)

UPDATE IDS WITH NEW INFORMATION (AND KEEPING HISTORY)

Did the student change districts?

No

New Process for Collecting and updating Data (such as Race and Ethnicity)

NOTIFY LAST/ PREVIOUS AND CURRENT DISTRICTS (FOR CURRENT YEAR)

OF RECORD OF A DIFFERENCE (e.g., GENERATE A REPORT SHOWING THE DISTRICT THE DIFFERENCE IN DATA) (ideally on a daily basis)

THIS COULD BE A WARNING OR INFORMATIONAL TO DISTRICT

UPDATE INTEGRATED DATA STORE (IDS) WITH ANY NEW INFORMATION (INCLUDING DISTRICT CHANGE, IDENTIFICATION DATA, OR OTHER DATA) (AND KEEP HISTORY BY DISTRICT)

Exact Match?

No

Near Match?

No

Yes

Send to District to Validate

Did District Validate?

No

Can District Correct/ Rectify?

CDE Case Management Resolves

Yes

No

Create New Student

Change identification data (any combination of the 5 fields) and keep history of the change

Load identification data (5 fields)

New process benefits

Allows districts to have more control over validations and frees up time from CDE for most ID management processes

More communications to check identifying changes in data to districts and pointing out warnings

For Charles, no exact match, but partial match

So, Database record in CDE says:

IDENTIFICATION DATA
SASID: 1234567890
First Name: Charles
Middle Name: 
Last Name: Smithe
Birth date: 1/5/2000
Gender: Male
OTHER DATA
Race: White
District: Douglas County

Note: Beyond the pilot, we will need to document reporting rules, for example, when data is different by district and/or time

NOTE: Do we want to notify the district when there are changes? Should we keep score on what changes or what data is used to generate the warning?
6. Lessons Learned

Some of the key lessons that the team learned during the pilot are:

- Defining data elements by committee is very time consuming. This task needs to be done by individuals empowered to do this, and then have the definition reviewed and input by various parties and approved by the accountable person.

- The core team needs to be extremely clear about the vision of EIM, what is expected, roles and responsibilities and the path moving forward.

- Management support, commitment and communication is needed and there needs to be a clear directive so that participants know that they have responsibilities and approved time allocated to enterprise-wide data governance activities.

At a data governance team meeting, the following points were documented regarding what worked and what we could improve as we move forward with the data governance program:

- What worked on the pilot:
  - Small teams meeting on metadata
  - Collective interest in EIM
  - Overall data governance processes and RACI identified
  - Data Governance and Data Modeling Classes/Training
  - Setting small scope
  - Starting out with definitions
  - Cross unit teams

- What to improve –
  - The EIM core team should set clearer expectations of the data governance team
  - Need to better, more frequently, and more specifically communicate the vision of EIM and what will be in the future (business processes and technical infrastructure) as well as be clear about roles and responsibilities. These communications need to be tailored to different audiences
  - Need to improve speed of data governance effort (may be alleviated by having specific roles assigned)
  - Need to get the right people involved
  - Define stakeholders
  - Need to have very specific timelines
  - Need resources from the application development group
7. RECOMMENDATIONS

- Communication and direction from management of allocation of certain percentages of time for data stewards to devote to enterprise wide data stewardship.

- Create an EIM Program handbook, including step-by-step instructions for each role in the EIM process.

- Further develop presentations for various audiences and work on continuously communicating the vision of EIM.

- Need to evaluate CODEX (ADE 2.0) and investigate how this project affects EIM
8. NEXT STEPS

- Develop EIM handbook (in Word) which include:
  - Data strategy - our approach moving forward for the EIM program
  - Enterprise data model - documentation and continued development of this model
  - Data governance (includes data stewardship guide) - a data governance guide
  - Data Architecture - documentation of our future data architecture
  - Change Management - how to manage and implement change

The audience for this document is everyone involved in data management in every role. Stacie Demchak will take the lead on this document.

- Develop presentation materials using Powerpoint for difference audiences in order to provide ongoing clarification and vision for the EIM program including:
  - Data Steward presentation - everything a data steward wants to know about data stewardship such as role description, responsibilities, tool available, etc.
  - EIM Management presentation - this is for CDE management to understand the future direction for the EIM program
  - EIM Business audience within CDE - This presentation will show how the EIM program will affect business representatives at CDE
  - EIM IT audience - This presentation will show how CDE IT can be involved in this effort and how it will affect IT
  - EIM External (districts) - This presentation will show how the EIM program affects parties outside of CDE
  - Provide input regarding communication with executives - perhaps we review presentation materials and review what will be presented to executives

The team has many materials that we can use for these presentations and we will re-use and reformat existing presentations.

- Evaluate and strategize how we appropriately establish accountabilities and responsibilities

- Need to continue developing the EIM Site Sharepoint

- Need to investigate CODEX (ADE 2.0) and how architecture and business processes are affected

- Need data governance plan and timeline, including clear planning for the necessary resources involved